Transcriptomics of C4 photosynthesis in rice paddy

Covshoff, Sarah; Szecowka, Marek; Hughes, Thomas; Smith-Unna, Richard; Kelly, Steve; Bailey, Karen; Sage, Tammy; Pachebat, Justin; Leegood, Richard; Hibberd, Julian

Published in:
Plant Physiology

DOI:
10.1104/pp.15.00889

Publication date:
2016

Citation for published version (APA):
Covshoff, S., Szecowka, M., Hughes, T., Smith-Unna, R., Kelly, S., Bailey, K., Sage, T., Pachebat, J., Leegood, R., & Hibberd, J. (2016). Transcriptomics of C4 photosynthesis in rice paddy. *Plant Physiology*, *170*(1), 57-73. https://doi.org/10.1104/pp.15.00889

C_4 Photosynthesis in the Rice Paddy: Insights from the Noxious Weed *Echinochloa glabrescens*[1][OPEN]

Sarah Covshoff, Marek Szecowka, Thomas E. Hughes[2], Richard Smith-Unna, Steven Kelly, Karen J. Bailey, Tammy L. Sage, Justin A. Pachebat[3], Richard Leegood, and Julian M. Hibberd*

Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, United Kingdom (S.C., M.S., T.E.H., R.S.-U., J.A.P., J.M.H.); Department of Plant Sciences, University of Oxford, Oxford OX1 3RB, United Kingdom (S.K.); Department of Animal and Plant Sciences, Alfred Denny Building, University of Sheffield, Western Bank, Sheffield S10 2TN, United Kingdom (K.J.B., R.L.); and Department of Ecology and Evolutionary Biology, 25 Willcocks Street, University of Toronto, Toronto, Ontario, Canada M5S 3B2 (T.L.S.)

ORCID IDs: 0000-0001-8721-7197 (R.S.-U.); 0000-0003-0662-7958 (J.M.H.).

The C_4 pathway is a highly complex trait that increases photosynthetic efficiency in more than 60 plant lineages. Although the majority of C_4 plants occupy disturbed, arid, and nutrient-poor habitats, some grow in high-nutrient, waterlogged conditions. One such example is *Echinochloa glabrescens*, which is an aggressive weed of rice paddies. We generated comprehensive transcriptome datasets for C_4 *E. glabrescens* and C_3 rice to identify genes associated with adaption to waterlogged, nutrient-replete conditions, but also used the data to better understand how C_4 photosynthesis operates in these conditions. Leaves of *E. glabrescens* exhibited classical Kranz anatomy with lightly lobed mesophyll cells having low chloroplast coverage. As with rice and other hygrophytic C_3 species, leaves of *E. glabrescens* accumulated a chloroplastic phospho*enol*pyruvate carboxylase protein, albeit at reduced amounts relative to rice. The arid-grown species *Setaria italica* (C_4) and *Brachypodium distachyon* (C_3) were also found to accumulate chloroplastic phospho*enol*pyruvate carboxylase. We identified a molecular signature associated with C_4 photosynthesis in nutrient-replete, waterlogged conditions that is highly similar to those previously reported from C_4 plants that grow in more arid conditions. We also identified a cohort of genes that have been subjected to a selective sweep associated with growth in paddy conditions. Overall, this approach highlights the value of using wild species such as weeds to identify adaptions to specific conditions associated with high-yielding crops in agriculture.

All photosynthetic organisms rely on the primary carboxylase activity of Rubisco to produce carbohydrates for life. However, since Rubisco evolved with a limited capacity to distinguish between CO_2 and oxygen (Sage, 2004), rates of oxygenation can have a significant impact on photosynthetic efficiency (Ehleringer and Monson, 1993). In response to evolutionary pressures, approximately 30 million years ago a complex phenotype known as C_4 photosynthesis evolved in more than 60 plant lineages (Sage et al., 2011, 2012). C_4 photosynthesis is a carbon-concentrating mechanism that allows Rubisco to operate in a CO_2-enriched environment, thus avoiding the problems of oxygenation. Phospho*enol*pyruvate carboxylase (PEPC), a cytosolic enzyme that does not react with oxygen and has a higher affinity for carbon than Rubisco, performs the initial carboxylation reaction (Hatch, 1987). The resulting C_4 acids are subsequently decarboxylated in the vicinity of compartmentalized, chloroplastic Rubisco. At least 25 anatomical types capable of achieving the necessary compartmentation for C_4 photosynthesis have been described (Edwards and Voznesenskaya, 2011). In most C_4 species, specialized cell-types within the leaf are arranged in "Kranz" formation (Sage, 2004). Veins are surrounded by concentric rings of bundle sheath (BS) and then mesophyll (M) cells, with PEPC and Rubisco operating in M and BS cells, respectively, and organic acids moving between them.

The C_4 biochemical cycle also varies between lineages of C_4 plants and has commonly been categorized into three subtypes (Fig. 1). These biochemical subtypes are based on whether NADP-dependent malic enzyme (NADP-ME), NAD-dependent malic enzyme (NAD-ME), or phospho*enol*pyruvate carboxykinase (PEPCK) releases CO_2 from the C_4 acids. However, it is now recognized that mixed subtypes utilizing multiple C_4 acid

[1] This work was supported by the Bill and Melinda Gates Foundation and UKAID.

[2] Present address: Department of Plant Sciences, University of Oxford, Oxford OX1 3RB, UK.

[3] Present address: Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Penglais, Aberystwyth, Ceredigion SY23 3FL, UK.

* Address correspondence to jmh65@cam.ac.uk.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Julian M. Hibberd (jmh65@cam.ac.uk).

J.M.H. and S.C. conceived the research plans; J.M.H. and R.L. supervised the experiments; S.C., M.S., T.E.H., R.S.-U., S.K., K.J.B., T.L.S., and J.A.P. performed the experiments. S.C., R.S.-U., S.K., T.L.S., and J.M.H. wrote the article.

[OPEN] Articles can be viewed without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.15.00889

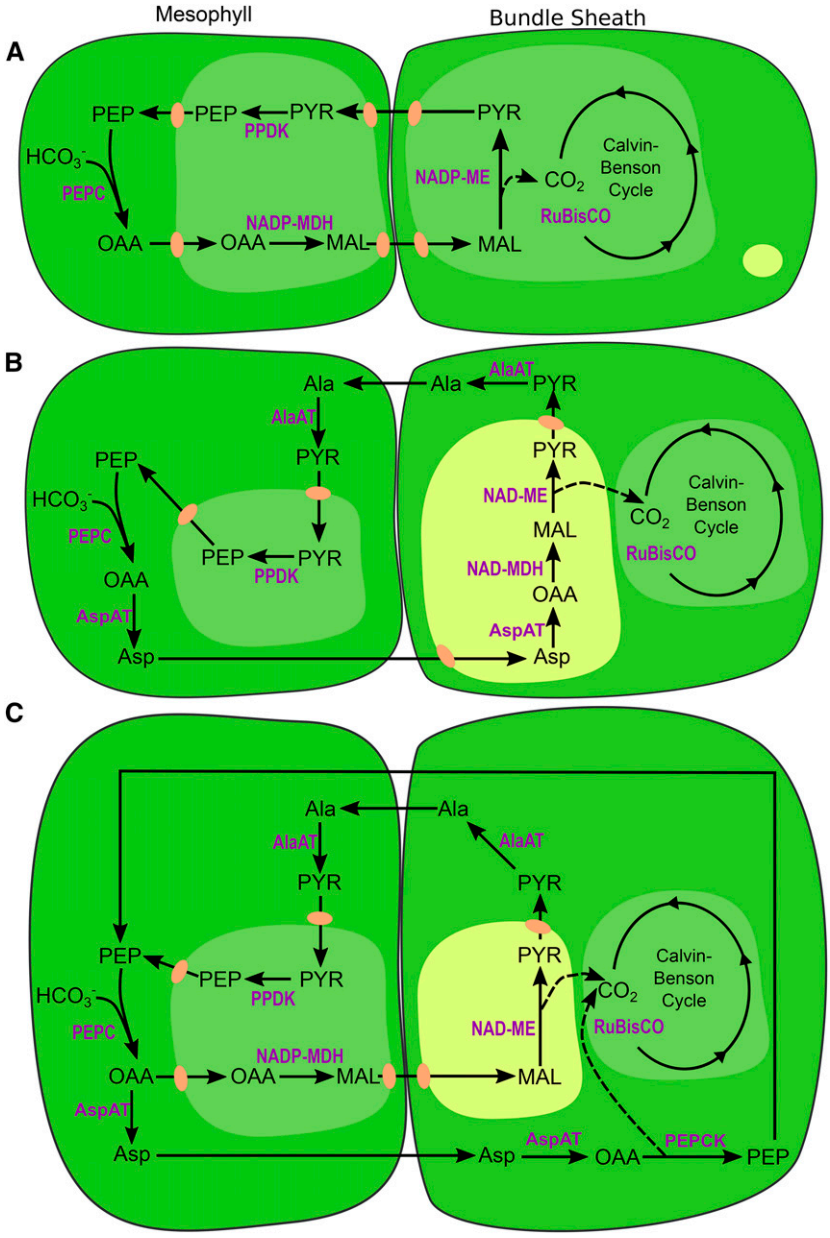


Figure 1. C_4 biochemistry has traditionally been categorized into three subtypes according to the decarboxylase enzyme used in the BS. A, NADP-ME operates in chloroplasts. B, NAD-ME operates in mitochondria. C, PEPCK is cytosolic. In all three, carbon is initially fixed by PEPC in M cells to OAA, which is then converted to MAL or Asp before transfer to the BS for decarboxylation. Released CO_2 in the BS is fixed by Rubisco in the Calvin-Benson cycle. PYR is cycled back to the M and converted to PEP. Abbreviations: Ala, alanine; AlaAT, Ala aminotransferase; Asp, aspartate; AspAT, Asp aminotransferase; MAL, malate; NAD-MDH, NAD-malate; NADP-MDH, NADP-malate dehydrogenase; OAA, oxaloacetate; PEP, phospho*enol*pyruvate; PEPC, phospho*enol*pyruvate carboxylase; PPDK, pyruvate, orthophosphate dikinase; PYR, pyruvate; Rubisco, Rubisco.

decarboxylases are common and provide selective advantage (Furbank, 2011; Bellasio and Griffiths, 2014; Wang et al., 2014a). Maize (*Zea mays*), for example, uses both NADP-ME and PEPCK to release CO_2 in BS cells (Hatch, 1971; Chapman and Hatch, 1981; Furumoto et al., 1999; Wingler et al., 1999; Furumoto et al., 2000; Majeran et al., 2010; Pick et al., 2011). Models suggest this flexibility helps maintain energy balances and allows for reduced diffusion requirements of acids between M and BS cells in a fluctuating environment (Wang et al., 2014a).

The C_4 carbon-concentrating mechanism has a wide-ranging impact on plant physiology. Due to the high affinity of PEPC for carbon, C_4 leaves maintain lower stomatal conductance than C_3 leaves, and have lower rates of water loss per unit of carbon fixed and higher water-use efficiencies (Taylor et al., 2010). Photosynthetic efficiency per unit nitrogen is also increased in C_4 species (Long, 1999) because the enriched CO_2 environment surrounding Rubisco allows for a reduction in the amount of Rubisco for the Calvin-Benson cycle. These increases in water- and nitrogen-use efficiencies are thought to contribute to the preponderance of C_4 species in relatively dry environments that can also be nutrient poor and disturbed (Osborne and Sack, 2012; Sage et al., 2012). However, the C_4 trait is broadly adaptive, with, for example, some C_4 species thriving in nutrient-rich, waterlogged conditions. In fact, the most

productive native species on the planet, *Echinochloa polystachya*, is an aquatic C_4 grass that occupies monospecific stands in the nutrient-rich Amazonian basin (Piedade et al., 1991). There is little information on photosynthesis in these highly productive aquatic C_4 species.

Other C_4 *Echinochloa* species are also highly productive in waterlogged conditions and some of the worst weeds in the world. One such species is *Echinochloa glabrescens* (barnyard grass), which is an aggressive weed of paddy rice (Fig. 2A). *Echinochloa* species have operated the C_4 pathway in waterlogged conditions for a significant period of time. Archaeological evidence documents the occurrence of *Echinochloa* species in rice paddies 8,000 years ago (Gross and Zhao, 2014), and although *E. glabrescens* is phylogenetically distant to rice (Fig. 2B), selective forces imposed over time from hand weeding are thought to have led to it becoming a rice biomimic (Barrett, 1983). *E. glabrescens* therefore provides an opportunity to investigate both how C_4 photosynthesis operates in waterlogged, nutrient-replete conditions generally and in the rice paddy specifically. It also allows for an investigation of genetic adaptation to paddy growth conditions irrespective of the photosynthetic state of a species.

