Reconstructed protein sequence evolution consistent with the evolution of C_4 photosynthesis via a C_2 ancestor in the Paniceae

Daniel S. Carvalho[1], Sunil Kumar Kenchanmane Raju[2], Yang Zhang[1,3] & James C. Schnable[1]

1. Center for Plant Science Innovation, Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE, USA.
2. Department of Plant Biology, Michigan State University, East Lansing, MI, USA.
3. Present Address: St. Jude Children's Research Hospital, Memphis, TN, USA.

ORCID: 0000-0001-6590-4287 (DSC), 0000-0001-8960-094X (SKKR), 0000-0003-1712-7211 (YZ), 0000-0001-6739-5527 (JCS).

Abstract

The grass tribe Paniceae includes a monophyletic subclade of species, the MPC clade, which specialize in each of the three primary C_4 sub-pathways NADP-ME, NAD-ME and PCK. The evolutionary history of C_4 photosynthesis in this subclade remains ambiguous. Leveraging newly sequenced grass genomes and syntenic orthology data, we estimated rates of protein sequence evolution on ancestral branches for both core enzymes shared across different C_4 sub-pathways and enzymes specific to C_4 sub-pathways. While core enzymes show elevated rates of protein sequence evolution in ancestral branches consistent with a transition from C_3 to C_4 photosynthesis in the ancestor for this clade, no subtype specific enzymes showed similar patterns. At least one protein involved in photorespiration also showed elevated rates of protein sequence evolution in the ancestral branch. The set of core C_4 enzymes examined here combined with the photorespiratory pathway are necessary for the C_2 photosynthetic cycle, a previously proposed intermediate between C_3 and C_4 photosynthesis. The patterns reported here are consistent with, but not conclusive proof that, C_4 photosynthesis in the MPC clade of the Paniceae evolved via a C_2 intermediate.

Introduction

C_4 plants are responsible for over a quarter of world terrestrial photosynthetic productivity (Gillon and Yakir, 2001; Sage, 2001). C_4 grasses alone account for approximately 18% of global productivity (Ehleringer et al., 1997; Wand et al., 1999). The high productivity of C_4 plants is linked to their ability to increase the ratio of CO_2 to O_2 around RuBisCO, lowering the net energy negative process of photorespiration (Bull, 1969; Keys, 1986), and reducing water losses to transpiration. C_4 plants are able to keep their stomata closed for longer because they are less sensitive to declines in intraleaf CO_2 concentrations than plants dependent on C_3 photosynthesis. C_4 photosynthesis is thought to have arisen in parallel in over 60 lineages approximately 30 million years ago in responses to a drop in CO_2 levels (Vicentini et al., 2008; Sage et al., 2011).

While exceptions exist (Akhani et al., 2003; Offermann et al., 2015), many lineages utilizing C_4 photosynthesis split the process between two different cell types: mesophyll (M) and bundle-sheath (BS) cells. In the M cells, CO_2 is fixed into bicarbonate to form oxaloacetate (OAA). OAA is converted to either malate or aspartate or both depending on the species and the C_4 pathway being used. Malate or aspartate is then transported to the BS cells where it is decarboxylated, releasing CO_2 which is then fixed via the conventional Calvin-Benson cycle (Kanai and Edwards, 1999). While this broad pattern is consistently found in a wide range of C_4 using species, the C_4 photosynthesis cycle can be divided into three subtypes based on 1) the decarboxylase enzyme which releases CO_2 within the bundle sheath and 2) the linked property of where decarboxylation occurs. The three enzyme families known to act as decarboxylases for C_4 as NAD-malic

enzyme (NAD-ME) which decarboxylates in the mitochondria, NADP-malic enzyme (NADP-ME) which decarboxylates in the chloroplast and PEP carboxykinase (PCK) which decarboxylates in the cytosol (Kanai and Edwards, 1999; Furbank, 2011). Individual species might utilize a single decarboxylase and hence a single C_4 pathway or multiple decarboxylases and multiple C_4 pathways (Walker et al., 1997; John et al., 2014; Huang et al., 2016; de Oliveira Dal'Molin et al., 2016; Washburn et al., 2017).

The tribe Paniceae within the grasses includes 84 genera (Morrone et al., 2012). Notably the clade encompasses different species which utilize each of the the three C_4 pathways as their primary carbon fixation mechanism. These species share a single common ancestor to the exclusion of any know C_3 lineage (Sage et al., 2011; Washburn et al., 2015).

This clade containing only C_4 photosynthesizers encompasses the subtribes Melinidinae (PCK subtype), Panicinae (NAD-ME subtype) and Cenchrinae (NADP-ME subtype) and is also referred to as the MPC clade. Ancestral state reconstruction based on phylogenies and current state data identified either NAD-ME subtype of C_4 photosynthesis or C_3 photosynthesis as the likely ancestral state of the MPC clade (Washburn et al., 2015). Ancestral state reconstruction based on expression data suggested that the common ancestor may have been a species using all three C_4 pathways simultaneously (Washburn et al., 2017).

