

Ancestral Reconstruction and C₃ Bundle Sheath Transcript Abundance in the Paniceae Grasses Indicate the Foundations for all Three Biochemical C₄ Sub-Types Were Likely Present in the Most Recent Ancestor

Jacob D. Washburn¹, Satya S. Kothapalli¹, Julia M. Brose¹, Sarah Covshoff², Julian M. Hibberd², Gavin C. Conant^{3,4}, J. Chris Pires¹

1) Division of Biological Sciences, University of Missouri, 311 Bond Life Science Center, Columbia, MO 65211 USA

2) Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA UK

3) Division of Animal Sciences, 920 East Campus Drive, University of Missouri-Columbia, USA 65211

4) Program in Genetics, Bioinformatics Research Center, Department of Biological Sciences, 356 Ricks Hall, North Carolina State University, Raleigh, NC, USA 27695

ABSTRACT

In C₄ plants the enzymatic machinery underpinning photosynthesis can vary, with for example, three distinct C₄ acid decarboxylases being used to release CO₂ in the vicinity of RuBisCO. For decades, these decarboxylases have been used to classify C₄ species into three biochemical sub-types. However, more recently the notion that C₄ species mix and match C₄ acid decarboxylases has increased in popularity and, as a consequence, the validity of specific biochemical sub-types has been questioned. Using species from the grass tribe Paniceae we show that whilst transcripts encoding multiple C₄ acid decarboxylases accumulate in bundle sheath cells in some species, in others, transcripts encoding only one enzyme are detected. In addition, a method that allows isolation of bundle sheath cells from a C₃ species within the Paniceae, *Sacciolepis indica*, was developed. Deep sequencing of bundle sheath preparations from all four species combined with ancestral state reconstruction support the notion that the three biochemical C₄ sub-types found in the Paniceae existed together in their most recent common ancestor. Thus, these species likely inherited the functional building blocks of all three C₄ pathways. We conclude that classification of C₄ plants into the classical biochemical sub-types is still appropriate for some species, and that evolution of this trait has been facilitated by characteristics of the ancestral C₃ bundle sheath and made use of multiple convergent routes involving either one or multiple C₄ acid decarboxylases.

INTRODUCTION

C₄ photosynthesis is considered the most productive mechanism by which plants convert sunlight into chemical energy. In all cases, the C₄ pathway leads to increased photosynthetic efficiency because high concentrations of CO₂ are supplied to RuBisCO. Since its discovery in the 1960s (Hatch and Slack, 1966), a unified understanding of the biochemistry underpinning C₄ photosynthesis has emerged. This basic system comprises a biochemical pump that initially fixes HCO₃⁻ into C₄ acids in mesophyll (MS) cells. Subsequently, diffusion of these C₄ acids into a separate compartment, followed by their decarboxylation, generates high concentrations of CO₂ around RuBisCO. In many plants, the release of CO₂ 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, the C₄ pathway increases photosynthetic efficiency compared with the ancestral C₃ state.

Elucidation of the C₄ pathway was based on analysis of sugar-cane and maize which both use the chloroplastic NADP-dependent malic enzyme (NADP-ME) to release CO₂ 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 phosphoenolpyruvate carboxykinase (PCK) (Edwards et al., 1971) to release CO₂ in the BS, whereas *Atriplex spongiosa* and *Panicum miliaceum* contained high activities of the mitochondrial NAD-dependent malic enzyme (NAD-ME) (Hatch and Kagawa, 1974). These findings led to the consensus that different C₄ species made preferential use of one C₄ acid decarboxylase, and the classification of C₄ plants into one

of three distinct biochemical pathways ensued (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. Corn (*Zea mays*), for example, was traditionally classified as using NADP-ME but evidence has mounted that it also has high activities of PCK (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). This blurring of the NADP-ME C_4 sub-type coincided with observations that many plants with high amounts of PCK also contained either NADP-ME or 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 PCK decarboxylation (Wang et al., 2014). It has therefore been suggested that PCK 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 small 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 + PCK, and NAD-ME + PCK sub-types (Wang et al., 2014; Washburn et al., 2015; Rao and Dixon, 2016). At present, none of these classification schemes has been widely

63 adopted by the community. Moreover, convincing experimental evidence (i.e., transcriptomic,
64 or proteomic data from multiple species) of traditionally PCK sub-type species using another
65 decarboxylation enzyme at a higher level than PCK is lacking, while enzyme activity
66 measurements in the older literature indicate PCK predominance for several species (Gutierrez
67 et al., 1974; Prendergast et al., 1987; Lin et al., 1993).