In this study, we define the anatomical arrangement of M and BS cells in *E. glabrescens* and the biochemical subtype associated with release of CO_2 in the BS. In addition, two M characteristics previously described in rice were investigated to better understand whether they are general adaptations to aquatic growth or are species specific. First, we assessed the degree of M-cell-lobing and chloroplast coverage in *E. glabrescens* to determine whether biomimicry extends to cellular ultrastructure. Rice M cells are heavily lobed with chloroplasts lining the entire cell periphery (Sage and Sage, 2009). Second, we investigated whether *E. glabrescens* possesses a chloroplast-localized PEPC. Rice uses a chloroplastic PEPC to balance carbon and nitrogen metabolism in paddy conditions where most nitrogen is supplied as ammonium (Masumoto et al., 2010), but it is not known whether C_4 photosynthesis can operate in the presence of a chloroplastic PEPC. In addition to addressing these specific questions, we generated comprehensive transcriptome data for *E. glabrescens* to provide unbiased insight into its photosynthetic characteristics specifically and patterns of gene expression more generally. Comparison with deep sequencing data from C_3 rice and C_4 maize identified repeated alterations to patterns of gene expression associated with the evolution of life in waterlogged conditions or the complex C_4 phenotype.

RESULTS

Anatomical and Physiological Traits of *E. glabrescens*

To investigate Kranz anatomy and lobing in C_4 *E. glabrescens*, transverse and tangential sections of *E. glabrescens* were generated. In cross section, *E. glabrescens* leaves exhibited classical Panicoid-type C_4 anatomy with M cells surrounding BS cells in concentric circles around the vasculature (Fig. 3). M cells were minimally lobed in both cross (Fig. 3A) and tangential (Fig. 3B) sections, and no lobing was apparent on the BS. On average, M and BS cells were similar in size (Fig. 3D). However, chloroplast area in BS cells was approximately twice that found in M cells on both a cell area and percentage of cell area basis (Fig. 3D). M chloroplasts were distributed close to the cell walls and contained extensive granal stacking (Fig. 3C). BS chloroplasts were centrifugally distributed, accumulated starch, and contained mostly thylakoids with little to no granal stacking (Fig. 3C).

We next quantified photosynthetic characteristics of both species (Fig. 4). The presence of PEPC in leaves of C_4 species leads to lower carbon isotope discrimination (von Caemmerer et al., 2014) and lower compensation points (Sage, 2004). Photosynthesis in C_4 species also saturates at higher light levels than in C_3 plants (Ehleringer and Monson, 1993). Carbon isotope discrimination, photosynthetic assimilation rates under the conditions of growth (A_{net}), CO_2 compensation points, carboxylation efficiencies, maximum photosynthetic rates (A_{max}), and assimilation rate of *E. glabrescens* and rice were consistent with C_4 and C_3 photosynthesis, respectively (Fig. 4E). A_{net} and the carboxylation efficiency of C_4 *E. glabrescens* were approximately 2.7 and 13 times higher than in C_3 rice, respectively. These data, taken together with the structural traits, confirm that *E. glabrescens* was running an effective C_4 cycle. We

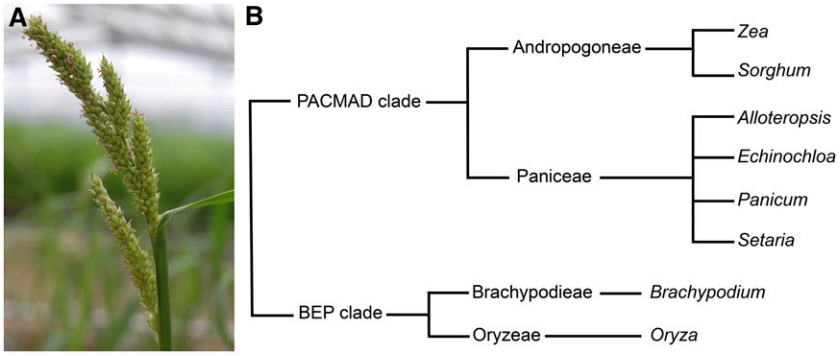
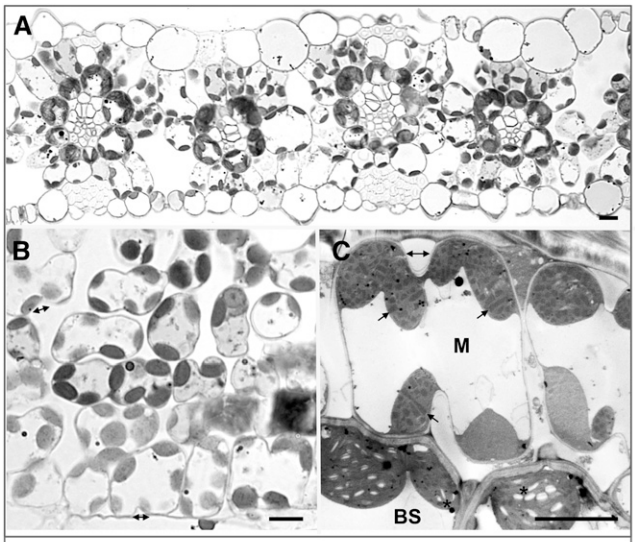


Figure 2. *Echinochloa glabrescens* is a weed of rice paddies that uses C_4 photosynthesis. A, Image of *E. glabrescens*. B, Schematic indicating phylogenetic relationships between rice (BEP clade), *E. glabrescens*, and maize (PACMAD clade).



D	M	BS
Cell size (μm²)	252 ± 27	209 ± 15
Chloroplast area per cell (μm²)	50 ± 5	107 ± 10
Chloroplast area per cell (%)	21 ± 2	49 ± 2

Figure 3. *Echinochloa glabrescens* leaves exhibit classical Panicoid Kranz anatomy and possess M cells with minimal lobing. A, Transverse section of *E. glabrescens* leaf showing Panicoid-type Kranz anatomy. Vein density was 10.36 ± 0.040 veins/mm. B, Tangential sections showing M cells are minimally lobed (double-headed arrows). C, M chloroplasts have extensive granal stacking (black arrows). BS chloroplasts show little granal stacking and contain starch accumulation (asterisk). D, Chloroplast area is greater in BS than M cells. Bars represent 10 μm.

next sought to better understand the patterns of gene expression associated with C_4 photosynthesis in *E. glabrescens* and also those genes associated with waterlogged conditions in the paddy environment.

De Novo Assembly and Annotation of the *E. glabrescens* Transcriptome

Flow cytometric analysis was used to determine that the size of the *E. glabrescens* genome is approximately 14 times larger than that of *Arabidopsis thaliana* and approaching the size of the maize genome (Table I). The *E. glabrescens* genome was approximately twice the size of other species in its genus (Zonneveld et al., 2005), some of which have previously been reported to be allohexaploids (Hilu, 1994; Ye et al., 2014). As there is neither a reference genome for *E. glabrescens* nor a genome from a closely related species that can be used for scaffolding, we used de novo assembly followed by contig annotation to define the gene complement and then quantify gene expression in this species (Aubry et al., 2014).

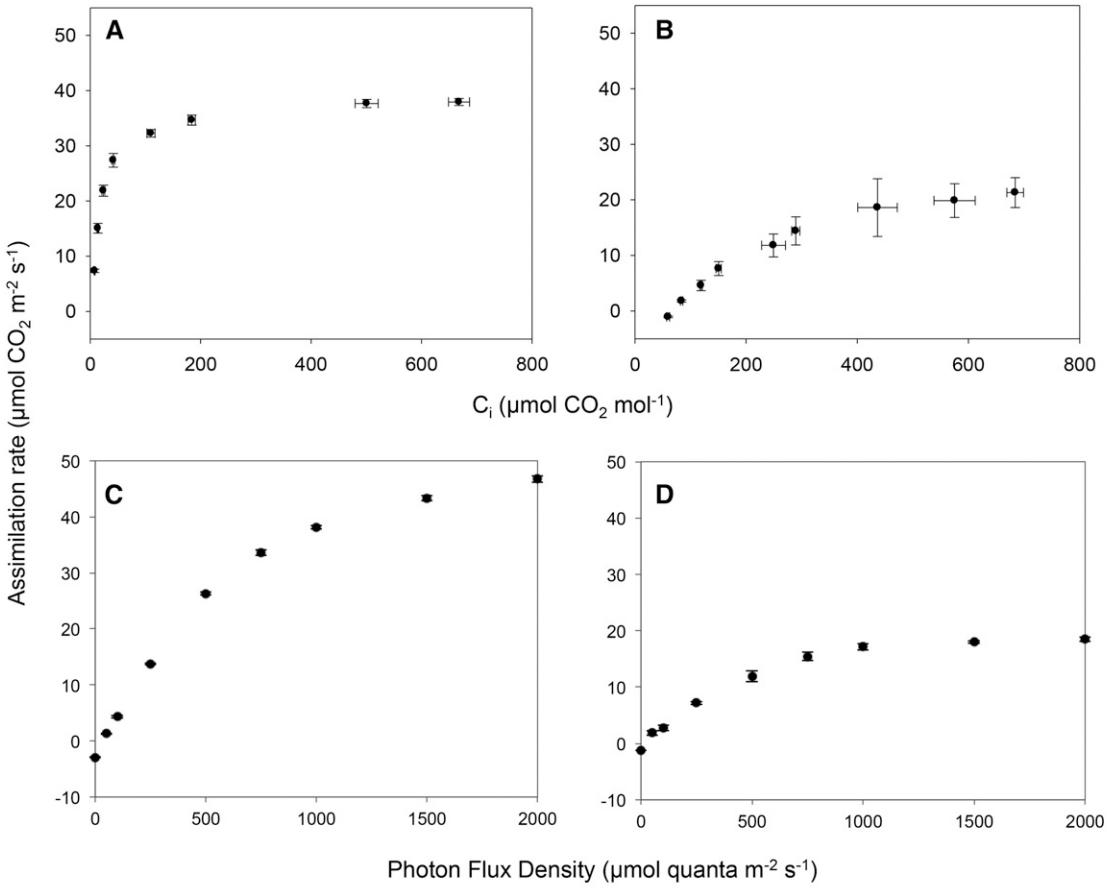
Sequencing of RNA from *E. glabrescens* and rice resulted in 74 million and 70 million reads, respectively (Table II). De novo assembly of *E. glabrescens* generated 50,226 contigs, 31,151 of which could be annotated to 15,212 orthologous genes in rice (Table II). A total of 15,147 orthologous transcript clusters had nonzero expression in at least one sample after quantification and were used in subsequent analyses. Spearman's rank correlation of transcript abundance indicated good agreement between biological replicates (Supplemental Fig. S1A).

Global Analysis of Transcript Abundance in *E. glabrescens* and Rice

To investigate major differences in transcript accumulation between rice and *E. glabrescens*, 15,147 orthologous transcript clusters were categorized into four possible statistical patterns consisting of those that showed equal expression in both species (posterior probability of differential expression [PPDE] $\leq$ 0.01), those that were more or less abundant in *E. glabrescens* relative to rice (both at PPDE $\geq$ 0.99), and those with no significant pattern (0.01 $<$ PPDE $<$ 0.99; Fig. 5A; Supplemental Table S1). Genes equally expressed in both species were included as a category to identify genes likely to be involved in fundamental processes not affected by photosynthetic type.

To obtain an overview of alterations in transcript abundance between C_3 rice and C_4 *E. glabrescens* gene ontology (GO) terms were used. Overall, 152 GO terms were statistically overrepresented across all four transcript patterns ($P < 0.05$). Transcripts that were more or less abundant in *E. glabrescens* compared with rice accounted for 74% of the overrepresented terms (Supplemental Table S2). GO terms encapsulating electron carrier activity (GO:0009055), thylakoids (GO:0009579), and oxidoreductase activity acting on NAD(P)H, quinone or similar acceptors (GO:0016655), were overrepresented in the genes up-regulated in C_4 *E. glabrescens* relative to rice. Genes more abundant in rice were associated with GO terms for ribosomes (GO:0003735, GO:0005840) and translation (GO:0008135, GO:0006412), reflecting the relatively large investment in Rubisco, Calvin-Benson cycle, and photorespiratory pathway necessary in C_3 species (Bräutigam et al., 2011). GO terms related to basic processes such as the cell-cycle (GO:0007049), histone-binding (GO:0042393), and multicellular organismal development (GO:0007275) were associated with transcripts that had equal expression between species.