Here we sought to employ a third approach based on the reconstruction of rates of protein sequence evolution along ancestral branches of the phylogeny. Estimates of changes in protein sequence evolution rates between C_3 lineages and ancestral branches along the phylogeny of the MPC clade were used to evaluate several potential models for the photosynthetic state of the common ancestor of the MPC clade. While several core enzymes shared by all three C_4 pathways did indeed show elevated rates of protein sequence evolution in the ancestral branch leading to the MPC clade, none of the three decarboxylases showed elevated rates of protein sequence evolution along this branch. At least one gene involved in photorespiration also showed elevated rates of protein sequence evolution in the ancestral MPC branch. Our findings are suggestive of a C_2 pathway intermediate ancestor at the base of the MPC clade. The C_2 uses the photorespiratory pathway as a CO_2 carbon pump and has been proposed as a potential intermediate state between the conventional C_3 and C_4 photosynthetic pathways (Tolbert, 1997; Mallmann et al., 2014; Edwards, 2019). Taken together, these results may represent a "ghost of C_2 past" in the genomes of this C_4 clade Edwards (2019).

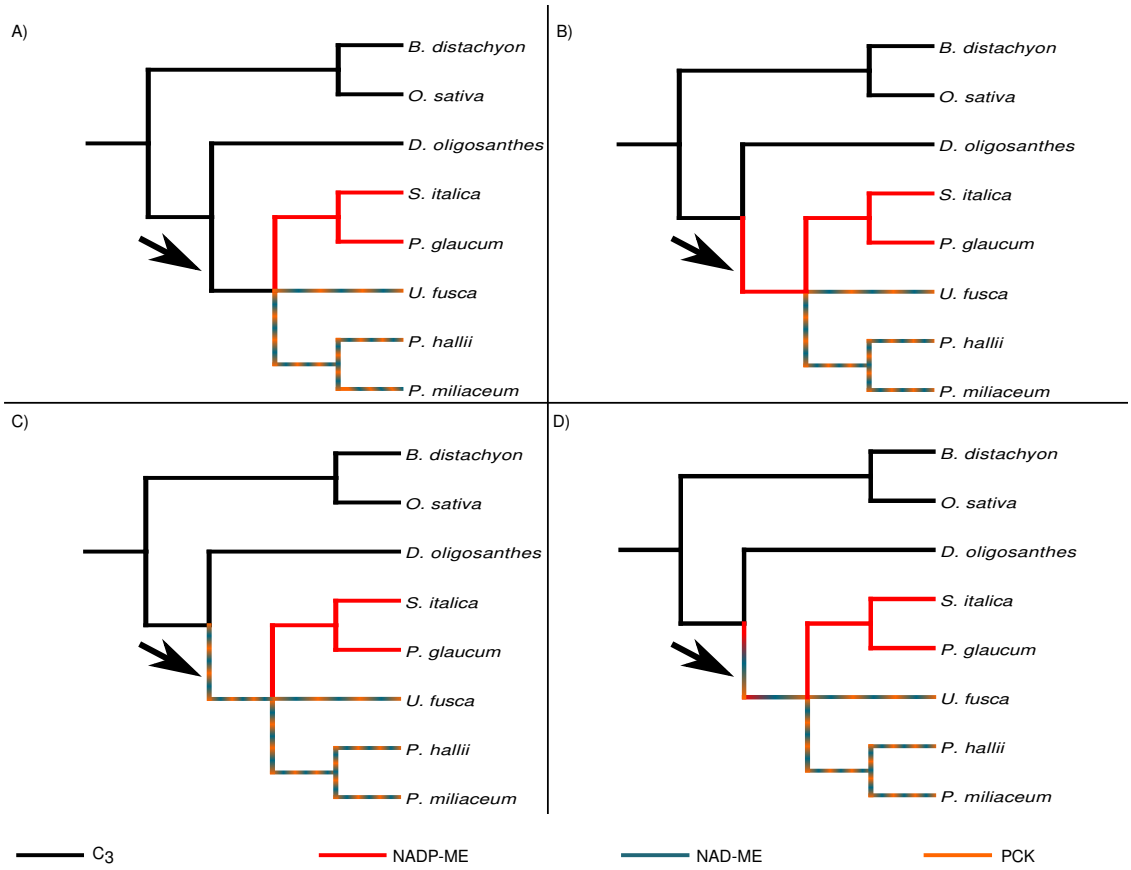

Figure 1: **Potential models for the evolution of C_4 photosynthesis in the MPC clade.** A) The common ancestor of the MPC clade utilized C_3 or another non-C_4 pathway and C_4 evolved independently in the three subtribes utilizing different C_4 pathways Washburn et al. (2015); B) The common ancestor of the MPC clade utilized the C_4 NADP-ME pathway and the NAD-ME and PCK clades represent later evolutionary changes from one C_4 subtype to another; C) The common ancestor of the MPC clade utilized either the C_4 NAD-ME and PCK pathways or a mix of both; D) The common ancestor of the MPC clade utilized utilized all three pathways simultaneously (Washburn et al., 2017). Black arrow indicates the ancestral branch for the MPC subclade within the Paniceae.