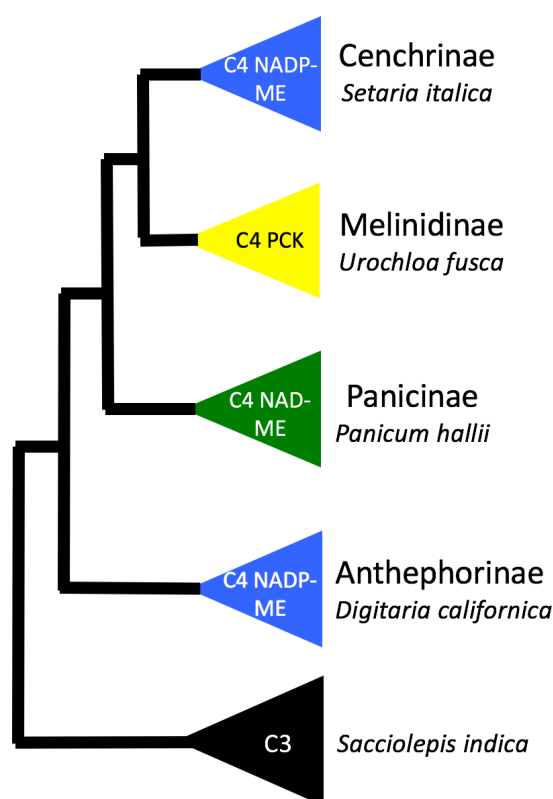


Figure 1. Species used in this study as well as their phylogenetic placement within the tribe Paniceae (Poaceae) and four of its sub-tribes. The photosynthetic type (C₃ or C₄) and C₄ sub-type of each species is labeled in the colored triangle next to it. NADP-ME = NADP-dependent malic enzyme, PCK = phosphoenolpyruvate carboxykinase, NAD-ME = NAD-dependent malic enzyme. Phylogenetic placement based on Vicentini et al. (2008) and Washburn et al. (In Review).

To better understand whether the PCK pathway represents a true biochemical sub-type, and more generally to investigate the extent to which C₄ species make use of mixtures of C₄ acid decarboxylases, global patterns of mRNA abundance were assessed from BS and MS enriched samples across phylogenetically-spaced C₄ plants that are traditionally defined as exclusively using one of each of the C₄ sub-types. These species belong to the grass tribe Paniceae (Poaceae), the only known group previously documented as containing all three biochemical sub-types of C₄ photosynthesis (Sage et al., 2011). The C₃ species *Sacciolepis indica*, another member of the Paniceae, was also included in this analysis to provide insight into the ancestral state in the tribe and the evolutionary transition from C₃ to different C₄ sub-types.

RESULTS

Distribution of Transcripts Encoding Components of The Core C₄ Cycle

Four C₄ species from the Paniceae tribe that have been reported to use different C₄ acid decarboxylases were selected for analysis (Figure 1). *Setaria italica* is a member of the Cenchrinae subtribe within the Paniceae and belongs to the classical NADP-ME C₄ sub-type. *Urochloa fusca* is in the subtribe Melinidinae which is classically defined as using the PCK sub-type. *Panicum hallii* belongs to the Panicinae subtribe and is reported to use NAD-ME, and lastly *Digitaria californica*, a member of the Anthephorineae subtribe, is defined as conforming to the NADP-ME sub-type (Gutierrez et al., 1974; Prendergast et al., 1987; Lin et al., 1993). Although belonging to distinct C₄ biochemical sub-types these species are derived from a common ancestor (Figure 1). *Sacciolepis indica* represents the C₃ ancestral state, and is phylogenetically sister to the other four species while still being part of the tribe Paniceae.

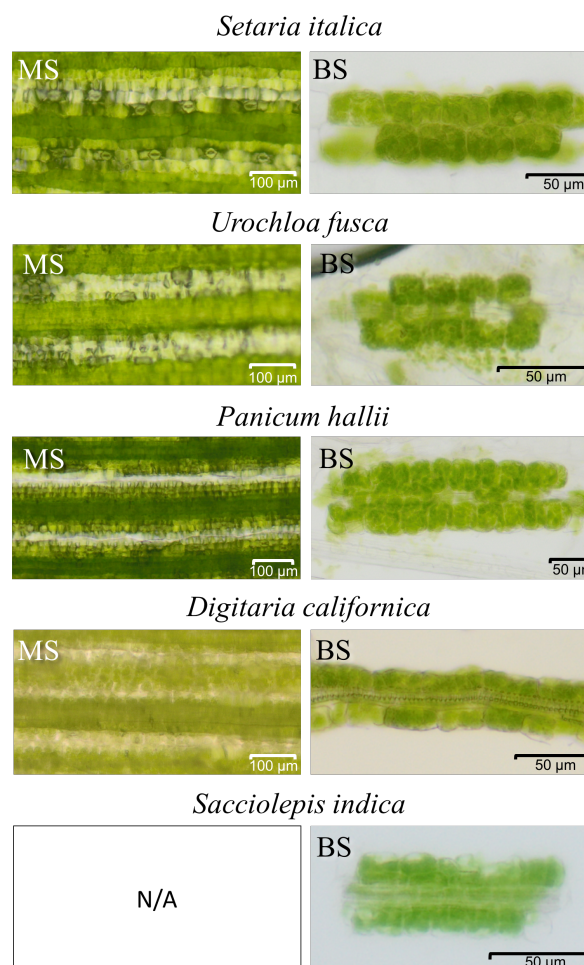


Figure 2. Representative images from leaves that have been rolled to remove mesophyll (MS) sap (Left) or bundle sheath (BS) strands after isolation (Right). All species use the C₄ pathway except *Sacciolepis indica* which is a C₃ plant. The bands of cells with low chlorophyll content in images on the left-hand side represent the position of mesophyll cells that have collapsed and sap expelled during the rolling procedure. Scale bars are depicted.