To provide insight into differences in regulators of gene expression in leaves of rice and *E. glabrescens*, we investigated expression of transcription factors (TF) in each species. 1092 TFs distributed across 87 classes were detected, of which 891 were significantly differentially expressed between C_3 rice and C_4 *E. glabrescens* or equally expressed in both species and potentially involved in basic metabolism. A previous comparison of developing leaves of C_3 rice and C_4 maize led to the identification of 118 TFs proposed to play a role in regulating C_4 photosynthesis (Wang et al., 2014b), and



E

Physiological Measurement	*Echinochloa glabrescens* Value	*n*	*Oryza sativa* Value	*n*
Mean photosynthetic rate, A_{net} (µmol m^{-2} s^{-1})	32.3 ± 0.7	5	11.8 ± 2.1	3
CO_2 compensation point (µmol CO_2 mol^{-1})	2.1 ± 0.7	4	68.6 ± 2.7	3
Carboxylation efficiency (mol m^{-2} s^{-1})	1.2 ± 0.2	4	0.09 ± 0.01	3
A_{max} (µmol CO_2 m^{-2} s^{-1})	37.9 ± 0.6	5	20.6 ± 2.7	3
Light compensation point (µmol photons m^{-2} s^{-1})	37.2 ± 1.3	5	16.5 ± 3.8	3
Quantum yield (initial slope, dimensionless)	0.07 ± 0.001	5	0.03 ± 0.001	3
δ13C‰	-14.2 ± 0.2	3	-31.6 ± 0.1	3

Figure 4. *Echinochloa glabrescens* shows classical characteristics of C_4 photosynthesis. A–D, Plots of net photosynthesis *versus* internal CO_2 concentration (C_i) for (A) *E. glabrescens* and (B) *Oryza sativa* (rice) show higher maximum rates of photosynthesis and a lower compensation point in *E. glabrescens* than in rice. Plots of net photosynthesis *versus* photon flux density for (C) *E. glabrescens* and (D) rice indicate that rice saturates at lower light intensities. E, Summary of photosynthetic characteristics for each species.

91 of these TFs were differentially expressed in our dataset (Supplemental Table S3). Of these, 31 were more abundant in C_4 *E. glabrescens* than in C_3 rice, and analysis of publicly available datasets showed that 90% of these TFs had cell-enriched expression patterns in maize or *Setaria viridis* (Li et al., 2010; Chang et al., 2012; John et al., 2014; Wang et al., 2014b). 34 TFs were less abundant in C_4 *E. glabrescens* than C_3 rice, but 82% of these also exhibited cell-enriched expression in M or BS cells of *S. viridis* or maize (John et al., 2014;

Table I. *Echinochloa glabrescens has a relatively large genome size as measured by flow cytometric analysis of isolated nuclei*

Species	Relative Genome Size	1C (pg)	1C (Mbp)	2C (pg)	2C (Mbp)
Arabidopsis	—	0.16	156	0.32	313
Oryza sativa	3.1	0.50	489	1.00	978
Echinochloa glabrescens	14	2.24	2193	4.49	4386
Echinochloa crus-galli	8.4	1.35	1320	2.70	2641
Echinochloa colonum	8.4	1.35	1320	2.70	2641
Echinochloa frumentacea	8.3	1.33	1296	2.65	2592
Setaria italica	3.3	0.53	513	1.05	1027
Setaria viridis	5	0.80	782	1.60	1565
Sorghum bicolor	4.7	0.75	734	1.50	1467
Zea mays	17.1	2.73	2665	5.45	5330

Values from other species were obtained from the Plant DNA C-values Database at Kew Gardens (http://data.kew.org/cvalues/)

Wang et al., 2014b). Quantitative PCR (Q-PCR) analysis of mRNA isolated from M and BS cells of *E. glabrescens* showed that seven of the eight transcription factors assessed were preferentially expressed in either cell type (Supplemental Fig. S2), although not always in the same cell-type as has been found in *S. viridis* and maize. This finding indicates that these separate C_4 lineages up-regulate homologous transcription factors, but interestingly not always in the same cell-type.

C_4 Cycle Proteins and Related Isoforms in *E. glabrescens*

To investigate the C_4 cycle in *E. glabrescens*, the transcriptome data were mined for homologs to known C_4 proteins. As expected, transcripts related to the core C_4 cycle accumulated at higher levels in *E. glabrescens* than rice, and multiple isoforms were detected for those genes that are part of multigene families (Supplemental Table S4).

Five homologs of the rice PEPC gene family were detected in the *E. glabrescens* transcriptome. To investigate whether any of these transcripts encoded a chloroplast targeted protein, the predicted amino acid sequences were aligned against the rice chloroplastic PEPC (Os01g11054). *E. glabrescens* transcript_8661 possessed an N-terminal extension, and the predicted amino acid sequence did not contain an Ala to Ser substitution highly correlated with the C_4 enzymatic isoform of PEPC (Bläsing et al., 2000; Christin et al., 2007). Transcript_8661 was approximately 4-fold less abundant in C_4 *E. glabrescens* than C_3 rice and 286 times less abundant than the most highly expressed PEPC in *E. glabrescens* (Supplemental Table S4). A translational fusion between the full coding sequence of transcript_8661 and GFP was found to be chloroplast targeted when tested for subcellular localization *in planta* (Fig. 6). To investigate whether chloroplastic PEPC is only present in hygrophytic species (Masumoto et al., 2010; Fukayama et al., 2014), publicly available databases were searched for PEPC proteins with an N-terminal extension. Fifteen annotated sequences in addition to the rice protein were detected. ChloroP suggested that PEPC proteins from C_4 *Setaria italica* and C_3 *Brachypodium distachyon* were chloroplast targeted (Emanuelsson et al., 1999; Supplemental Table S5). Translational fusions of the first 55 residues of Si000160m and the first 48 residues of Bradi2g06620 to GFP were sufficient for chloroplast localization (Fig. 6). It therefore appears that chloroplastic PEPCs are not restricted to species adapted to waterlogged conditions and are found in multiple C_3 (rice, *B. distachyon*) and C_4 (*E. glabrescens*, *S. italica*) species (Fig. 6).

Transcripts for multiple isoforms of the C_4 decarboxylase enzymes NADP-ME and NAD-ME and a single isoform

Table II. *Illumina technology was used to sequence the leaf transcriptomes of paddy grown Echinochloa glabrescens* and *Oryza sativa (IR64)*

	Echinochloa glabrescens	*Oryza sativa*
Raw reads	74,105,468	70,089,448
Filtered reads for quantification	67,074,916	64,664,795
Processed reads for assembly	27,689,019	NA
Contigs assembled	50,226	NA
Transcripts annotated to rice genes	35,151	NA
Transcripts annotated %	50.1%	NA
Rice orthologs represented	15,212	NA
Reads mapped in quantification	54,524,488	56,228,702
Reads mapped in quantification %	81.3%	86.9%

NA, Not applicable because reads were mapped to rice reference transcriptome.

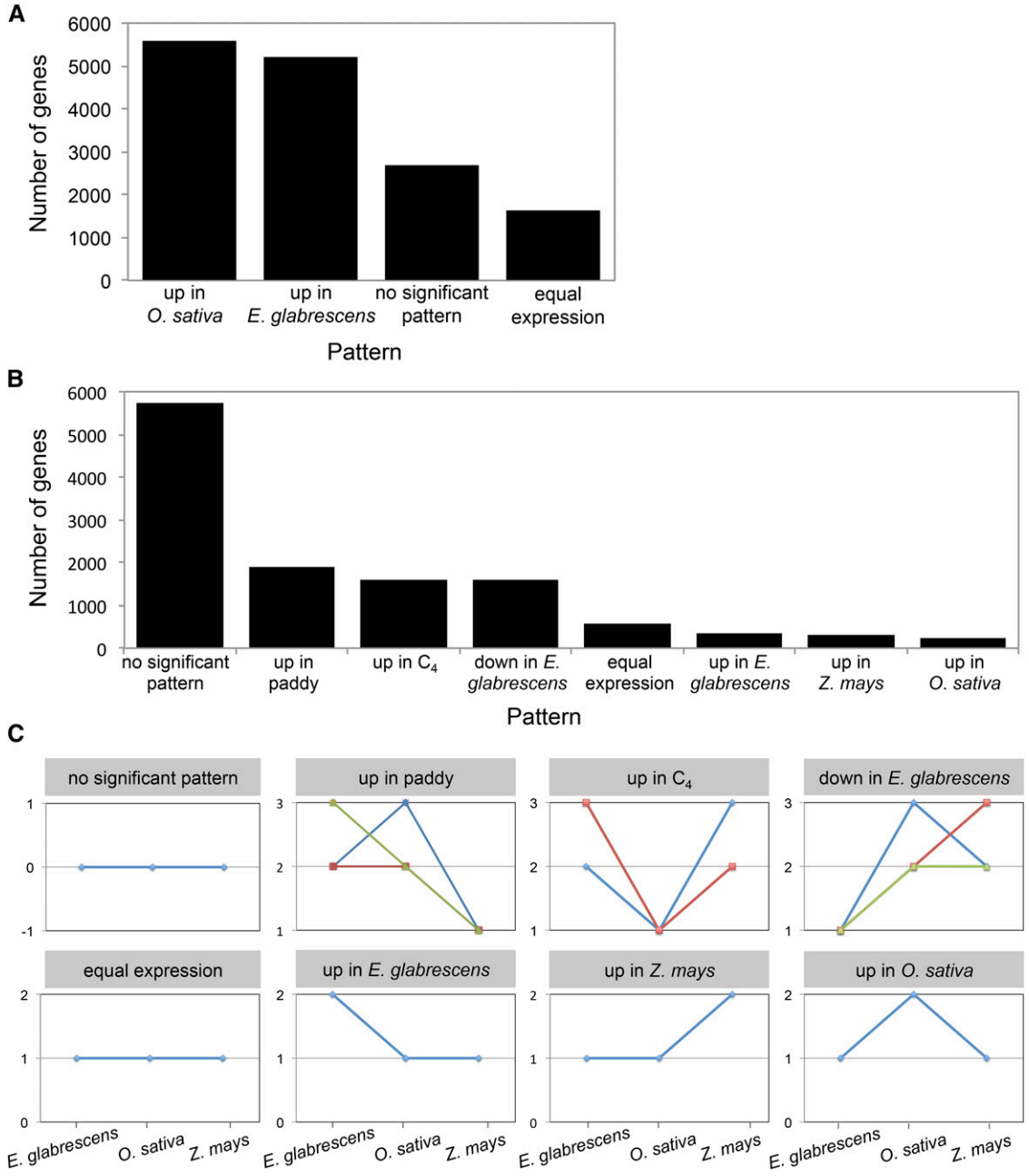


Figure 5. Differentially expressed transcripts as classified by pattern of expression. A, The distribution of detected transcripts during differential expression analysis of *Oryza sativa* (rice) and *Echinochloa glabrescens* was studied across four statistical categories for patterns of expression. B, Three-way differential expression analysis of rice (C_3), *E. glabrescens* (C_4), and *Zea mays* (C_4) was studied across eight statistical categories for patterns of expression. C, Pictorial representation of the possible expression patterns for genes statistically analyzed in the three-species comparison. Colored lines highlight the different relative expression patterns considered for each category as displayed by the squares placed along the *x* axis.

of PEPCK were also detected (Supplemental Table S4). While NAD-ME encoding transcripts were either up in rice relative to *E. glabrescens* or had no significant pattern of expression, NADP-ME and PEPCK encoding transcripts were significantly more abundant in *E. glabrescens*. In particular, *E. glabrescens* transcripts most similar to Os01g09320 (NADP-ME) and Os03g15050 (PEPCK) were 43 and 124 times more abundant in *E. glabrescens* relative to rice, respectively. Immunoblotting confirmed that the NADP-ME, PEPCK, and other C_4 cycle enzymes were abundant in *E. glabrescens* (Fig. 7). PEPC and pyruvate, orthophosphate dikinase (PPDK), which had transcripts 423 and 26 times more abundant in *E. glabrescens* than in rice,