Material and methods

Plant growth and RNA-Seq data generation for *Urochloa fusca*

Urochloa fusca seeds were planted and grown in a Percival (Percival model E-41L2) growth chamber with target conditions of 111 μmol m^{-2} s^{-1}, 60% relative humidity, a 12 hour/12 hour day night cycle with a target temperature of 29°C during the day and 23°C at night. Under the growing conditions employed, twelve days after planting (DAP) plants were collected. The whole seedlings (except root) were used for RNA extraction. RNA isolation and library construction followed the protocol described by Zhang *et al.* (Zhang et al., 2015). The library was sequenced using 2x100 bp paired end Illumina sequencing, and transcripts were generated using Trinity (v2.0.6) with default parameters with the exception of "–seqType fq –max_memory 20G –CPU 16" (Grabherr et al., 2011).

Sequence data set

Coding Sequences (CDS) for the transcript annotated as "primary" for each gene in *Brachypodium distachyon* (Initiative et al., 2010), *Oryza sativa* (rice) (Kawahara et al., 2013), *Panicum hallii* (Lovell et al., 2018) and *Setaria italica* (foxtail millet) (Bennetzen et al., 2012) were obtained

from Phytozome 12 (https://phytozome.jgi.doe.gov/pz/portal.html). For *Dichanthelium oligosanthes* the CDS were retrieved from version v1.001 in CoGe OrganismView, genome ID 28856 (https://genomevolution.org/CoGe/OrganismView.pl) (Studer et al., 2016). CDS from *Pennisetum glaucum* (pearl millet) were obtained from (Varshney et al., 2017). CDS from *Panicum miliaceum* (proso millet) were downloaded from CoGe (Genome ID: 52484), version v1 (Lyons and Freeling, 2008; Zou et al., 2019). Open reading frames for *Urochloa fusca* were obtained from the transcriptome assembly described above. As a number of transcript assemblies were not full length in manual proofing, open reading frames were constrained to those ending in a stop codon, but were not required to start with an ATG.

Orthology assignment

LASTZ (Harris, 2007) was used to perform all by all comparisons of coding sequence from the primary transcript of each gene as downloaded from from Phytozome 12 with the following parameters –identity=70 –coverage=50 –ambiguous=iupac, –notransition, and –seed=match12. LASTZ output was parsed to identify syntenic orthologs using QuotaAlign with the additional parameters –tandemNmax=10, cscore=0.5, –merge and –Dm=20 (Tang et al., 2011). To reverse the collapse of tandem gene clusters which are part of the QuotaAlign algorithm, final syntenic orthologs were assigned based on the gene copy with the highest LASTZ alignment score within 20 genes up or downstream of the original syntenic location predicted by quota align. Synteny analysis of proso and pearl millet could not retrieve all C_4 related genes possibly due to larger amounts of genomic rearrangement and smaller blocks of conserved synteny relative to other grass species included in this study. When no syntenic ortholog could be identified in pearl millet and/or proso millet based on synteny, the best reciprocal LASTZ hit to the foxtail millet gene copy was included as a presumed ortholog, even in the absence of synteny. A similar process was employed for four genes from the *Dichanthelium oligosanthes* genome, as this genome was assembled using purely short read technology and remains fragmented into a large number of scaffolds. In all cases, orthology for *Urochloa fusca* sequences was inferred based on reciprocal best LASTZ hits with the foxtail millet C_4 genes. The list of syntenic orthologous and presumed orthologous gene sequences analyzed here is provided in supplemental material 2.

dN/dS calculation and evolutionary analyses

The DNA sequence of the coding sequence region of each gene was translated to an amino acid sequence. Amino acid sequences for groups of orthologous and presumed orthologous genes were aligned using Kalign version 2.04 (Lassmann and Sonnhammer, 2005). This amino acid multiple sequence alignment was used to create a codon level DNA sequence alignment. The codon alignment and a guide phylogenetic tree created based on the known relationships of all species included in the analysis (Edwards et al., 2011) were used to calculate the nonsynonymous and synonymous substitution rates (dN/dS) for each branch of the tree using codeml from the PAML package version 4.09 (Yang, 2007).

Statistical comparison of branch dN/dS values and photosynthetic trait values

A Fisher Exact Test was performed in order to test whether significant differences in evolutionary rate existed between between branches leading to species utilizing C_3 photosynthesis and branches leading to species utilizing C_4 photosynthesis. For each comparison, a contingency table was constructed comparing the separate estimated number of synonymous and nonsynonymous substitutions in two branches or groups of branches. For each gene, each one of the C_4 branches was tested for an elevated ratio of nonsynonymous to synonymous substitutions relative to the aggregate background C_3 branches. C_3 background values for both non-synonymous and synonymous substitutions were summed across three branches each leading to a single C_3 species: *D. oligosanthes*, *B. distachyon*, and *O. sativa*. A Fisher Exact Test was used to compare each tested branch to the aggregate C_3 branch values.