Microscopic examination of leaves of *S. italica*, *U. fusca*, *P. hallii* and *D. californica* from which MS cell sap had been extracted using leaf rolling techniques, indicated bands of cells containing low chlorophyll content (Figure 2) a phenotype consistent with efficient removal of MS sap (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. Separation of BS strands was also successful for the C_3 species *Sacciolepis indica* (Figure 2). Analysis of transcripts derived from core C_4 genes showed clear differences in abundance between MS and BS samples. For example, transcripts derived from carbonic anhydrase (CA), phosphoenolpyruvate carboxylase (PEPC) and pyruvate, orthophosphate dikinase (PPDK) genes preferentially accumulated in MS cells (Figure 3A). In contrast, transcripts derived from the rubisco small

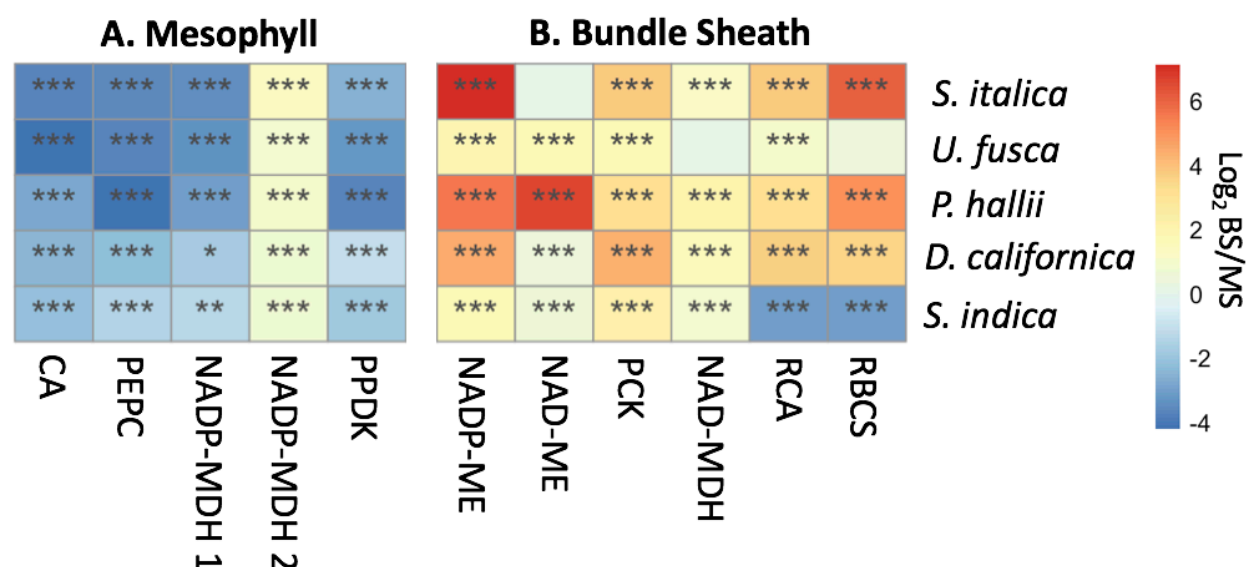


Figure 3. Log₂ fold change between mesophyll (MS) and bundle sheath (BS) enriched mRNA transcripts of *Setaria italica*, *Urochloa fusca*, *Panicum hallii* and *Digitaria californica*, or between whole leaf and BS strands for C_3 *Sacciolepis indica*. Genes depicted encode proteins of the core C_4 cycle that are known to be preferentially expressed in either: A) MS, 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 1 or 2 = NADP-dependent malate dehydrogenase (two different isoforms), PPDK = pyruvate, orthophosphate dikinase, NADP-ME = NADP-dependent malic enzyme, PCK = phosphoenolpyruvate carboxykinase, NAD-ME = NAD-dependent malic enzyme, NAD-MDH = NADP-dependent malate dehydrogenase, RCA = rubisco activase, RBCS = rubisco small subunit.

subunit (RBCS) and rubisco activase (RCA) as well as either NADP-ME, NAD-ME or PCK were more abundant in BS strands (Figure 3B). These differences are consistent with our current understanding of transcript distribution between MS and BS cells in C_4 grasses (Covshoff et al., 2013; John et al., 2014). The abundance of transcripts relating to C_4 photosynthesis in the C_3 species *S. indica* is also consistent with current knowledge of metabolism in the BS of C_3 species. For example, transcripts of RBCS and RCA were less abundant in BS strands than in whole leaf samples, which supports the notion that the C_3 BS being less photosynthetic than C_3 MS cells (Kinsman and Pyke, 1998; Leegood, 2002; Aubry et al., 2014). Furthermore, although not to the same degree as in the C_4 species sampled, it was also the case that transcripts encoding each of the C_4 acid decarboxylases were more abundant in the BS strands than in the whole leaf of *S. indica* (Figure 3). This is consistent with the finding that cells near veins of C_3 species contain relatively high activities of these C_4 enzymes (Hibberd and Quick, 2002; Brown et al., 2010).

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

Setaria italica has classically been defined as belonging to the NADP-ME C_4 sub-type (Gutierrez et al., 1974; Prendergast et al., 1987; Lin et al., 1993). Accumulation of the NADP-ME and NADP-MDH enzymes in BS and MS cells respectively is indicative of this sub-type, and their transcript distributions confirmed this expectation. In addition, consistent with knowledge of the NADP-ME sub-type, in BS strands of *S. italica* transcripts encoding PCK, NAD-ME, NAD-MDH, ASP-AT, and ALA-AT were detected at low levels. The data therefore indicate that under the

121 conditions we used for sampling, *S. italica* was likely making near exclusive use of NADP-ME to
 122 decarboxylate malate in the BS (Figure 4A).