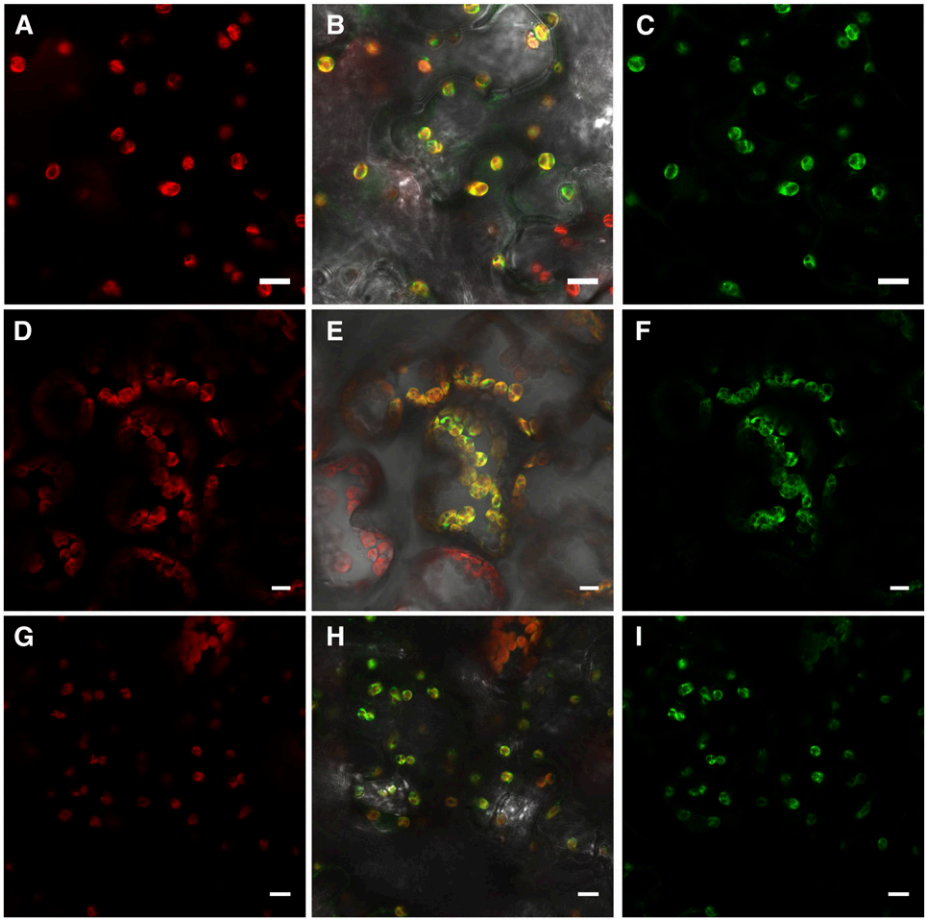


Figure 6. *Echinochloa glabrescens* contains a chloroplastic PEPC protein. *E. glabrescens* transcript_8661 was translationally fused to the GFP reporter. Subcellular localization was assessed *via* confocal scanning microscopy after infiltration into leaves of *Nicotiana benthamiana*. (A-C) Transcript_8661 fused to GFP shows chloroplast localization. (D-F) The PEPC proteins encoded by Si000160m from *Setaria italica* and (G-I) Bradi2g06620 from *Brachypodium distachyon* are also chloroplast-localized. Chlorophyll autofluorescence is seen in red (A, D, G), GFP signal in green (C, F, I), and both channels are merged onto bright field images to show cellular context (B, E, H). Bars represent 10 μm.

were also more abundant in *E. glabrescens* at the protein level. In contrast, carbonic anhydrase (CA) protein was more abundant in rice, while transcripts for two detected CA genes were up-regulated in *E. glabrescens* (Supplemental Table S4). Malate dehydrogenase (MDH) encoding transcripts were nearly 10 times more abundant in *E. glabrescens* but did not differ dramatically between species at the protein level (Supplemental Table S4; Fig. 7).

Transcripts encoding metabolite transporters associated with the C_4 pathway were more abundant in *E. glabrescens* than in C_3 rice, with the exception of *DICARBOXYLATE TRANSPORTER2* (*DCT2*), which had no significant pattern of expression (Supplemental Table S4). The transporters analyzed are known to be more abundant in maize relative to C_3 *Pisum sativum* (Bräutigam et al., 2008) and enriched in M or BS cells of maize (Taniguchi et al., 2004; Majeran et al., 2005; Majeran et al., 2008; Majeran et al., 2010). Consistent with the lower amounts of Rubisco in C_4 leaves and reduced rates of photorespiration, *RBCS* transcripts (encoding the small subunit of Rubisco) as well as transcripts related to photorespiration were reduced in *E. glabrescens* compared with rice (Supplemental Table S4). Overall, we conclude from the transcript and protein data that the aquatic C_4 species *E. glabrescens* is operating a C_4 pathway that relies on both NADP-ME and PEPCK to provide CO_2 to Rubisco in the BS. The data also indicate that a chloroplastic PEPC in C_4 leaves is compatible with operation of the C_4 cycle, although in C_4 *E. glabrescens* transcripts encoding the chloroplastic PEPC are down-regulated relative to C_3 rice (Supplemental Table S4).

Genes Associated with Paddy Conditions and C_4 Photosynthesis

To provide broader insight into genes associated with either the paddy environment or C_4 photosynthesis, we next included publicly available data for the C_4 grass maize grown in nonpaddy conditions in our analysis (Wang et al., 2013a). These data allowed a three-way comparison consisting of two species that evolved the C_4 pathway independently (*E. glabrescens* and maize; Giussani et al., 2001; Christin et al., 2008; Sage et al., 2011), and also two species that independently adapted to life in waterlogged paddy conditions (*E. glabrescens* and rice). Transcript abundance was highly correlated between biological replicates of each species (Supplemental Fig. S1B), but maize and *E. glabrescens*, both members of the PACMAD clade (Fig. 2), showed higher correlations with each other than with rice.

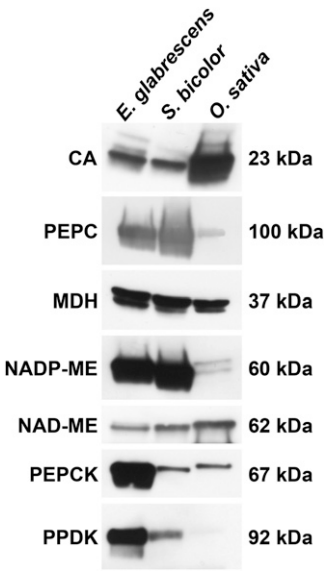


Figure 7. *Echinochloa glabrescens* accumulates significant levels of known C_4 cycle proteins in leaves, including both NADP-ME and PEPCK. Samples were loaded on a fresh weight basis from 1 mg leaf tissue. Estimated molecular weights are indicated to the right of each blot. Abbreviations: CA, Carbonic Anhydrase; PEPC, phospho*enol*pyruvate carboxylase; MDH, malate dehydrogenase; PPDK, pyruvate, orthophosphate dikinase.

In total, 12,256 rice loci were detected in all three species and analyzed for statistical differences in transcript accumulation according to eight possible patterns of expression (Fig. 5, B and C; Supplemental Table S6): up in both C_4 species relative to C_3 rice, up in rice relative to the C_4 species (equivalent to down in C_4), up in both paddy species relative to maize, up in maize relative to the paddy species (equivalent to down in paddy), up and down in *E. glabrescens* relative to both rice and maize (representing this species' unique expression profile), equal expression between species (representing genes involved in basic metabolism), and no significant pattern between species. While nearly half the genes had no significant pattern of expression (Fig. 5B), putative paddy- or C_4-related genes constituted 18% and 15% of the detected rice loci, respectively, and 16% of the genes corresponded to an expression profile that was unique to *E. glabrescens*. Approximately 5% of genes were equally expressed between species and likely represent basic metabolism.

E. glabrescens and maize had similar, but not identical, molecular signatures for the core C_4 pathway, metabolite transport, and photorespiratory process (Supplemental Tables S4 and S6). For the core C_4 pathway genes, with the exception of CA, the gene family member having the most abundant expression in *E. glabrescens* relative to rice in the two-species analysis was also up-regulated in both C_4 species in the three-species analysis. Similarly, at least one isoform of the transporters *PHOSPHO*ENOL*PYRUVATE/PHOSPHATE TRANSLO*CATOR (*PPT*), *DI*CA*RBOXYLATE TRANSPORTER1* (*DCT1*), *METHYL ERYTHRITOL PHOSPHATE2* (*MEP2*), and *TRIOSE PHOSPHATE/PHOSPHATE ANTIPORTER* (*TPT*) was up-regulated in both C_4 species. However, the CA and *MEP3b* homologs most abundant in *E. glabrescens* relative to rice were up-regulated in the paddy species relative to maize, *DCT2* was down-regulated in the paddy species relative to maize, and no significant pattern of expression was found for *2-OXOGLUTARATE/MALATE TRANSPORTER* (*OMT1*) and *MEP3a*. Although the photorespiratory genes were not statistically classified as up in rice relative to the C_4 species (i.e. down in C_4), transcript accumulation was typically at least an order of magnitude higher for one isoform per gene in rice relative to the C_4 species (Supplemental Tables S4 and S6), consistent with a more active photorespiratory pathway in C_3 rice.

Fifty-five GO terms were statistically overrepresented across the eight expression categories, of which 17 and 12 were associated with genes up-regulated in paddy or C_4 species, respectively (Supplemental Table S7). Terms related to the nonoxidative branch of the pentose phosphate pathway and/or the Calvin-Benson cycle were associated with genes up-regulated in the paddy (GO:0009052, GO:0004751), possibly indicating a stronger repression of these pathways in maize with respect to rice and *E. glabrescens*. With respect to C_4, two GO terms annotated as ATP proton transport (GO:0015991, GO:0046961) were associated with up-regulated genes. This is consistent with the higher demand for ATP generated by the C_4 cycle relative to C_3 biochemistry. The Glu biosynthetic process (GO:0006537) was also associated with the C_4 species, providing further support for an operational PEPCK C_4 cycle in *E. glabrescens*.

TFs were analyzed to gain insight into regulatory networks that might be operating on the C_4- or paddy-related gene sets. 967 TFs belonging to 86 families were identified, with 117 and 132 TFs more abundant in the C_4 and paddy species, respectively. To determine whether the C_4- or paddy-associated TFs might be coregulated, a motif search of the first 1000 bp of the rice homologs for each set of TFs using ELEMENTS was performed (Mockler et al., 2007). Seven and 16 motif clusters, up in C_4 and up in paddy respectively, were found ($P < 0.05$; Supplemental Table S8). The TFs identified in the three-species analysis were also compared with the 118 TFs previously proposed to be associated with C_4 photosynthesis (Wang et al., 2014b). The same 91 TFs present in the rice vs. *E. glabrescens* analysis were distributed across five of the patterns of expression present in the three-species analysis (Supplemental Table S9), but only 12 TFs were classified as up in both C_4 species. These 12 TFs were also up-regulated in *E. glabrescens* in the two-species analysis, suggesting they may play a strong role in C_4 regulation. Indeed, the rice homolog of *SCARECROW*, a BS-enriched gene proposed to be involved Kranz anatomy development (Slewinski et al., 2012; Slewinski, 2013;

Wang et al., 2013a), was part of this list. It was also notable that 9 of 12 TFs up in the C_4 species have BS-enriched expression in the tip of maize leaves (Li et al., 2010), and when a motif search was performed using ELEMENT, an exact match for a candidate BS cell cis-element from Wang et al. (2014b) was found (Supplemental Tables S10 and S11), suggesting that a egulatory network present in C_3 rice continues to function in maize and *E. glabrescens* but in a cell-enriched manner. It is also notable that 14 of the TFs previously identified as having a role in C_4 photosynthesis (Wang et al., 2014b) were up in both paddy species and that they were roughly split between BS- and M-enriched expression in maize (Li et al., 2010). Motif analysis using ELEMENT (Supplemental Table S10) yielded three candidate cis-elements that were present in both the Wang et al. (2014b) dataset and the TFs up-regulated in the paddy species (Supplemental Table S11).

The GLK TF family is required in plants for chloroplast development and photosynthetic gene expression (Hall et al., 1998; Rossini et al., 2001; Fitter et al., 2002; Waters et al., 2009; Powell et al., 2012; Wang et al., 2013b; Nguyen et al., 2014) and also plays a role in regulating nitrogen metabolism in Arabidopsis by inducing Gln synthetase expression and repressing Glu dehydrogenase (Gutiérrez et al., 2008). *GLK*s are usually present as a duplicated gene pair whose members are subfunctionalized in C_4 species between M and BS cells but operate redundantly in all photosynthetic cells of C_3 plants (Wang et al., 2013b). A *GLK* gene pair previously identified as *OsGLK1* (Os06g24070) and *OsGLK2* (Os01g13740; Wang et al., 2013b) was up in both paddy species in the three-species analysis and up in rice relative to *E. glabrescens* in the two-species analysis. A second related gene pair, Os04g56990 and Os05g40960, was up-regulated in *E. glabrescens* relative to rice in the two-species analysis, with Os04g56990 up in both C_4 species and Os05g40960 having no significant pattern in the three-species analysis. These two pairs of genes related to *GLK* may reflect subfunctionalization needed to perform the dual roles of chloroplast development and nitrogen assimilation in both paddy (Pair 1: Os06g24070, Os01g13740) and C_4 (Pair 2: Os04g56990, Os05g40960) contexts. Homologs of Os06g24070 and Os04g56990 are M-enriched and Os01g13740 and Os05g40960 are BS-enriched in maize (Li et al., 2010).