Results

The evolutionary patterns of ten genes encoding enzymes either unique to individual pathways of C_4 photosynthesis or shared by two or more pathways were investigated (Figure 2). The selection of genes for investigation was guided both by the strength of literature supporting their specific roles in one or more pathways of C_4 photosynthesis and the identification of high confidence orthologs in as many of the grass species included in this analysis (Table 1). Two different signatures were possible for genes which experienced positive selection as part of a transition from C_3 to C_4 photosynthesis. The first of these two potential signatures is an unambiguous signal where the rate of nonsynonymous substitutions (dN) exceeds the rate of synonymous substitutions (dS) on a given branch (i.e. dN/dS >1) (Figure 3). This signal can be observed either on extremely short branches, or when positive selection continued for long periods of time. The second potential signal is more ambiguous. When an interval of positive selection occurs on a given branch but is preceded and/or followed by purifying within the same branch, the effects of these two intervals cannot be disambiguated without including additional outgroups to break up the branch into short segments. As a result, in branches that include both positive and purifying selection dN/dS will be elevated relative to branches experiencing purely purifying selection. However, despite the presence of positive selection, the dN/dS ratio for the entire branch may still still be less than 1 (Figure 3). Unfortunately, the same signature (elevated dN/dS ratios relative to other branches but not exceeding a ratio of one) is also produced in cases of relaxed selection of pseudogenization. Here we considered either case to be potentially informative about the origins of C_4 photosynthesis in the MPC clade, however the interpretation of all results presented here must be in the context of both potential explanations for elevated substitution rates.

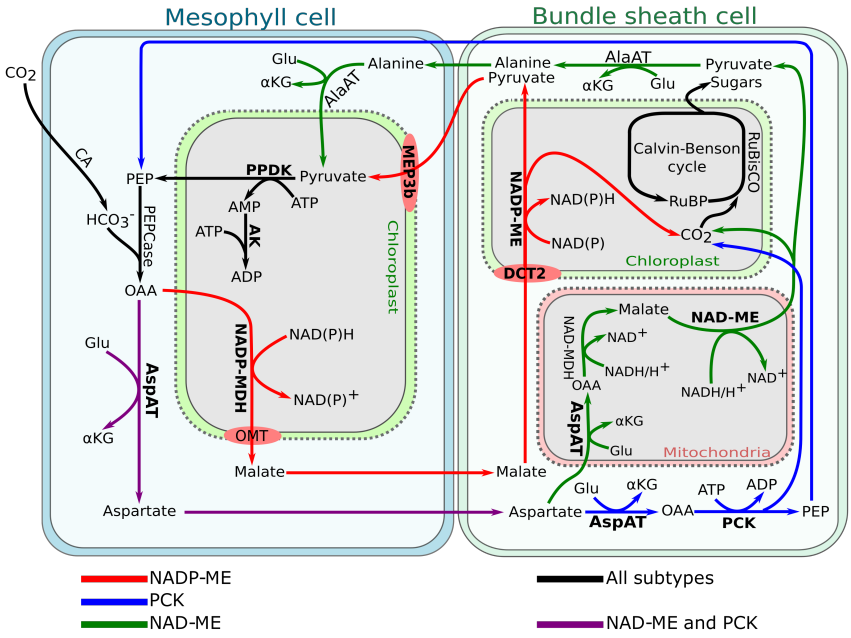

Figure 2: **Simplified pathway representation of the three main C_4 photosynthesis subtypes.** Enzymes where protein sequence evolutionary rates were calculated as part of this study are indicated in bold.

Enzyme name	C_4 subtype	Reference
AK	All subtypes	Kanai and Edwards (1999)
AspAT	NAD-ME & PEPCK	Hatch et al. (1975); Kanai and Edwards (1999)
DCT2	NADP-ME	Huang et al. (2016)
MEP3b	NADP-ME	John et al. (2014)
NAD-ME	NAD-ME	Hatch et al. (1975); Kanai and Edwards (1999); Wang et al. (2011)
NADP-MDH	NADP-ME & PEPCK	Kanai and Edwards (1999)
NADP-ME	NADP-ME	Hatch et al. (1975); Kanai and Edwards (1999)

PCK	PCK	Hatch et al. (1975); Kanai and Edwards (1999)
PPDK	All subtypes	Hatch et al. (1975); Kanai and Edwards (1999); Furbank (2011)
PPDK-RP	All subtypes	

Table 1: **List of enzymes investigated as part of this study, and the set of C_4 subpathways each enzyme was inferred to contribute to based on the literature.**