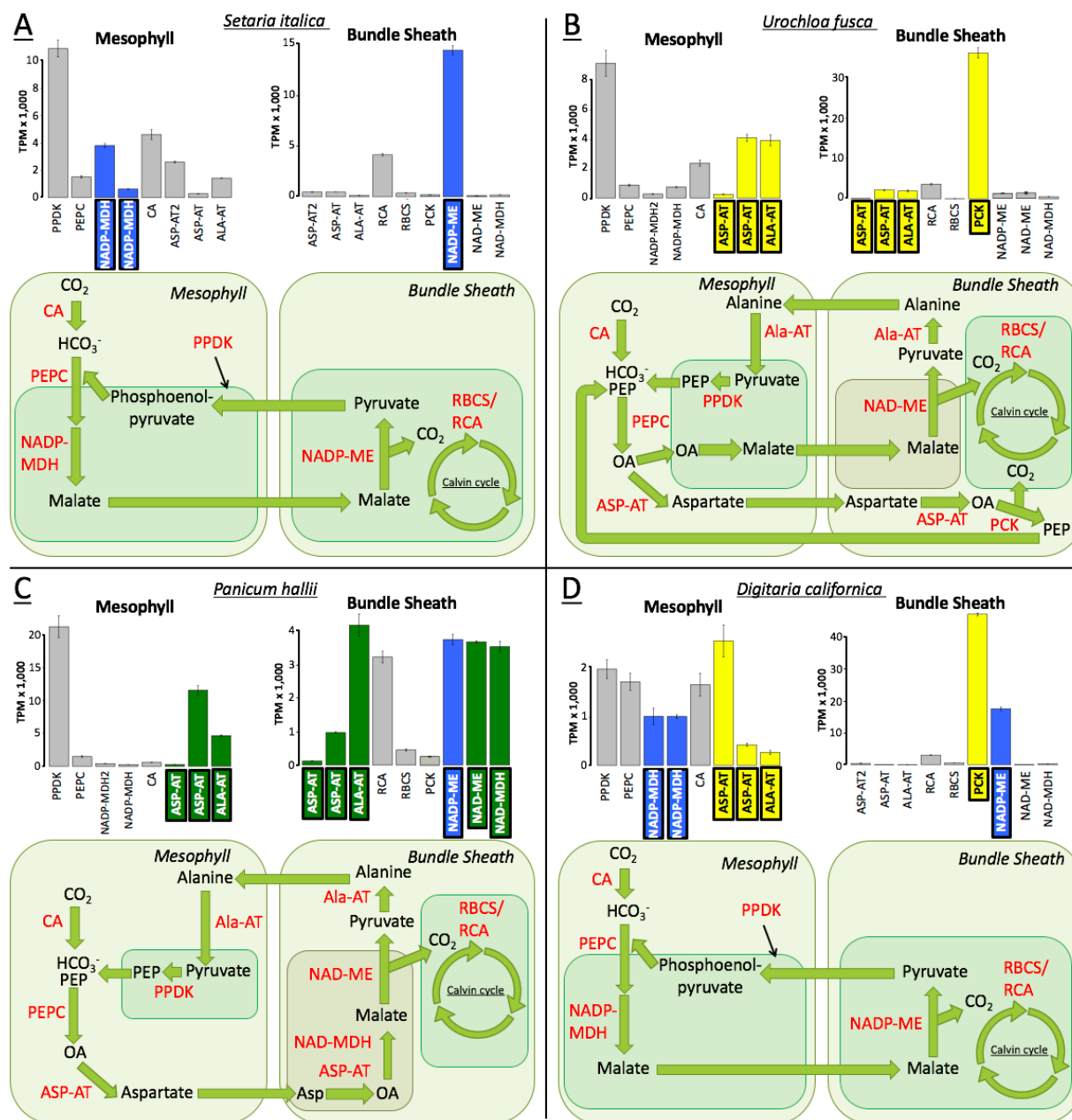


Figure 4. Relative transcript abundance between core C₄ pathway enzymes within mesophyll (MS) and bundle sheath (BS) of: A) *Setaria italica*, B) *Urochloa fusca*, C) *Panicum hallii*, and D) *Digitaria californica*. The schematics below each histogram indicate the enzyme complement associated with each of the three biochemical sub-types. CA = carbonic anhydrase, PEPC = phosphoenolpyruvate carboxylase, NADP-MDH 1 or 2 = NADP-dependent malate dehydrogenase (two different isoforms), PPDK = pyruvate,orthophosphate dikinase, NADP-ME = NADP-dependent malic enzyme, PCK = phosphoenolpyruvate carboxykinase, NAD-ME = NAD-dependent malic enzyme, NAD-MDH = NADP-dependent malate dehydrogenase, RCA = rubisco activase, RBCS = rubisco small subunit, asparagine-aminotransferase (ASP-AT) and alanine-amino transferase (ALA-AT)

Urochloa fusca is thought to exclusively use PCK to release CO₂ in the BS (Gutierrez et al., 1974; Prendergast et al., 1987; Lin et al., 1993). The patterns of transcript accumulation in MS and BS strands of *U. fusca* are consistent with PCK functioning in this species without supplemental decarboxylation from either NADP-ME or NAD-ME (Figure 4B). For example, BS stands contained barely detectable levels of transcripts encoding NADP-ME and NAD-ME, but very high levels of those encoding PCK. 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 both cell-types (Figure 4B).

In contrast to the above analysis of *S. italica* and *U. fusca*, which supports the contention that they make almost exclusive use of NADP-ME and PCK respectively, analysis of *P. hallii* and *D. californica* indicates that they likely use multiple C₄ acid decarboxylases during photosynthesis. Although *P. hallii* is classically considered as using NAD-ME (Gutierrez et al., 1974; Prendergast et al., 1987; Lin et al., 1993) in addition to high levels of transcripts encoding NAD-ME, NAD-MDH, ASP-AT and ALA-AT, unexpectedly high levels of transcripts encoding NADP-ME were detected in the BS (Figure 4C). In the case of *Digitaria californica* which is thought to belong to the NADP-ME sub-type, although transcripts encoding NADP-ME and NADP-MDH were abundant in BS and MS samples respectively, PCK levels were more than double those of NADP-ME in the BS. ASP-AT levels in the MS cells were also higher than expected for an NADP-ME sub-type plant. Overall, these data are consistent with *S. italica* and *U. fusca* using a single C₄ acid decarboxylase in the BS, whereas *P. hallii* and *D. californica* do not.