Transcripts Up-Regulated in Paddy Conditions Are Evolving More Rapidly

The analysis above identified 1610 transcripts putatively important to C_4 function in both dry and wet conditions and also 1905 genes with a potential role in paddy life. To investigate whether these groups of genes have been evolving at different rates compared with the general population of genes, we assessed the number of nonsynonymous substitutions per site (Fig. 8). Interestingly, genes up-regulated in the paddy species have been evolving approximately 10% faster than those that are either down-regulated or nondifferentially expressed (P = 0.004; Fig. 8). This suggests that genes up-regulated in the two paddy species may have evolved to adapt to waterlogged conditions. An analogous effect is not seen for genes that are up- or down-regulated in C_4 species compared to C_3. However, we note that genes that are up-regulated in C_4 species may be evolving faster (P = 0.018) than those that are not significantly differently expressed under any other condition.

DISCUSSION

Photosynthetic Characteristics of *E. glabrescens*

M cells of *E. glabrescens* were similar to arid C_4 species, having minimal lobing as well as low chloroplast coverage of the cell area and periphery (Fig. 3; Panteris and Galatis, 2005; Stata et al., 2014). As such, although *E. glabrescens* seedlings are highly effective at mimicking rice leaf architecture at a gross level, M cell ultrastructure was similar to C_4 maize (Panteris and Galatis, 2005) rather than rice. In addition, M chloroplasts of *E. glabrescens* were distributed along the cell periphery and, as with other C_4 species, are partially recessed from the cell wall to create a cytoplasmic channel where PEPC can fix carbon (Stata et al., 2014). This is unlike

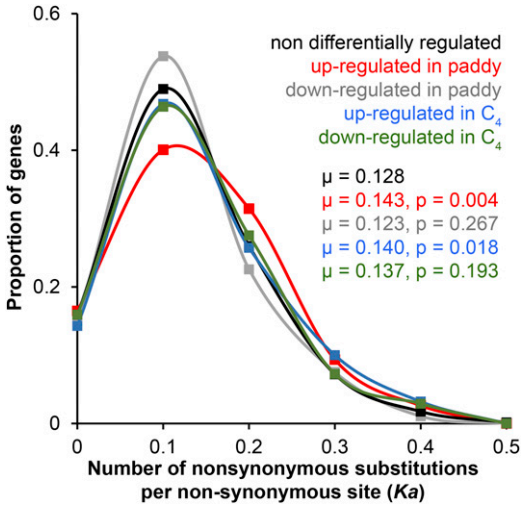


Figure 8. Genes up-regulated in paddy grown rice and *Echinochloa glabrescens* show faster rates of evolution than homologs in maize, while those up-regulated in the C_4 species (*E. glabrescens* and maize) relative to rice do not. The distribution of evolutionary rates as calculated by the number of amino acid substitutions per site per gene at the amino acid level are shown for genes up-regulated in paddy (red), down-regulated in paddy (gray), up-regulated in C_4 (blue), and down-regulated in C_4 (green). Non differentially regulated genes are shown in black and include all those genes from the three-species analysis that are not present in the four categories under assessment.

both domesticated and wild rice species, which exhibit extensive M-cell-lobing with chloroplasts and stromules covering over 97% of the cell periphery (Sage and Sage, 2009). It also contrasts with C_3 and C_2 (C_3-C_4 intermediate) species, which have more M chloroplasts than C_4 species and a high degree of appression against the cell wall (Stata et al., 2014). Placement of M cell chloroplasts against the cell wall is thought to act as a partial photorespiratory CO_2-scavenging mechanism by forcing CO_2 to pass through chloroplasts when diffusing both into and out of the cell (Sage and Sage, 2009; Stata et al., 2014). Such an arrangement would be unnecessary in C_4 M cells since PEPC acts as carboxylase in the cytosol and Rubisco operates in the BS (Stata et al., 2014). Physiological characteristics were all in the C_4 range, and gas exchange analysis of *E. glabrescens* showed high rates of photosynthesis relative to rice (Fig. 4). Taken together, the anatomical and physiological features of *E. glabrescens* were those of a classical Panicoid C_4 species similar to other arid C_4 grass species.

Chloroplast-Localized PEPC

Our data suggest a chloroplastic PEPC that provides carbon skeleton for ammonium assimilation is important to paddy life for both C_3 and C_4 species. Chloroplastic PEPC is present in two monocot C_3 hygrophytic species, rice (Masumoto et al., 2010) and *Monochoria vaginalis* (water hyacinth; Fukayama et al., 2014), as well as C_4 *E. glabrescens* (Fig. 6). In these paddy species, chloroplastic PEPC is likely an adaptation to the reductive environment of the waterlogged growth habitat where all inorganic nitrogen is available as NH_4^+ (Masumoto et al., 2010). However, the presence of a chloroplast-localized PEPC in both the arid-growing C_3 weed *B. distachyon* and the cultivated C_4 *S. italica* suggests that either dry environments also lead to the evolution of a chloroplastic PEPC functioning in ammonium metabolism, or that this enzyme has undiscovered roles. The rapidly growing list of species with published transcriptomes can be used to explore the presence and purpose of chloroplast-targeted PEPC across a broad phylogenetic context.

It was previously uncertain whether chloroplastic PEPC in rice would complicate engineering efforts to introduce C_4 photosynthesis into rice in order to increase yields (Kajala et al., 2011). Although transcripts encoding chloroplastic PEPC were less abundant in *E. glabrescens* than in rice (Supplemental Table S4), chloroplastic PEPC localization in multiple highly successful C_4 species (*S. italica* and *E. glabrescens*; Fig. 6) suggests its activity is compatible with the C_4 cycle. This is likely aided by spatial restriction of most PEPC accumulation to M cells where both cytosolic and chloroplastic PEPCs would positively contribute to C_4 shuttling (Fig. 9). In addition, any PEPC activity in C_4 BS cells, if expression occurs in these cells, would likely be limited to non-C_4 related pathways because PPDK has M-enriched accumulation in C_4 species. Of further benefit to engineering efforts, rice chloroplastic PEPC appears to behave very differently from the cytosolic isoform (Fukayama et al., 2006; Masumoto et al., 2010; Fukayama et al., 2014; Muramatsu et al., 2015). Relative to cytosolic PEPC, the rice chloroplastic PEPC has very low affinity for PEP; low sensitivity to the inhibitors Asp, Glu, and malate; and it is activated by Glc-1-phosphate, Fru-1,6-bisphosphate, 3-phosphoglyceric acid, and 2- phosphoglyceric acid, but not by Glc-6-phosphate or alkaline pH (Muramatsu et al., 2015). Cytosolic and chloroplastic PEPCs are also nocturnally and likely diurnally regulated respectively (Fukayama et al., 2006; Fukayama et al., 2014). As such, it is possible that when cytosolic and chloroplastic PEPCs are both present in C_4 species, they are active at different times of day and have different enzymatic properties, allowing for balancing of C_4 and nitrogen metabolism. Activity assays and temporal expression analyses of the chloroplast targeted PEPC gene in the C_4 species *E. glabrescens* and *S. italica* are required to better understand how the operation of these two PEPCs is integrated.

Balancing Carbon and Ammonium Metabolism in a C_4 Paddy Species

A mixed NADP-ME/PEPCK C_4 cycle may play an important role in linking carbon and nitrogen metabolism in *E. glabrescens* and in maintaining adequate C_4 flux in a paddy environment (Fig. 9). Presence of a mixed NADP-ME/PEPCK cycle is supported by centrifugally positioned BS cell chloroplasts with reduced granal stacking (Gutierrez et al., 1974; Nelson and Dengler, 1992; Fig. 3), high protein abundance of NADP-ME and PEPCK (Fig. 7), and high transcript levels of NADP-ME, PEPCK, *ASP AMINOTRANSFERASE* (*Asp-AT*) and *ALA AMINOTRANSFERASE* (*Ala-AT*) relative to both maize and rice (Supplemental Table S6). In addition, *E. glabrescens* possesses a highly abundant chloroplastic CA (Os01g45274) that could provide bicarbonate to the chloroplastic PEPC. Transcripts encoding this CA isoform was approximately 9-fold and 2-fold more abundant than in maize and rice, respectively (Supplemental Table S6), and the corresponding *E. glabrescens* transcript_4427 only differed by one residue to the peptide sequence of Os01g45274, which is known to be chloroplast-localized (Suzuki and Burnell, 1995).

A mixed NADP-ME/PEPCK cycle would be able to provide PEP to carbon and nitrogen metabolism through a number of routes (Fig. 9). Both cytosolic and chloroplastic PEPCs in the M could be directly provided with PEP by PEPCK activity in BS cells. In support of this notion, PEPCK (Os03g15050) transcripts are approximately 4-fold and 129-fold more abundant in *E. glabrescens* relative to maize and rice, respectively (Supplemental Table S6). In addition, PEP could be generated from pyruvate in M cell chloroplasts via PPDK operating in the NADP-ME C_4 cycle. Pyruvate could be transferred from BS to M cells either directly or

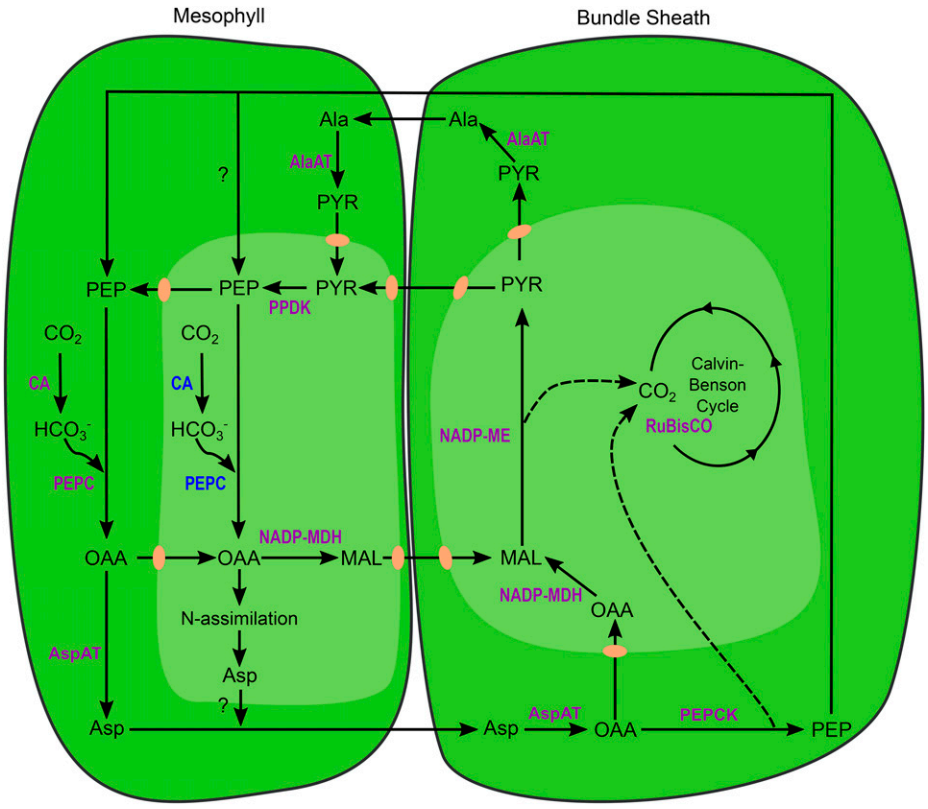


Figure 9. A model of the interaction of the mixed NADP-ME/PEPCK C_4 photosynthetic cycle and specialized ammonium metabolism operating in *Echinochloa glabrescens*. Cytosolic and chloroplastic PEPC isoforms are present in M cells, where they convert PEP to OAA. Cytosolic and chloroplastic CAs could provide both PEPCs with the necessary bicarbonate substrate in situ. The cytosolic pool of M cell OAA could be converted to Asp, move to the BS, be converted back to OAA, and then used as substrate by PEPCK to generate PEP and CO_2. The PEP would travel to the M cell to complete C_4 cycles for possibly both the cytosolic and chloroplastic PEPCs, while Rubisco fixes CO_2 in the BS chloroplast. The BS cytosolic pool of OAA could also be directly imported into the BS chloroplast, converted to MAL, and decarboxylated by NADP-ME. Resulting CO_2 would be used by Rubisco as part of the Calvin-Benson cycle, and PYR could either directly travel to the M cell chloroplast or cycle through Ala en route before being converted to PEP. The M cell chloroplastic OAA pool could be part of both C_4 and ammonium metabolism. As part of the C_4 cycle, OAA would be converted to MAL, move to the BS chloroplast, and become decarboxylated by NADP-ME. Alternatively, M chloroplast OAA could remain in the M chloroplast, enter N-metabolism to achieve ammonium assimilation, and/or generate Asp which could also be cycled into C_4 PEPCK metabolism. Abbreviations: Ala, Ala; AlaAT, Ala aminotransferase; Asp, Asp; AspAT, Asp aminotransferase; CA, carbonic anhydrase; MAL, malate; NADP-MDH, NADP-malic dehydrogenase; OAA, oxaloacetate; PEP, phospho*enol*pyruvate; PEPC, phospho*enol*pyruvate carboxylase; PPDK, pyruvate, orthophosphate dikinase; PYR, pyruvate; Rubisco, rubisco.

via Ala. *E. glabrescens* transcripts most similar to NADP-ME genes Os01g09320 and Os05g09440 are approximately 4-fold and 8-fold more abundant than in maize and 45-fold and 4-fold more abundant than in rice (Supplemental Table S6), and transcripts encoding ADENYLATE KINASE (AK) (Os08g01770), which drives the PPDK reaction toward PEP (Burnell, 2011), accumulate approximately eight times and 11 times higher in *E. glabrescens* relative to maize and rice (Supplemental Table S6). Since maize also operates a mixed NADP-ME/PEPCK cycle (Hatch, 1971; Chapman and Hatch, 1981; Furumoto et al., 1999; Wingler et al., 1999; Furumoto et al., 2000; Majeran et al., 2010; Pick et al., 2011), the abundant accumulation of NADP-ME and *AK* transcripts in *E. glabrescens* suggests that increased pyruvate and PEP in M cells is both possible and necessary for the operation of C_4 and ammonium metabolism.