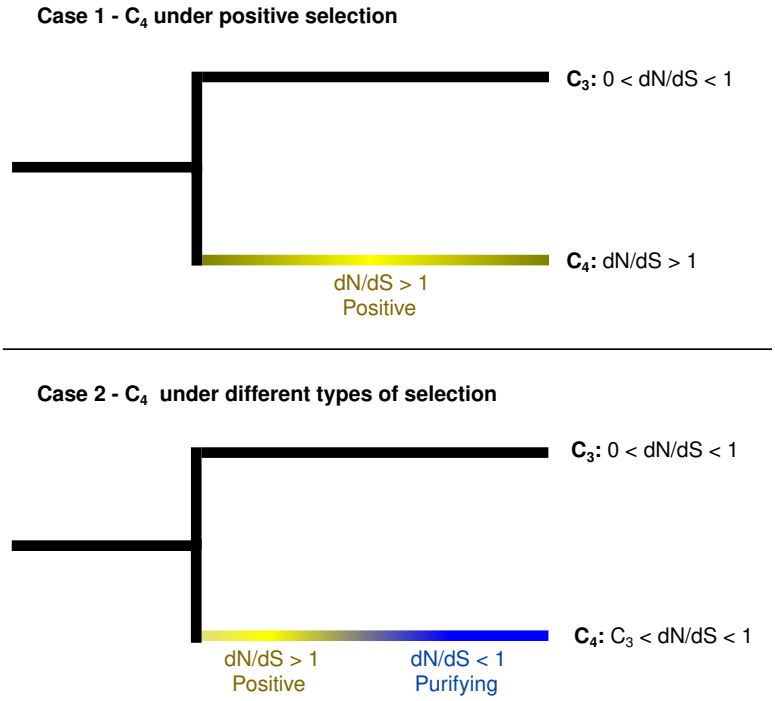

Figure 3: **Models for different histories of selection and the predicted outcomes on dN/dS ratios.** Case 1 shows a classical case of positive selection leading to change in function, while Case 2 shows a case where an enzyme might have gone through a mixture of positive and purifying selection leading to a change in function and a final purifying selection period to maintain the new enzymatic function.

Core C_4 enzymes, those which are utilized by all three C_4 photosynthetic subtypes, showed distinct patterns of change in synonymous/nonsynonymous substitution rates relative to enzymes used in specific subtypes. The analysis of the phylogenetic tree was performed on the branches leading to C_4 species as well as the ancestor branch of the Paniceae, which is the common ancestor of all C_4 species in this study (Figure 1). Both PPDK and PPDK-RP, core enzymes, showed a significantly faster rates of protein sequence evolution in the common ancestor of C_4 species branch relative to C_3. Adenylate kinase (AK) did not show accelerated protein sequence evolution in the ancestral branch, but showed significantly faster rates on the proso and pearl millet branches compared to their C_3 counterparts (Figure S1).

None of the subtype specific enzymes showed significantly higher dN/dS values than the background C_3 genes in the common ancestor of C_4 species branch (Figure S2). Dicarboxylic acid transporter 2 (DCT2), NADP-ME and MEP3b are all employed in the NADP-ME C_4 subtype. Both NADP-ME and MEP3b enzymes showed a significantly faster evolutionary rate in all branches of the NADP-ME subtype (foxtail millet, pearl millet and their ancestral branch) and most other C_4 branches compared to the background C_3 rate. The C_4 branches showing no significant differences in dN/dS relative to the background C_3 rate were: *P. hallii* for the NADP-ME enzyme, and both *P. hallii* and NAD-ME ancestor for MEP3b. DCT2 exhibited a similar pattern to the other two enzymes, with the exception of the branch leading to *S. italica*. An evolutionary pattern shared

among NADP-ME, MEP3b and DCT2 enzymes was a significantly faster evolutionary rate in *U. fusca* and proso millet, which perform PCK and NAD-ME subtypes, respectively, compared to the background C_3 rate. NADP-MDH only showed significantly faster evolutionary rates than C_3 branches in pearl millet, NADP-ME species, and the ancestral branch of NAD-ME subtype. Both AspAT and NAD-ME showed significantly higher dN/dS ratios in branches leading to NAD-ME and PCK subtype species compared with C_3 species (Table 2).

Enzyme name	**Lineages showing accelerated protein sequence evolution**	**Tree branches**
AK	NADP-ME & NAD-ME	pearl millet & proso millet
AspAT	NAD-ME & PCK	proso millet, *P. hallii* & urochloa*
DCT2	NADP-ME, NAD-ME & PCK	ancestral NADP-ME branch, proso millet & urochloa
MEP3b	NADP-ME, NAD-ME & PCK	ancestral C_4 & NADP-ME branches
NAD-ME	NAD-ME & PCK	ancestral NAD-ME branch & urochloa
NADP-MDH	NADP-ME & NAD-ME	pearl millet & ancestral branch of NAD-ME
NADP-ME	NADP-ME, NAD-ME & PCK	all NADP-ME branches, NAD-ME ancestral branch & urochloa
PCK	PCK & NAD-ME	urochloa & proso millet
PPDK	All subtypes	ancestral branch of all C_4
PPDK-RP	All subtypes	ancestral branch of all C_4

Table 2: **List of enzymes employed in this study which showed elevated rates of protein sequence evolution relative to the C_3 background in one or more branches leading to C_4 species.** Cases highlighted in red showed elevated rates of protein sequence evolution in at least one branch which is inconsistent with the canonical assignment of MPC grasses into three clades each utilizing a single distinct C_4 pathway. *this branch was marginally insignificant $p = 0.062$.