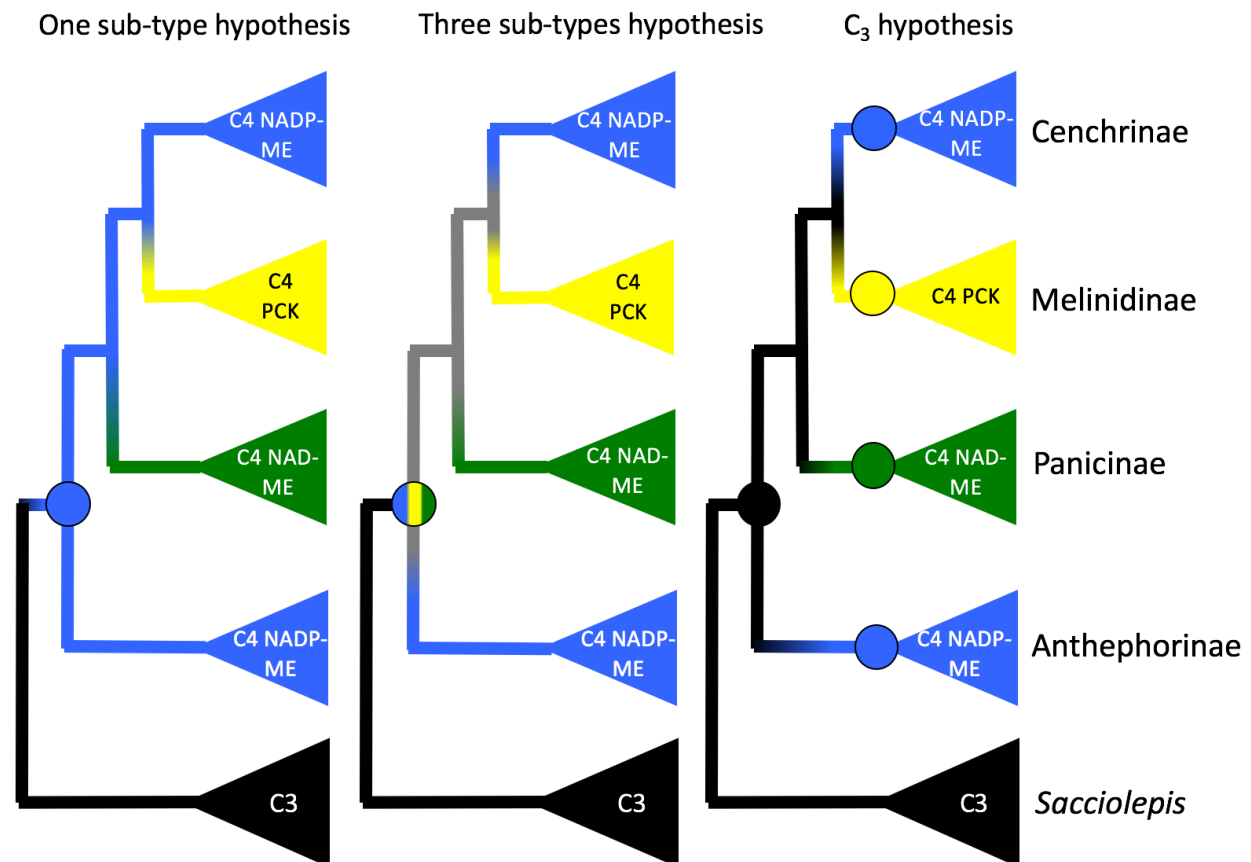


Figure 5. Three hypotheses for the evolution of C₄ sub-types within the MPCA clade (Melinidinae, Panicinae, Cenchrinae, Anthephorinae) of the tribe Paniceae. The one sub-type hypothesis posits that the most recent common ancestor (MRCA) of the group utilized one sub-type exclusively, and the other types evolved from it in a step-wise fashion. The three sub-type hypothesis suggests that all three sub-types existed in the MRCA and then each has become dominant in one clade or another over time. The C₃ hypothesis is based on the idea that each of the sub-types evolved independently from a C₃ ancestor. Hypothesizes and phylogeny based on Washburn et al. (2015) and Washburn et al. (In Review). NADP-ME = NADP-dependent malic enzyme, PCK = phosphoenolpyruvate carboxykinase, NAD-ME = NAD-dependent malic enzyme.

145

146 Ancestral State Reconstruction of the MPCA Clade

147 Several hypotheses relating to how the MPCA (Melinidinae, Panicinae, Cenchrinae, and
148 Anthephorinae) clade may have evolved have been proposed (Figure 5). While they have been
149 examined using categorical data, this is the first time that ancestral state reconstructions of the
150 MPCA have been performed using transcript abundance levels. The results at each node of the

MPCA phylogeny show mixtures of all three primary sub-type enzymes (Figure 6). Although this approach indicates that transcripts for NADP-ME and PCK would have been the most abundant in the last common ancestor of these C₄ species, it was also the case that substantial levels of NAD-ME are predicted.