To support sufficient levels of metabolite flux for both C_4 and ammonium metabolism, abundant expression of transporters might be expected in *E. glabrescens*. Indeed, transcripts for PPT (PEP transporter), TPT (3-phosphoglycerate/triosephosphate exchanger), DCT1 (Glu/malate exchanger), MEP2 (putative OAA/malate shuttle), and OMT1 (2-oxoglutarate/malate exchanger) are all more abundant in *E. glabrescens* relative to maize and rice. Of particular note, MEP3b, a putative pyruvate transporter, is 40 times more abundant in *E. glabrescens* than in maize. Although the Arabidopsis homolog of MEP3b acts as a glycolate glycerate transporter (Pick et al., 2013), the role of MEP proteins in maize has, to our knowledge, not yet been defined. Overall, the data imply that *E. glabrescens* likely has sufficient flexibility to provide flux to both C_4 and ammonium metabolism as needed and so is able to maintain high photosynthetic rates (Fig. 4).

E. glabrescens Transcriptome Yields C_4- and Paddy-Associated Genes

Differential expression analysis of rice (paddy C_3), *E. glabrescens* (paddy C_4), and maize (arid C_4) identified 1610 genes, including 117 TFs putatively important to C_4 function. *E. glabrescens* and maize are descended from two independent lineages of C_4 evolution (Giussani et al., 2001; Christin et al., 2008), but both are members of the same PACMAD clade (Fig. 2). As a result, it is

presently unclear whether the up-regulation of those genes in the two C_4 species relative to rice is due to independent recruitment to the C_4 pathway or phylogenetic signature. In either case, it is interesting to note that 91 of the TFs have previously been characterized as putative C_4 regulators (Wang et al., 2014b), but only 12 of those TFs were up-regulated in both *E. glabrescens* and maize (Supplemental Table S9). These 12 TFs form key targets for C_4 functional analyses since they exhibit similar behavioral patterns across at least two C_4 species. In addition, *E. glabrescens* and maize had unique molecular signatures with 3% and 2% of detected genes respectively up-regulated solely in one species. Some of these genes may be specifically required for either wet or dry C_4 photosynthesis, and future comparative studies should consider further wet C_4 species to address this. Interestingly, the majority of core C_4 genes and transporters appear to have similar expression profiles, with the exception of genes playing a dual role in ammonium metabolism, suggesting genes associated with recruitment to C_4 photosynthesis can dynamically interface across multiple pathways.

A putatively paddy-related signature of 1905 genes was also identified in the three-species analysis. This gene suite appears to be evolving more rapidly relative to genes that were nondifferentially expressed or down-regulated in paddy (Fig. 8) and so can be used to identify essential paddy genes by comparative expression analysis with rice varieties adapted to dry upland growth conditions and/or other paddy weeds. Major alterations in transcript abundance associated with life in paddy conditions could be mediated by convergent alterations to elements in cis in each genome, changes to transfactors, or a combination of both. Without *E. glabrescens* genome information, it is not possible to comprehensively investigate alterations in cis, but the 132 TFs identified as under selection constitute a manageable starting point for characterizing genes of possible benefit to rice-breeding programs. As rice agriculture must rapidly adjust to changes in climate that affect paddy conditions, e.g. using reduced water levels, these TFs and other genes identified as being under selection in the paddy species could be mined for diversity that would help improve rice stocks.

Conclusion

In summary, the analysis presented provides, to our knowledge, the first insight into leaf transcript populations of a species performing C_4 photosynthesis in wet paddy conditions. The data indicate that *E. glabrescens* runs a mixed C_4 biochemical pathway dependent on both chloroplastic NADP-ME and cytosolic PEPCK decarboxylases. We propose that a chloroplastic PEPC combined with abundant PEPCK allows for efficient metabolic flux through both C_4 and ammonium metabolism in *E. glabrescens*. In addition, comparative expression analysis between the three species *E. glabrescens*, maize, and rice has laid the foundation for identifying a molecular signature for paddy life.

MATERIALS AND METHODS

Anatomy

Echinochloa glabrescens was grown from seed in a University of Toronto greenhouse in 20 L pots of sandy-loam. The approximate day/night temperature of the greenhouse was 27/26°C, and peak midday irradiance exceeded 1600 μmol photons m^{-2} s^{-1}. Fertilizer was supplied as a 50:50 mixture of Miracle-Gro 24-10-10 All Purpose Plant Food and Miracle Gro Evergreen Food (30-110-20) at the recommended dosage (22 mL of fertilizer salt per 6 L; Scotts Micracle-Gro). Plants were sampled between 10:00 and 11:00 AM. Internal leaf anatomy was assessed on sections harvested from the middle of the leaf (1 leaf per plant; 3 plants) and prepared for light and transmission electron microscopy as described previously (Sage and Williams, 1995). Quantification of transmission electron microscopy (TEM) images was prepared from three leaf sections from three separate plants. TEM images were analyzed using ImageJ software (National Institutes of Health, Bethesda, MD, USA) to quantify cell area (n = 11), chloroplast area per cell (M n = 11; BS n = 10), and the percent chloroplast area per cell (M n = 11; BS n = 10). Cross sections of cleared *E. glabrescens* leaf tissue from plants grown at the University of Cambridge (see Physiology below, same seed stock) were used to calculate vein density. Tissue was cleared of chlorophyll by immersion in 70% ethanol at 65°C for 24 h followed by 5% w/v NaOH at room temperature for 24 h. Samples were washed three times in distilled water before examination. Images were taken on a light microscope using Leica Application Suite (Leica, Germany) at 4× magnification and analyzed using ImageJ (National Institutes of Health, Bethesda, MD, USA).

Physiology

Seeds of *E. glabrescens* and *Oryza sativa* L. ssp. *indica* cv IR64 were germinated at the University of Cambridge in a growth chamber at 28°C, 50% humidity, 400 ppm CO_2, and 400 μmol photons m^{-2} s^{-1} light in a 16 h photoperiod. Seedlings were transferred from trays to pots filled with compost and medium vermiculite (2:1) and watered to saturation. After two weeks, plants were transferred to a greenhouse with a temperature of 26°C and 800 μmol m^{-2} s^{-1} light in a 16 h photoperiod. Plants were allowed to acclimate for one week before physiological measurements were taken. For $\delta^{13}C‰$ measurements, leaf samples of *E. glabrescens* and rice were prepared using a modified procedure based on Smith and Brown (1973). Leaf samples were dried at 70°C, ground by a ball mill for 6 min (frequency 30/s), and then 0.5 mg was weighed into tin capsules. Caffeine and USGS40 standards were used to compare the $^{13}C/^{12}C$ ratios. Samples were run using a Costech Elemental Analyzer attached to a Thermo MAT253 mass spectrometer in continuous flow mode, and values were calculated using the formula $\delta^{13}C‰ = \left[\frac{Rsample}{Rstandard} - 1\right] x1000$. Photosynthesis measurements ($n > 3$) were generated using an infrared gas analyzer (Li-6400XT, LICOR, USA) with a 2 cm^2 leaf chamber fluorometer head (6400-40, LICOR, USA). Plants were well watered before measurements, and the first reading was taken after conductance and humidity stabilized and VpdL dropped to near 1 kPa. Mean photosynthetic rate was measured at 400 ppm CO_2, 1000 μmol photons m^{-2} s^{-1}, humidity > 65%, and leaf temperature 27°C. For A/C_i curve generation, chamber conditions were 1000 μmol photons m^{-2} s^{-1}, 27°C and relative humidity > 60%. A CO_2 mixer was used to control chamber CO_2 concentrations through a series of measurements at 400, 50, 100, 150, 200, 400, 800, and 1000 ppm CO_2. Measurements were taken at steady state, normally achieved after 4 min. Light response curves were constructed at 27°C at relative humidity > 60% in a saturating chamber with CO_2 of 800 ppm. Leaves were allowed to acclimate at 2000 μmol photons m^{-2} s^{-1} before being measured at 2000, 1500, 1000, 750, 500, 250, 100, 50, and 0 μmol photons m^{-2} s^{-1}. For each measurement, the leaf was allowed to equilibrate, normally taking 4 min.

Genome Size

E. glabrescens was grown as described for the physiological measurements. Nuclear DNA content was determined by flow cytometric analysis of isolated nuclei from 50 mg of leaf material. The protocol was modified from Arumuganathan et al. (1991) and performed at the Benaroya Research Institute (www.benaroyaresearch.org) as a pay-per-use service. Chicken erythrocyte nuclei (2.5 pg/2C) were used as a standard. Suspensions of sample nuclei were spiked with suspension of standard nuclei and analyzed with a FACScalibur flow cytometer (Becton-Dickinson, USA). The mean positions of the G0/G1 (nuclei) peak of the sample and internal standard were determined by CellQuest

software (Becton-Dickinson, USA). The mean nuclear DNA content of each plant sample, measured in pg, was based on 1000 scanned nuclei. Genome size was calculated using 1 pg = 978 Mbp as a conversion coefficient. Values for genome size in pg and Mbp for other species in Table I are from the RGB Kew DNA C-values database (data.kew.org/cvalues).

RNA Extraction and Sequencing

E. glabrescens and *O. sativa* cv IR64 were germinated and grown simultaneously in pots in paddy conditions in a screenhouse at the International Rice Research Institute in the Philippines between December 2010 and January 2011. Plants were grown in local soil. Three biological replicates were created for RNAseq of each species. Each biological replicate was pooled from sampling the youngest mature leaf from the main tiller of two individual plants at the vegetative stage, midribs were removed, and each was harvested at similar time points. For *E. glabrescens*, tissue was collected at 10:30, eight days after transplanting from the third leaves of the main tiller of fourth leaf emerging plants. For rice, tissue was sampled at 11:00 AM on the same day from the newest fully expanded leaves of the main tiller. RNA was extracted from each biological replicate using 3 mL of TRIzol (Life Technologies, USA) and the standard protocol. After extraction, RNA was purified and DNase-treated using an RNeasy column (Qiagen, Netherlands) and checked for quality using a Bioanalyzer (Agilent Technologies, USA). Illumina RNAseq libraries were prepared by the Beijing Genomics Institute (www.genomics.cn) and sequenced using HiSeq2000 Illumina 100 bp paired end technology (Illumina, USA). 4G clean data were prepared per sample.

For Q-PCR on mRNAs from M and BS cells, RNA was isolated using leaf rolling and mechanical isolation, respectively (Covshoff et al., 2012). cDNA synthesis was performed using the Invitrogen SuperScript II Reverse Transcriptase kit with 300 ng of RNA. Q-PCR reactions were performed on Rotor-Gene Q real-time PCR cycler (Qiagen) with an annealing temperature of 58°C (30 s) and extension temperature of 72°C (30 s) using Sigma SYBR Green-JumpStart. Cycle threshold (Ct) values were then calculated using actin as a control. Data are presented as means of three replicates.