In addition to the 10 C_4 photosynthesis related enzymes studied here, the approach used to find genes involved in the C_4 cycle was used to retrieve the sequences of genes encoding two enzymes involved in the photorespiratory pathway: glycolate oxidase (GOX) and serine hydroxymethyltransferase (SHMT). Both enzymes exhibit faster rates of protein sequence change in branches leading to C_4 species than in branches leading to C_3 species. GOX is localized in the peroxisome while SHMT is localized in the mitochondria (Figure 4). The evolutionary pattern of these enzymes were different. GOX showed a faster evolutionary rate than C_3 species branches in the ancestral branch of all Paniceae, the branch leading to *U. fusca*, the ancestral branch of *S. italica* and *P. glaucum* and the branch leading to *S. italica*. On the other hand, SHMT showed fast evolving branches in *U. fusca* branch, the ancestral branch of both *P. miliaceum* genes and one of the *P. miliaceum* genes. In the branch leading to the common ancestor of the MPC clade, GOX, but not SHMT, also showed elevated rates of protein sequence changes relative to branches leading to C_3 species S3).

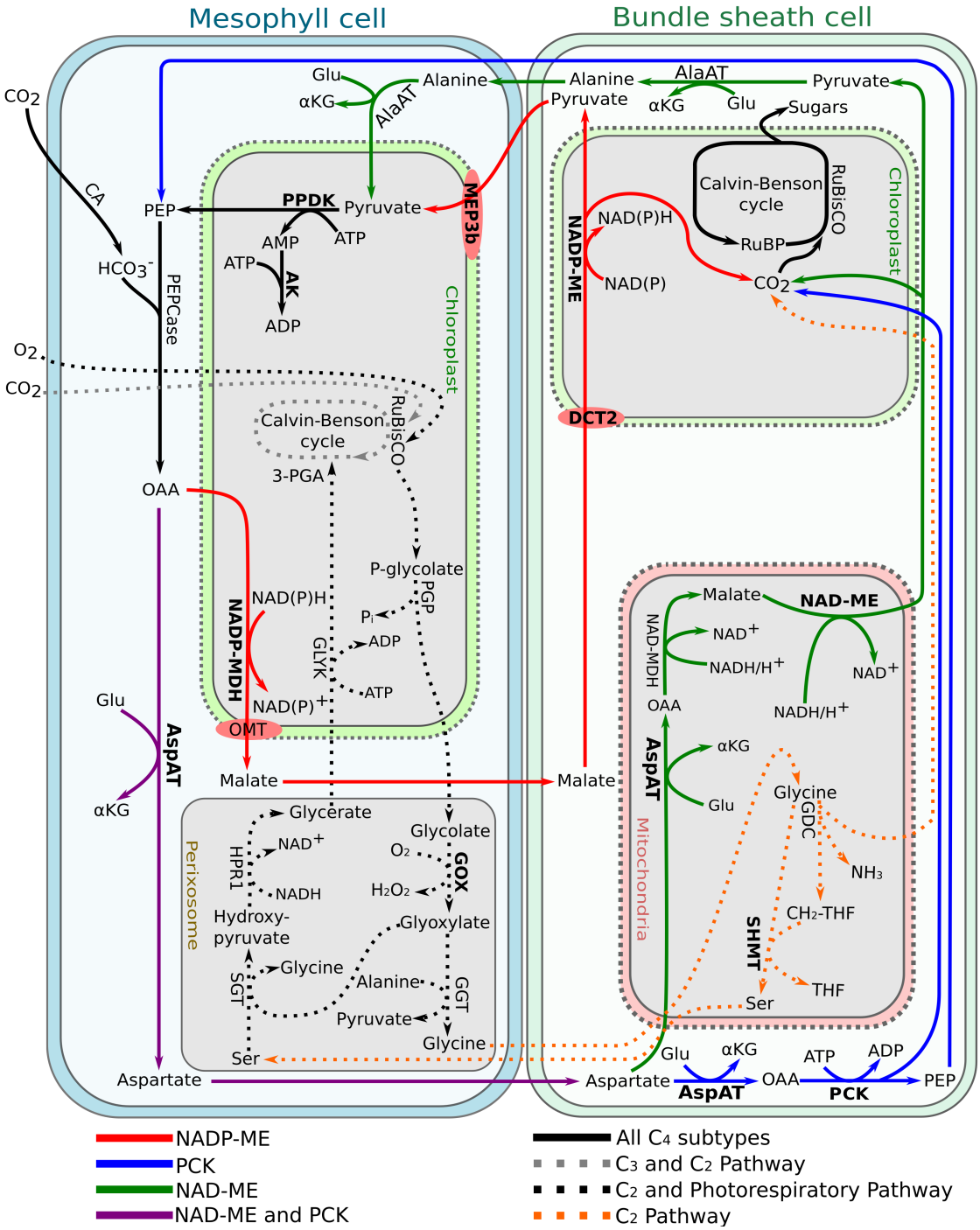

Figure 4: **Simplified pathway representation of the three main C_4 photosynthesis subtypes including the C_3, C_2 and photorespiratory pathways.** Enzymes studied here are represented in bold. Mitochondrial pathway of the C_2 cycle is the same as the mitochondrial photorespiratory cycle. However the mitochondrial pathway occurs in the bundle sheath cell in the C_2 cycle and in the mesophyll cell in the photorespiratory cycle.