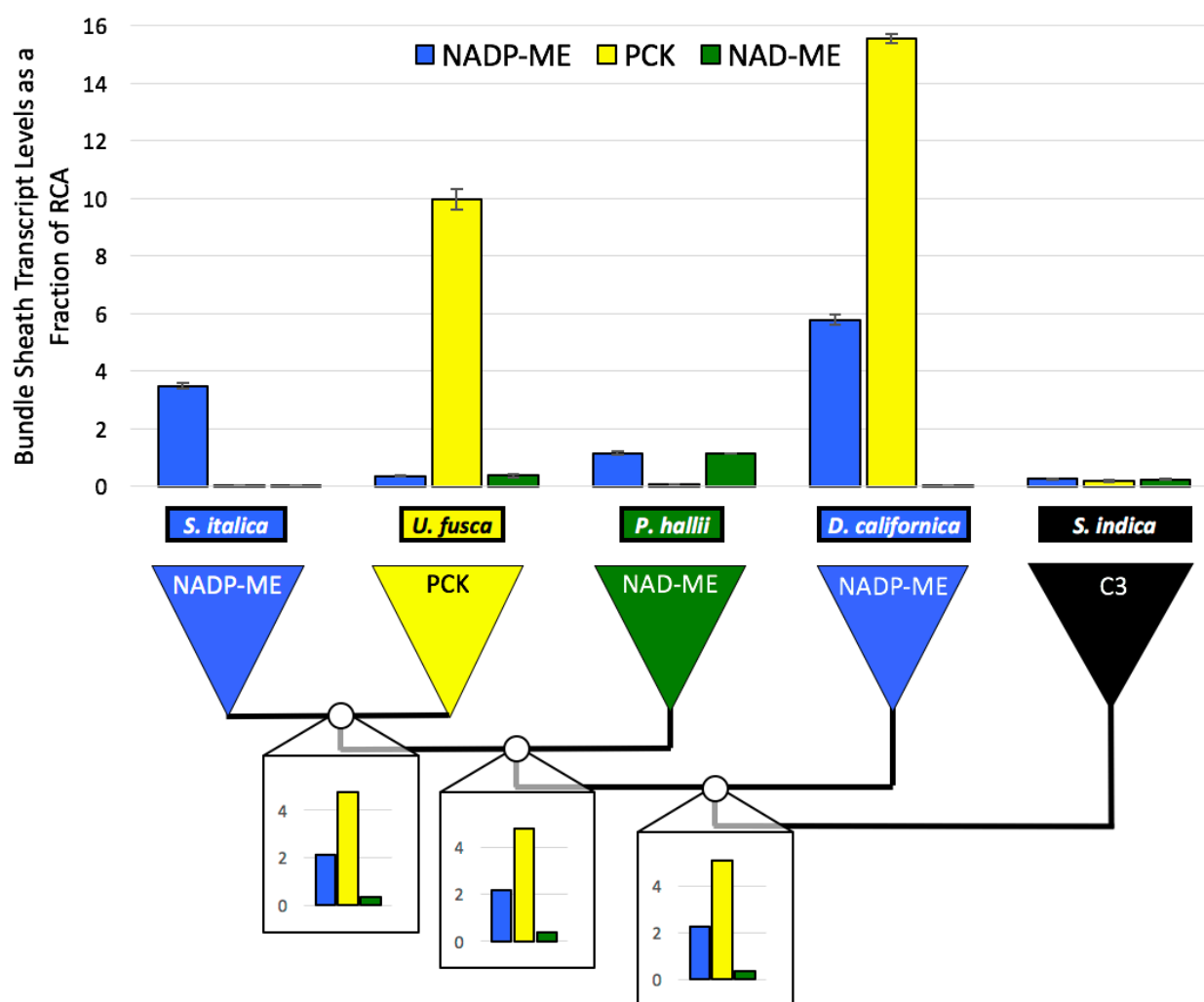


Figure 6. Comparisons between species for transcript abundance of NADP-dependent malic enzyme (NADP-ME), phosphoenolpyruvate carboxykinase (PCK), and NAD-dependent malic enzyme (NAD-ME) transcripts in Bundle Sheath strands. Below that, the nuclear gene phylogenetic relationships and ancestral state reconstruction of transcript abundance levels at each node are displayed. Transcript levels are normalized to those of Rubisco activase (RCA). Phylogenetic relationships based on Vicentini et al. (2008) and Washburn et al. (In Review).

DISCUSSION

The PCK Sub-Type

The dominance of PCK transcripts over NADP-ME and NAD-ME in *U. fusca* provide evidence for the biological relevance of the classical PCK sub-type. These data are in direct contrast to recent proposals that PCK cannot function on its own but rather is always ancillary to one of the other two C₄ 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 PCK 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.

In the case of *U. fusca*, it appears that the use of alternative sub-type classification systems where PCK is considered ancillary to either NADP-ME or NAD-ME (Wang et al., 2014; Washburn et al., 2015; Rao and Dixon, 2016) is not useful. In fact, NAD-ME and NADP-ME transcript abundance in *U. fusca* were statistically indistinguishable making this species impossible to classify into the proposed two and four sub-type systems.

C₄ Sub-Type Mixing

Of the four C₄ species examined, *P. hallii* and *D. californica* show evidence of extensive 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 evidence of sub-type mixing. In this case, NADP-ME and PCK transcripts were both reasonably abundant. Although this species is classically considered to belong to the NADP-ME sub-type, transcripts encoding PCK were more than double the abundance of those of NADP-ME. ASP-AT levels, which are also associated with the PCK sub-type, were high as well. This suggests that *D. californica* may use PCK as the dominant decarboxylation enzyme and NADP-ME secondarily.