De Novo Assembly, Annotation, and GO Term Analysis

Paired end reads from all three biological replicates were subjected to quality-based trimming using the FASTX toolkit (Goecks et al., 2010), setting the PHRED quality threshold at 20 and discarding reads less than 50 nucleotides in length. Further processing was then performed to remove reads corresponding to poly-A tails and reads containing more than 75% of any single nucleotide. Processed reads were then subjected to read-error correction using ALLPATHS-LG (Maccallum et al., 2009) and filtered to remove all redundant read-pairs. These reads were then filtered to remove rRNA using SortMeRNA (Kopylova et al., 2012) before reads that contained only unique k-mers were discarded. The processed read set was then subjected to de novo assembly using Velvet/Oases (Zerbino and Birney, 2008; Schulz et al., 2012) using four different k-mer lengths (k = 31, 41, 51, 61) and merged using Oases. Redundant transcripts and partial transcripts (for which a longer transcript was present that contained >95% of the nucleotides of the shorter) were discarded using USEARCH (Edgar, 2010).

To estimate transcript abundances, the original unprocessed reads were subjected to quality-based trimming using the FASTX toolkit (Goecks et al., 2010), setting the PHRED quality threshold at 20 and discarding reads less than 21 nucleotides in length. RSEM was used on the trimmed reads to quantify the assembled transcripts for *E. glabrescens* and the reference transcripts for rice (Li and Dewey, 2011). Assembled transcripts were annotated with orthologous relationships to rice genes using the conditional reciprocal best-BLAST annotation method, which has ~82% accuracy (Aubry et al., 2014), and the rice representative gene models from Phytozome V9 (Goodstein et al., 2012).

For the two-way differential expression analysis, assembled *E. glabrescens* transcripts were grouped into clusters according to their predicted orthology to rice genes, and expected counts were summed within each cluster to give one expected count value per cluster for each sample. Data for assembled transcripts that did not map to a rice gene or that had zero expected count in all three *E. glabrescens* replicates were excluded from downstream analysis. All possible expression difference patterns between conditions, including equal expression, were assigned a posterior probability using EBSeq v1.4.0 (Leng et al., 2013), allowing 5 rounds of expectation maximization for the optimization of model parameters. GO terms from Phytozome V9 (Goodstein et al., 2012) were assigned to each gene, and the weighted Fisher's exact test with a *p*-value threshold of 0.05 for enrichment of GO terms was performed for each expression pattern using TopGO v2.20 (Alexa et al., 2006).

In the three-way species differential expression analysis, expression data for two replicates of maize whole-leaf were taken from Wang et al. (2013a). Maize expression data were first clustered by homeology, and expected counts were summed per homeologous cluster and then annotated with orthologous relationships to rice using the same procedure as for *E. glabrescens.* Only genes that were present in all three datasets were included in subsequent analysis. Differential expression analysis and GO term enrichment were performed as for the two-way species analysis with each of *E. glabrescens*, rice, and maize treated as a condition.

Immunoblotting

E. glabrescens, *Sorghum bicolor* cv Rooney, and *O. sativa* cv IR64 were grown at the University of Sheffield. Seeds were germinated in trays of Levington M3 compost (Scotts, Ipswich, UK) in a growth chamber at a light intensity of 450 μmol m^{-2} s^{-1} with a 12 h photoperiod (27°C day, 23°C night, 60% humidity). Seedlings were transferred to 100 mm pots with compost supplemented with 3 g Osmocote Exact standard pellets (Scotts, Ipswich, UK) after 7 d. Samples were taken from the fully expanded fourth leaf on the main tiller or stem of all plants as the fifth leaf was just emerging. Plant material (100 mg) was homogenized in a mortar containing 5 volumes of ice-cold 200 mM Bicine-KOH (pH 9.8), 50 mM DTT and clarified by centrifugation at 14,000 g for 3 min. Supernatants were added to an equal volume of SDS-PAGE solubilization buffer (62.5 mM Tris/HCl [pH 6.8], 10% [v/v] glycerol, 5% [w/v] SDS, 5% [v/v] 2-mercaptoethanol, 0.002% [w/v] bromophenol blue), placed at 100°C for 3 min and then frozen until use. Clarified supernatants were analyzed by SDS-PAGE using a 4.7% T/2.7% C stacking gel and either a 7.5% T/2.7% C or 12% T/2.7% C resolving gel. After electrophoresis, transfer of polypeptides from an SDS-PAGE gel to immunoblot PVDF membrane (Bio-Rad, Hemel Hempstead, UK) was performed with a Bio-Rad Tetra cell blotting apparatus. Immunoreactive polypeptides were visualized using the relevant antiserum in conjunction with an ECL kit and Hyperfilm ECL (GE Healthcare, UK). Antisera to PEPCK, NADP-ME, NAD-ME, PPDK, and MDH were affinity-purified and developed specifically to peptide sequences in rice. PEPC antiserum was raised from purified *Amaranthus edulis* protein and CA antiserum, a gift from Professor J. Burnell (James Cooke University, Australia), was raised from purified recombinant maize protein.

Subcellular Localization of PEPC

The full-length coding sequence of *E. glabrescens* transcript_8661 was amplified using the following primers: forward 5′CACCATGGTTTCCCTCCGTCT and reverse 5′GCCGGTGTTCTGCATGC. The first 55 residues of Si000160m and the first 48 residues of Bradi2g06620, encompassing the N-terminal extensions for these genes, were optimized for rice codon usage and synthesized by Life Technologies (UK) as Strings DNA fragments with a Gateway compatible CACC at the 5′ end of each sequence for pENTR/D-TOPO cloning following the manufacturer's protocol (Life Technologies, USA). Clones were sequence verified and transferred to the pGWB5 destination vector (GenBank accession no. AB289768) using Gateway LR Clonase II Enzyme mix (Life Technologies, USA). *Agrobacterium* strain GV 3101 was transformed with each vector and grown in 5 mL LB medium containing 50 μg/mL Kanamycin and 25 μg/mL Rifampicin overnight. Cultures were centrifuged at 4500 g for 15 min. The resulting pellet was resuspended to OD_{600} 1.0 in AS medium (10 mM $MgCl_2$, 10 mM MES-KOH pH 5.6, 0.15 mM Acetosyringone) and incubated for 4 h at 28°C. Three-week-old *N. benthamiana* leaves were infiltrated by syringe with culture. Plants were kept in the growth chamber for 48 h prior to visualization of GFP using a Leica TCS-SP5 microscope and Leica LAS AF software (Leica, Germany).

Estimation of Nonsynonymous Substitution Rates

To estimate the rates of gene sequence evolution, the assembled *E. glabrescens* transcript sequences were aligned to representative gene models from the rice genome using the MAFFT linsi method (Katoh et al., 2005). Only transcripts that were single copy in the transcriptome and contained no splice or sequence variants were selected for analysis. Pairwise sequence alignments were assessed, and only those with alignments covering >75% of the rice transcript sequence and >100 codons were selected for further analysis to prevent partial transcripts from biasing the results. The resulting alignments were then subject to molecular sequence evolution analysis using KaKs calculator (Wang et al., 2010). For the purposes of this test, non-differentially regulated genes included all those from the three-species analysis that were not in the other categories under assessment.

Accession Numbers

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers PRJEB9593 and KT963028.

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Correlation between biological replicates used in differential expression analysis.

Supplemental Figure S2. Cell specificity of transcription factors more from *Echinochloa glabrescens*.

Supplemental Table S1. Patterns of transcript abundance in *Echinochloa glabrescens* and *Oryza sativa*.

Supplemental Table S2. Statistically overrepresented Gene Ontology terms in the two-species analysis.

Supplemental Table S3. Transcription factors in the two-species analysis overlapping with previous studies.

Supplemental Table S4. Transcript abundance of core C4 pathway, metabolite transport, and photorespiration.

Supplemental Table S5. ChloroP predictions for PEPC proteins.

Supplemental Table S6. Patterns of transcript abundance in *Echinochloa glabrescens*, *Oryza sativa*, and *Zea mays*.

Supplemental Table S7. Gene Ontology terms from *Echinochloa glabrescens*, *Oryza sativa*, and *Zea mays*.

Supplemental Table S8. Motif analysis of transcription factors up-regulated in C4 or paddy species.

Supplemental Table S9. Transcription factors in *Echinochloa glabrescens*, *Oryza sativa*, and *Zea mays* and overlapping with previous studies.

Supplemental Table S10. Motif analysis for previously identified transcription factors that were up-regulated in C4 or paddy species.

Supplemental Table S11. Motifs detected overlapping with previous studies.

ACKNOWLEDGMENTS

We thank Paul Quick, Shanta Karki, and the C_4 Rice Team at the International Rice Research Institute for growing *Echinochloa glabrescens* and rice seedlings for comparative transcriptomics.

Received June 10, 2015; revised October 22, 2015; accepted November 2, 2015; published November 2, 2015.

LITERATURE CITED

Alexa A, Rahnenführer J, Lengauer T (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics **22:** 1600–1607

Arumuganathan K, Slattery JP, Tanksley SD, Earle ED (1991) Preparation and flow cytometric analysis of metaphase chromosomes of tomato. Theor Appl Genet **82:** 101–111

Aubry S, Kelly S, Kümpers BM, Smith-Unna RD, Hibberd JM (2014) Deep evolutionary comparison of gene expression identifies parallel recruitment of *trans*-factors in two independent origins of C_4 photosynthesis. PLoS Genet **10:** e1004365

Barrett SCH (1983) Crop mimicry in weeds. Econ Bot **37:** 255–282

Bellasio C, Griffiths H (2014) The operation of two decarboxylases, transamination, and partitioning of C_4 metabolic processes between mesophyll and bundle sheath cells allows light capture to be balanced for the maize C_4 pathway. Plant Physiol **164:** 466–480

Bläsing OE, Westhoff P, Svensson P (2000) Evolution of C_4 phospho*enol*pyruvate carboxylase in *Flaveria*, a conserved serine residue in the carboxyl-terminal part of the enzyme is a major determinant for C_4-specific characteristics. J Biol Chem **275:** 27917–27923

Bräutigam A, Hoffmann-Benning S, Weber AP, Weber AP (2008) Comparative proteomics of chloroplast envelopes from C_3 and C_4 plants reveals specific adaptations of the plastid envelope to C_4 photosynthesis and candidate proteins required for maintaining C_4 metabolite fluxes. Plant Physiol **148:** 568–579

Bräutigam A, Kajala K, Wullenweber J, Sommer M, Gagneul D, Weber KL, Carr KM, Gowik U, Mass J, Lercher MJ, et al (2011) An mRNA blueprint for C_4 photosynthesis derived from comparative transcriptomics of closely related C_3 and C_4 species. Plant Physiol **155:** 142–156

Burnell JN (2011) Hurdles to engineering greater photosynthetic rates in crop plants: C_4 rice. *In* A Raghavendra, R Sage, eds, C_4 photosynthesis and related CO_2 concentrating mechanisms, **Vol 32**. Springer, Dordrecht, Netherlands, pp 361-378

Chang YM, Liu WY, Shih AC, Shen MN, Lu CH, Lu MY, Yang HW, Wang TY, Chen SC, Chen SM, et al (2012) Characterizing regulatory and functional differentiation between maize mesophyll and bundle sheath cells by transcriptomic analysis. Plant Physiol **160:** 165–177

Chapman KSR, Hatch MD (1981) Aspartate decarboxylation in bundle sheath-cells of *Zea mays* and its possible contribution to C_4 photosynthesis. Aust J Plant Physiol **8:** 237–248

Christin PA, Besnard G, Samaritani E, Duvall MR, Hodkinson TR, Savolainen V, Salamin N (2008) Oligocene CO_2 decline promoted C_4 photosynthesis in grasses. Curr Biol **18:** 37–43

Christin PA, Salamin N, Savolainen V, Duvall MR, Besnard G (2007) C_4 Photosynthesis evolved in grasses via parallel adaptive genetic changes. Curr Biol **17:** 1241–1247

Covshoff S, Furbank RT, Leegood RC, Hibberd JM (2012) Leaf rolling allows quantification of mRNA abundance in mesophyll cells of sorghum. J Exp Bot **64:** 807–813

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics **26:** 2460–2461

Edwards GE, Voznesenskaya EV (2011) C_4 photosynthesis: Kranz forms and single-cell C_4 in terrestrial plants. *In* A Raghavendra, R Sage, eds, C_4 Photosynthesis and Related CO_2 Concentrating Mechanisms. Kluwer Academic Publishers, pp 29–60

Ehleringer JR, Monson RK (1993) Evolutionary and ecological aspects of photosynthetic pathway variation. Annu Rev Ecol Syst **24:** 411–439

Emanuelsson O, Nielsen H, von Heijne G (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Sci **8:** 978–984

Fitter DW, Martin DJ, Copley MJ, Scotland RW, Langdale JA (2002) *GLK* gene pairs regulate chloroplast development in diverse plant species. Plant J **31:** 713–727