Discussion

The emergence of C_4 photosynthesis in more than sixty plant lineages remains a fascinating example of the parallel evolutionary emergence of a complex trait. Reconstruction of protein sequence evolution on ancient branches provides a potential window into the mechanisms plants employed to cope with a comparatively rapid drop in CO_2 concentrations in the atmosphere. Here we used sequence data from eight grass species to examine the patterns of evolution in an ancestral lineage

that split from its closest sequenced C_3 relative 18 million years ago, and gave rise to a diverse set of NADP-ME, NAD-ME and PCK utilizing C_4 species 13 million years ago (Kumar et al., 2017). What happened over that five million year span? Did that ancestral lineage make the leap from C_3 photosynthesis to C_4, or was it a matter of luck or some genetic predisposition that lead to all of its extant descendants utilizing C_4 photosynthesis, admittedly diverse varieties of C_4 photosynthesis today?

If one were to focus only on core enzymes which are both low copy and clearly necessary for all forms of the C_4 photosynthetic pathway, it would appear that this ancient lineage did indeed make the transition from C_3 photosynthesis to C_4 (Figure S1). A focus on enzymes used in some C_4 pathways but not others instead is consistent with the common ancestor of the MPC clade still utilizing C_3 photosynthesis. This model also implies C_4 photosynthesis instead emerged later and independently in the diversification of this clade. The pattern we observed is also inconsistent with a clean separation into NADP-ME, NAD-NE, and PCK utilizing species (Figure S2). In fact only in two cases, AspAT and NAD-ME, where the branches that exhibited accelerated protein sequence evolution relative to C_3 outgroups, are entirely consistent with the primary C_4 pathways utilizing by species descended from those branches (Table 2). This complexity of C_4 pathway utilization suggested by which enzymes show accelerated rates of protein sequence evolution in which lineages is consistent with more recent studies. Multiple reports indicate that many species traditionally thought to employ only a single C_4 pathway may actually fix significant proportions of their total carbon through two or more C_4 pathways (Walker et al., 1997; John et al., 2014; Huang et al., 2016; de Oliveira Dal'Molin et al., 2016; Washburn et al., 2017).

The observation that glycolate oxidase, a critical component of the photorespiratory pathway, also shows accelerated protein sequence evolution in the lineage leading to the common ancestor of the MPC clade (Figure S3) in addition to PPDK and PPDK-RP suggests an intermediate hypothesis. Sometime in the five million year span between the divergence of the MPC lineage from *Dichanthelium oligosanthes* and the most recent common ancestor of the MPC clade, this lineage transitioned from conventional C_3 photosynthesis to the intermediate C_2 photosynthetic cycle where the photorespiratory pathway acts as carbon pump, rather than utilizing any of the three decarboxylation enzymes of classical C_4 photosynthesis (Figure 4). This model is consistent with (Washburn et al., 2015) where ancestral state reconstruction suggested the MPC species utilizing different subtypes of C_4 photosynthesis may have evolved from a non-C_4 ancestor.

It is intriguing to speculate that these ancient changes in protein sequence represents an evolutionary echo of the strong selection acting on plants to adapt to a dramatic decline in atmospheric CO_2 levels. However, it is important to keep in mind the same caveat discussed above: elevated dN/dS ratios which remain below one, even when statistically significant, can be explained by either a mixture of purifying and positive selection or by a simple relaxation of purifying selection. In addition, parallel substitution for the same amino acid substitutions in sister lineages can sometimes lead to a amino acid change being incorrectly inferred to have happened a single time in the common ancestor. This phenomenon has been observed in the study of the parallel evolution of C_4 grasses in the past (Christin et al., 2007, 2008). The findings we present here are suggestive of a C_2 intermediate in the evolution of C_4 photosynthesis in the MPC clade (Sage, 2001; Edwards, 2019), but they do not yet represent conclusive proof of such an ancestor. Given the accelerating pace of plant genome sequencing and improved automated methods for identifying orthologs even in the absence of high quality synteny data, future investigations may incorporate data from larger numbers of C_3 and C_4 grass species, as well as larger numbers of different genes known to be or believed to be involved in C_3, C_2, and/or C_4 photosynthesis.

Data availability

CDS sequences for *Urochloa fusca* are available at Zenodo with the identifier 10.5281/zenodo.3238541.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

DSC, SKKR, YZ and JCS wrote the paper; DSC and JCS designed and conducted the analyses; JCS and YZ collected the data.