Bundle Sheath Enriched Transcript Abundance Levels in C₃ Species and the Evolution of Three Sub-Types Within a Single Clade

Analysis of transcript abundance in MS and BS cells from C₃ species that are closely related to C₄ species is critical to understanding how C₄ photosynthesis evolved, and how it can be engineered for enhanced crop production. Although transcripts loaded onto ribosomes in the BS of C₃ *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 is no equivalent data from a monocotyledonous lineage in which both C₃ and C₄ species are found. The relatively high levels of C₄ related transcripts in the BS of *S. indica* are consistent with previous work on cells around the veins of C₃ plants (Hibberd and Quick, 2002; Brown et al., 2010; Shen et al., 2016). Together these data support the concept that C₃ species are in some ways pre-

adapted to adopt the C_4 mechanism (Gould, 1989; Christin et al., 2009; Christin et al., 2015; Washburn et al., 2016).

The results from ancestral state reconstruction favor the notion that the three C_4 sub-types in MPCA clade in the Paniceae may have existed together in their most recent common ancestor. The result that two of the four species sampled may be mixing sub-types is further evidence for this hypothesis. Additionally, the presence of transcripts associated with multiple enzymes involved in each of the three C_4 sub-types in the C_3 species *S. indica* supports all three sub-types existing together in the MRCA rather than each sub-type evolving independently, or sub-types evolving in a step-wise fashion one from another.

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/>.

All plant materials 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 Materials S1 for complete modified protocol) (Sheen and Bogorad, 1985; Chang et al., 2012; Covshoff et al., 2013; John et al., 2014).

The protocol used for obtaining BS strands in *S. indica* was the same as that used for the *C₄* 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 results were qualitatively similar samples with less apparent contamination. One step that may have been key to the isolation of *C₃* 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 MS sap prior to BS isolation.

MS enriched samples were not successfully obtained for *S. indica* in this study because of the sensitivity of the *C₃* leaves to rolling. Very small amounts of leaf rolling pressure resulted in the leaves becoming damaged to the point that the purity of MS sap obtained from them was called into question. Leaves that were rolled gently enough not to damage the BS strands and contaminate the MS sap resulted in sap with insufficient quantities of RNA for sequencing. It is our opinion that MS sap could be sampled using: 1) low input mRNA extraction and sequencing procedures, 2) a more precise instrument for leaf rolling such as that described by Leegood (1985), and/or 3) further experimentation with the developmental stage at which MS sap is extracted.

Sequencing

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

Analysis

Each mRNA sample was quality trimmed and mapped to the *Sorghum bicolor* genome (Paterson et al., 2009; DOE-JGI, 2017) using Trimmomatic and Trinity 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 RSEM and Bowtie2 for read mapping and counting as well as edgeR and DESeq 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 C₄ photosynthesis genes was compiled based on the literature; a custom script and BLAST were then used to find the appropriate homologous genes for each species 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). Ancestral state reconstructions were performed using the “Trace Character History” command in Mesquite (Maddison and Maddison, 2017).

Transcript Normalization and Transcriptome Size Estimation

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 DESeq was used. For comparisons of BS enhanced transcripts between the different species, transcript abundance levels were normalized to Rubisco activase (RCA). In the current study, BS cell analyses were tested separately using normalization to RCA and TMM and the results were qualitatively identical.

ACCESSION NUMBERS

Sequence data are available on NCBI SRA (<https://www.ncbi.nlm.nih.gov/sra>) under the following identifiers: XXX

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 (DEB Award no. 1501406).

LITERATURE CITED

Aubry S, Smith-Unna RD, Bournsnell 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 C4 metabolic processes between mesophyll and bundle sheath cells allows light capture to be balanced for the maize C4 pathway. *Plant Physiology*

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

Bräutigam A, Schliesky S, Külahoglu C, Osborne CP, Weber APM (2014) Towards an integrative model of C4 photosynthetic subtypes: insights from comparative transcriptome analysis of NAD-ME, NADP-ME, and PEP-CK C4 species. *Journal of Experimental Botany* **65**: 3579-3593

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) C4 acid decarboxylases required for C4 photosynthesis are active in the mid-vein of the C3 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

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

C4 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 C4 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 C4-dicarboxylic acid cycle of photosynthesis. Biochemical and Biophysical Research Communications **45**: 278-285

Furbank RT (2011) Evolution of the C4 photosynthetic mechanism: are there really three C4

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

Gould SJ (1989) *Wonderful life: the Burgess Shale and the nature of history*, Ed 1st. W.W. Norton, New York

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**: 644-652

Gutierrez M, Gracen VE, Edwards GE (1974) Biochemical and cytological relationships in C4 plants. *Planta* **119**: 279-300

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protocols* **8**: 1494-1512

Hatch M, Kagawa T, Craig S (1975) Subdivision of C4-Pathway Species Based on Differing C4 Acid Decarboxylating Systems and Ultrastructural Features. *Functional Plant Biology* **2**: 111-128

Hatch M, Slack C (1966) Photosynthesis by sugar-cane leaves. A new carboxylation reaction and 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 C4-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 C4 pathway photosynthesis. *Archives of Biochemistry and Biophysics* **175**: 39-53

Hibberd JM, Covshoff S (2010) The regulation of gene expression required for C4 photosynthesis. *Annu Rev Plant Biol* **61**: 181-207

Hibberd JM, Quick WP (2002) Characteristics of C4 photosynthesis in stems and petioles of C3 flowering plants. *Nature* **415**: 451-454