Fukayama H, Fujiwara N, Hatanaka T, Misoo S, Miyao M (2014) Nocturnal phosphorylation of phospho*enol*pyruvate carboxylase in the leaves of hygrophytic C_3 monocots. Biosci Biotechnol Biochem **78:** 609–613

Fukayama H, Tamai T, Taniguchi Y, Sullivan S, Miyao M, Nimmo HG (2006) Characterization and functional analysis of phosphoenolpyruvate carboxylase kinase genes in rice. Plant J **47:** 258–268

Furbank RT (2011) Evolution of the $C_{(4)}$ photosynthetic mechanism: are there really three $C_{(4)}$ acid decarboxylation types? J Exp Bot **62:** 3103–3108

Furumoto T, Hata S, Izui K (1999) cDNA cloning and characterization of maize phospho*enol*pyruvate carboxykinase, a bundle sheath cell-specific enzyme. Plant Mol Biol **41:** 301–311

Furumoto T, Hata S, Izui K (2000) Isolation and characterization of cDNAs for differentially accumulated transcripts between mesophyll cells and bundle sheath strands of maize leaves. Plant Cell Physiol **41:** 1200–1209

Giussani LM, Cota-Sánchez JH, Zuloaga FO, Kellogg EA (2001) A molecular phylogeny of the grass subfamily Panicoideae (Poaceae) shows multiple origins of C_4 photosynthesis. Am J Bot **88:** 1993–2012

Goecks J, Nekrutenko A, Taylor J; Galaxy Team (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol **11:** R86

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, et al (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res **40:** D1178–D1186

Gross BL, Zhao Z (2014) Archaeological and genetic insights into the origins of domesticated rice. Proc Natl Acad Sci USA **111:** 6190–6197

Gutierrez M, Gracen VE, Edwards GE (1974) Biochemical and cytological relationships in C_4 plants. Planta **119:** 279–300

Gutiérrez RA, Stokes TL, Thum K, Xu X, Obertello M, Katari MS, Tanurdzic M, Dean A, Nero DC, McClung CR, et al (2008) Systems

approach identifies an organic nitrogen-responsive gene network that is regulated by the master clock control gene *CCA1*. Proc Natl Acad Sci USA **105:** 4939–4944

Hall LN, Rossini L, Cribb L, Langdale JA (1998) *GOLDEN 2*: a novel transcriptional regulator of cellular differentiation in the maize leaf. Plant Cell **10:** 925–936

Hatch MD (1971) The C 4 -pathway of photosynthesis. Evidence for an intermediate pool of carbon dioxide and the identity of the donor C 4 -dicarboxylic acid. Biochem J **125:** 425–432

Hatch MD (1987) C_4 photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. Biochim Biophys Acta **895:** 81–106

Hilu KW (1994) Evidence from RAPD markers in the evolution of *Echinochloa* millets (Poaceae). Plant Syst Evol **189:** 247–257

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 Physiol **165:** 62–75

Kajala K, Covshoff S, Karki S, Woodfield H, Tolley BJ, Dionora MJA, Mogul R, Elmido-Mabilangan A, Danila F, Hibberd JM, et al (2011) Strategies for engineering a two-celled C4 photosynthetic pathway into rice. J Exp Bot **62:** 3001–3010

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

Kopylova E, Noé L, Touzet H (2012) SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics **28:** 3211–3217

Leng N, Dawson JA, Thomson JA, Ruotti V, Rissman AI, Smits BM, Haag JD, Gould MN, Stewart RM, Kendziorski C (2013) EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. Bioinformatics **29:** 1035–1043

Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics **12:** 323

Li P, Ponnala L, Gandotra N, Wang L, Si Y, Tausta SL, Kebrom TH, Provart N, Patel R, Myers CR, et al (2010) The developmental dynamics of the maize leaf transcriptome. Nat Genet **42:** 1060–1067

Long SP (1999) Environmental responses. *In* RF Sage, RK Monson, eds, C_4 plant biology. Academic Press, San Diego, CA, USA, pp 215–249

Maccallum I, Przybylski D, Gnerre S, Burton J, Shlyakhter I, Gnirke A, Malek J, McKernan K, Ranade S, Shea TP, et al (2009) ALLPATHS 2: small genomes assembled accurately and with high continuity from short paired reads. Genome Biol **10:** R103

Majeran W, Cai Y, Sun Q, van Wijk KJ (2005) Functional differentiation of bundle sheath and mesophyll maize chloroplasts determined by comparative proteomics. Plant Cell **17:** 3111–3140

Majeran W, Friso G, Ponnala L, Connolly B, Huang M, Reidel E, Zhang C, Asakura Y, Bhuiyan NH, Sun Q, et al (2010) Structural and metabolic transitions of C_4 leaf development and differentiation defined by microscopy and quantitative proteomics in maize. Plant Cell **22:** 3509–3542

Majeran W, Zybailov B, Ytterberg AJ, Dunsmore J, Sun Q, van Wijk KJ (2008) Consequences of C_4 differentiation for chloroplast membrane proteomes in maize mesophyll and bundle sheath cells. Mol Cell Proteomics **7:** 1609–1638

Masumoto C, Miyazawa S, Ohkawa H, Fukuda T, Taniguchi Y, Murayama S, Kusano M, Saito K, Fukayama H, Miyao M (2010) Phospho-*enol*pyruvate carboxylase intrinsically located in the chloroplast of rice plays a crucial role in ammonium assimilation. Proc Natl Acad Sci USA **107:** 5226–5231

Mockler TC, Michael TP, Priest HD, Shen R, Sullivan CM, Givan SA, McEntee C, Kay SA, Chory J (2007) The DIURNAL project: DIURNAL and circadian expression profiling, model-based pattern matching, and promoter analysis. Cold Spring Harb Symp Quant Biol **72:** 353–363

Muramatsu M, Suzuki R, Yamazaki T, Miyao M (2015) Comparison of plant-type phosphoenolpyruvate carboxylases from rice: identification of two plant-specific regulatory regions of the allosteric enzyme. Plant Cell Physiol **56:** 468–480

Nelson T, Dengler NG (1992) Photosynthetic tissue differentiation in C_4 plants. Int J Plant Sci **153:** S93–S105

Nguyen CV, Vrebalov JT, Gapper NE, Zheng Y, Zhong S, Fei Z, Giovannoni JJ (2014) Tomato *GOLDEN2-LIKE* transcription factors reveal molecular gradients that function during fruit development and ripening. Plant Cell **26:** 585–601

Osborne CP, Sack L (2012) Evolution of C_4 plants: a new hypothesis for an interaction of CO_2 and water relations mediated by plant hydraulics. Philos Trans R Soc Lond B Biol Sci **367:** 583–600

Panteris E, Galatis B (2005) The morphogenesis of lobed plant cells in the mesophyll and epidermis: organization and distinct roles of cortical microtubules and actin filaments. New Phytol **167:** 721–732

Pick TR, Bräutigam A, Schlüter U, Denton AK, Colmsee C, Scholz U, Fahnenstich H, Pieruschka R, Rascher U, Sonnewald U, et al (2011) Systems analysis of a maize leaf developmental gradient redefines the current C_4 model and provides candidates for regulation. Plant Cell **23:** 4208–4220

Pick TR, Bräutigam A, Schulz MA, Obata T, Fernie AR, Weber AP (2013) *PLGG1*, a plastidic glycolate glycerate transporter, is required for photorespiration and defines a unique class of metabolite transporters. Proc Natl Acad Sci USA **110:** 3185–3190

Piedade MTF, Junk WJ, Long SP (1991) The productivity of the C_4 grass *Echinochloa polystachya* on the Amazon floodplain. Ecology **72:** 1456–1463

Powell AL, Nguyen CV, Hill T, Cheng KL, Figueroa-Balderas R, Aktas H, Ashrafi H, Pons C, Fernández-Muñoz R, Vicente A, et al (2012) *Uniform ripening* encodes a *Golden 2-like* transcription factor regulating tomato fruit chloroplast development. Science **336:** 1711–1715

Rossini L, Cribb L, Martin DJ, Langdale JA (2001) The maize *golden2* gene defines a novel class of transcriptional regulators in plants. Plant Cell **13:** 1231–1244

Sage RF (2004) The evolution of C_4 photosynthesis. New Phytol **161:** 341–370

Sage RF, Christin PA, Edwards EJ (2011) The $C_{(4)}$ plant lineages of planet Earth. J Exp Bot **62:** 3155–3169

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

Sage TL, Sage RF (2009) The functional anatomy of rice leaves: implications for refixation of photorespiratory CO_2 and efforts to engineer C_4 photosynthesis into rice. Plant Cell Physiol **50:** 756–772

Sage TL, Williams EG (1995) Structure, ultrastructure, and histochemistry of the pollen tube pathway in the milkweed *Asclepias exaltata* L. Sex Plant Reprod **8:** 257–265

Schulz MH, Zerbino DR, Vingron M, Birney E (2012) Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics **28:** 1086–1092

Slewinski TL (2013) Using evolution as a guide to engineer kranz-type c_4 photosynthesis. Front Plant Sci **4:** 212

Slewinski TL, Anderson AA, Zhang C, Turgeon R (2012) Scarecrow plays a role in establishing Kranz anatomy in maize leaves. Plant Cell Physiol **53:** 2030–2037

Smith BN, Brown WV (1973) The Kranz syndrome in the gramineae as indicated by carbon isotopic ratios. Am J Bot **60:** 505–513

Stata M, Sage TL, Rennie TD, Khoshravesh R, Sultmanis S, Khaikin Y, Ludwig M, Sage RF (2014) Mesophyll cells of C_4 plants have fewer chloroplasts than those of closely related C_3 plants. Plant Cell Environ **37:** 2587–2600

Suzuki S, Burnell JN (1995) Nucleotide sequence of a cDNA encoding rice chloroplastic carbonic anhydrase. Plant Physiol **107:** 299–300

Taniguchi Y, Nagasaki J, Kawasaki M, Miyake H, Sugiyama T, Taniguchi M (2004) Differentiation of dicarboxylate transporters in mesophyll and bundle sheath chloroplasts of maize. Plant Cell Physiol **45:** 187–200

Taylor SH, Hulme SP, Rees M, Ripley BS, Woodward FI, Osborne CP (2010) Ecophysiological traits in C_3 and C_4 grasses: a phylogenetically controlled screening experiment. New Phytol **185:** 780–791

von Caemmerer S, Ghannoum O, Pengelly JJ, Cousins AB (2014) Carbon isotope discrimination as a tool to explore C_4 photosynthesis. J Exp Bot **65:** 3459–3470

Wang D, Zhang Y, Zhang Z, Zhu J, Yu J (2010) KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genomics Proteomics Bioinformatics **8:** 77–80

Wang L, Czedik-Eysenberg A, Mertz RA, Si Y, Tohge T, Nunes-Nesi A, Arrivault S, Dedow LK, Bryant DW, Zhou W, et al (2014b) Comparative analyses of C_4 and C_3 photosynthesis in developing leaves of maize and rice. Nat Biotechnol **32:** 1158–1165

Wang P, Fouracre J, Kelly S, Karki S, Gowik U, Aubry S, Shaw MK, Westhoff P, Slamet-Loedin IH, Quick WP, et al (2013b) Evolution of *GOLDEN2-LIKE* gene function in $C_{(3)}$ and $C_{(4)}$ plants. Planta **237:** 481–495

Wang P, Kelly S, Fouracre JP, Langdale JA (2013a) Genome-wide transcript analysis of early maize leaf development reveals gene cohorts associated with the differentiation of C_4 Kranz anatomy. Plant J **75:** 656–670

Wang Y, Bräutigam A, Weber AP, Zhu XG (2014a) Three distinct biochemical subtypes of C_4 photosynthesis? A modelling analysis. J Exp Bot **65:** 3567–3578

Waters MT, Wang P, Korkaric M, Capper RG, Saunders NJ, Langdale JA (2009) GLK transcription factors coordinate expression of the photosynthetic apparatus in *Arabidopsis.* Plant Cell **21:** 1109–1128

Wingler A, Walker RP, Chen ZH, Leegood RC (1999) Phospho*enol*pyruvate carboxykinase is involved in the decarboxylation of aspartate in the bundle sheath of maize. Plant Physiol **120:** 539–546

Ye CY, Lin Z, Li G, Wang YY, Qiu J, Fu F, Zhang H, Chen L, Ye S, Song W, et al (2014) *Echinochloa* chloroplast genomes: insights into the evolution and taxonomic identification of two weedy species. PLoS One **9:** e113657

Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res **18:** 821–829

Zonneveld BJ, Leitch IJ, Bennett MD (2005) First nuclear DNA amounts in more than 300 angiosperms. Ann Bot (Lond) **96:** 229–244