John CR, Smith-Unna RD, Woodfield H, Covshoff S, Hibberd JM (2014) Evolutionary Convergence of Cell-Specific Gene Expression in Independent Lineages of C4 Grasses. *Plant Physiology* **165**: 62-75

Kinsman EA, Pyke KA (1998) Bundle sheath cells and cell-specific plastid development in *Arabidopsis* leaves. *Development* **125**: 1815-1822

- Koteyeva NK, Voznesenskaya EV, Edwards GE** (2015) An assessment of the capacity for phosphoenolpyruvate carboxykinase to contribute to C4 photosynthesis. *Plant Science* **235**: 70-80
- 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
- Leegood RC** (2002) C4 photosynthesis: principles of CO2 concentration and prospects for its introduction into C3 plants. *Journal of Experimental Botany* **53**: 581-590
- 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**
- Maddison WP, Maddison DR** (2017) Mesquite: a modular system for evolutionary analysis. *In*, Ed Version 3.2
- Majeran W, Friso G, Ponnala L, Connolly B, Huang M, Reidel E, Zhang C, Asakura Y, Bhuiyan NH, Sun Q, Turgeon R, van Wijk KJ** (2010) Structural and Metabolic Transitions of C(4) Leaf Development and Differentiation Defined by Microscopy and Quantitative Proteomics in Maize. *The Plant Cell* **22**: 3509-3542
- McCarthy DJ, Chen Y, Smyth GK** (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research* **40**: 4288-4297

Meyer E, Aspinwall MJ, Lowry DB, Palacio-Mejía JD, Logan TL, Fay PA, Juenger TE (2014)

Integrating transcriptional, metabolomic, and physiological responses to drought stress and recovery in switchgrass (*Panicum virgatum* L.). *BMC Genomics* **15**: 527

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G,

Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T,

Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M,

Narechania A, Otiillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC,

Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R,

Peterson DG, Mehboob ur R, Ware D, Westhoff P, Mayer KFX, Messing J, Rokhsar DS

(2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature* **457**: 551-556

Pick TR, Brautigam A, Schluter U, Denton AK, Colmsee C, Scholz U, Fahnenstich H, Pieruschka

R, Rascher U, Sonnewald U, Weber AP (2011) Systems analysis of a maize leaf

developmental gradient redefines the current C4 model and provides candidates for regulation. *Plant Cell* **23**: 4208-4220

Prendergast H, Hattersley P, Stone N (1987) New structural/biochemical associations in leaf

blades of C4 grasses (Poaceae). *Functional Plant Biology* **14**: 403-420

Rao X, Dixon RA (2016) The Differences between NAD-ME and NADP-ME Subtypes of C4

Photosynthesis: More than Decarboxylating Enzymes. *Frontiers in Plant Science* **7**

Rao X, Lu N, Li G, Nakashima J, Tang Y, Dixon RA (2016) Comparative cell-specific

transcriptomics reveals differentiation of C4 photosynthesis pathways in switchgrass and other C4 lineages. *Journal of Experimental Botany* **67**: 1649-1662

- Robinson MD, McCarthy DJ, Smyth GK** (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**
- Robinson MD, Oshlack A** (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* **11**: R25
- Sage RF, Christin PA, Edwards EJ** (2011) The C4 plant lineages of planet earth. *Journal of Experimental Botany* **62**: 3155-3169
- 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** (1999) C4 GENE EXPRESSION. *Annual Review of Plant Physiology & Plant Molecular Biology* **50**: 187
- 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
- Shen W, Ye L, Ma J, Yuan Z, Zheng B, LV C, Zhu Z, Chen X, Gao Z, Chen G** (2016) The existence of C4-bundle-sheath-like photosynthesis in the mid-vein of C3 rice. *Rice* **9**: 20
- 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 C4-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 C4 photosynthesis. *Global Change Biology* **14**: 2963-2977
- von Caemmerer S, Ghannoum O, Furbank RT** (2017) C4 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 C4 Plants: Its Role and Regulation. *Functional Plant Biology* **24**: 459-468

Wang Y, Bräutigam A, Weber APM, Zhu X-G (2014) Three distinct biochemical subtypes of C4 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 (In Review) Genome-Guided Phylo-Transcriptomic Methods and the Nuclear Phylogentic Tree of the Paniceae Grasses.

Washburn JD, Schnable JC, Davidse G, Pires JC (2015) Phylogeny and photosynthesis of the grass tribe Paniceae. *American Journal of Botany* **102**: 1493-1505

Weissmann S, Ma F, Furuyama K, Gierse J, Berg H, Shao Y, Taniguchi M, Allen DK, Brutnell TP (2016) Interactions of C4 Subtype Metabolic Activities and Transport in Maize Are Revealed through the Characterization of DCT2 Mutants. *The Plant Cell* **28**: 466-484

Wingler A, Robert PW, Zhi-Hui C, Leegood RC (1999) Phosphoenolpyruvate Carboxykinase Is Involved in the Decarboxylation of Aspartate in the Bundle Sheath of Maize. *Plant Physiology* **120**: 539-545

Zhang J-Y, Lee Y-C, Torres-Jerez I, Wang M, Yin Y, Chou W-C, He J, Shen H, Srivastava AC, Pennacchio C, Lindquist E, Grimwood J, Schmutz J, Xu Y, Sharma M, Sharma R, Bartley LE, Ronald PC, Saha MC, Dixon RA, Tang Y, Udvardi MK (2013) Development of an

integrated transcript sequence database and a gene expression atlas for gene discovery
and analysis in switchgrass (*Panicum virgatum* L.). *The Plant Journal* **74**: 160-173