Acknowledgements

This study was supported by Science without Borders scholarship (214038/2014-9) to DSC, a Robert B. Daugherty Water for Food Institute research support to JCS and award 2016-67013-24613 from the USDA National Institute of Food and Agriculture to JCS. We thank Jyothi Kumar for critical reading and commentary on earlier versions of this manuscript.

Supplemental Data

Supplementary material:
File S1) Complete set of supplementary figures
File S2) The syntenic and and recriptocal best blast hit inferred orthologs used in this study

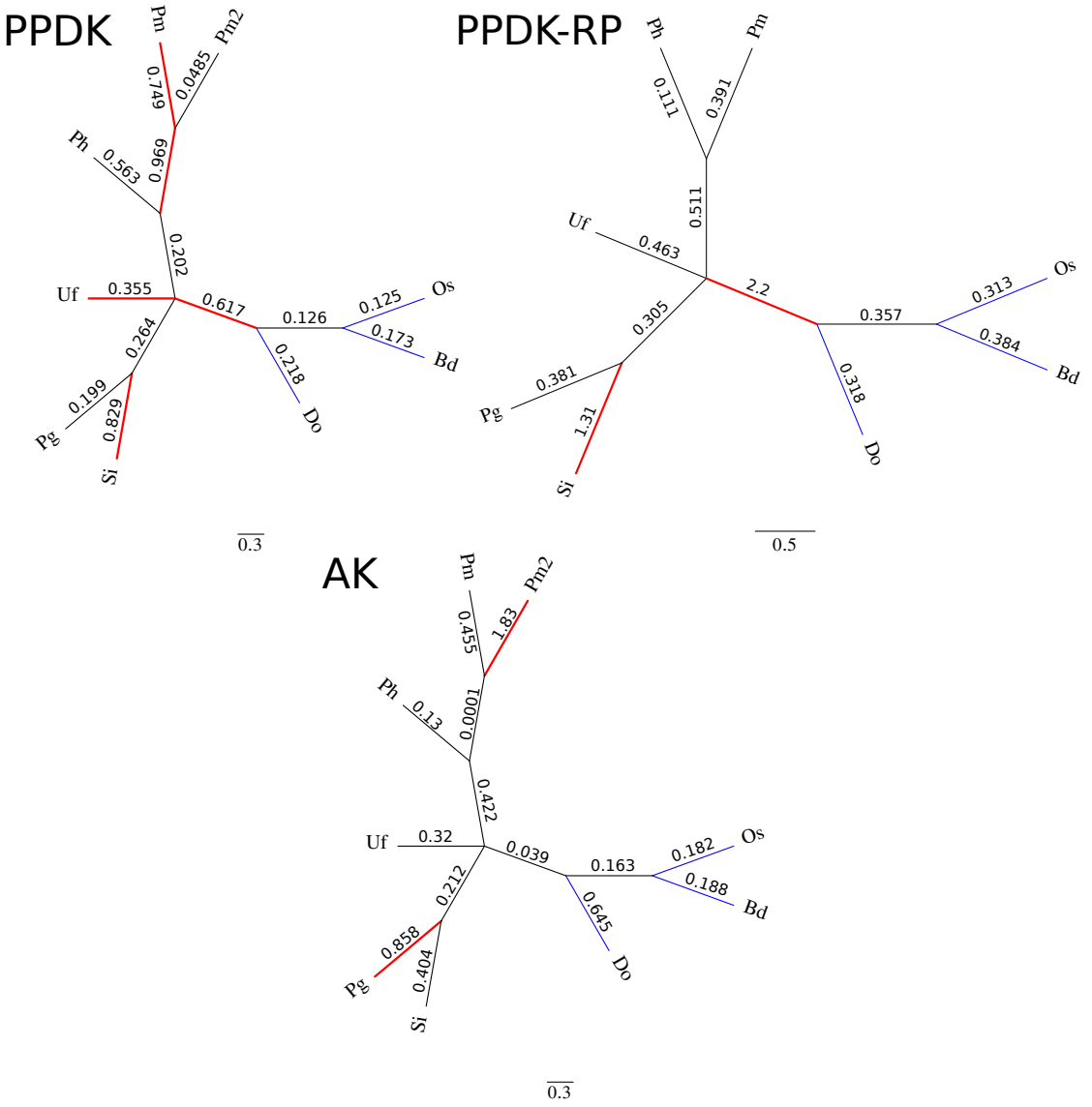

Figure S1: **Unrooted phylogenetic trees of C_4 photosynthesis core enzymes, present in all subtypes, according to citations in Table 1.** Branch lengths are equal. Thick red branches represent branches evolving significantly faster than background C_3 branches in blue. Abbreviations: Os = *Oryza sativa*, Bd = *Brachypodium distachyon*, Do = *Dichanthelium oligosanthes*, Si = *Setaria italica*, Pg = *Pennisetum glaucum*, Uf = *Urochloa fusca*, Ph = *Panicum hallii*, Pm = *Panicum miliaceum*.

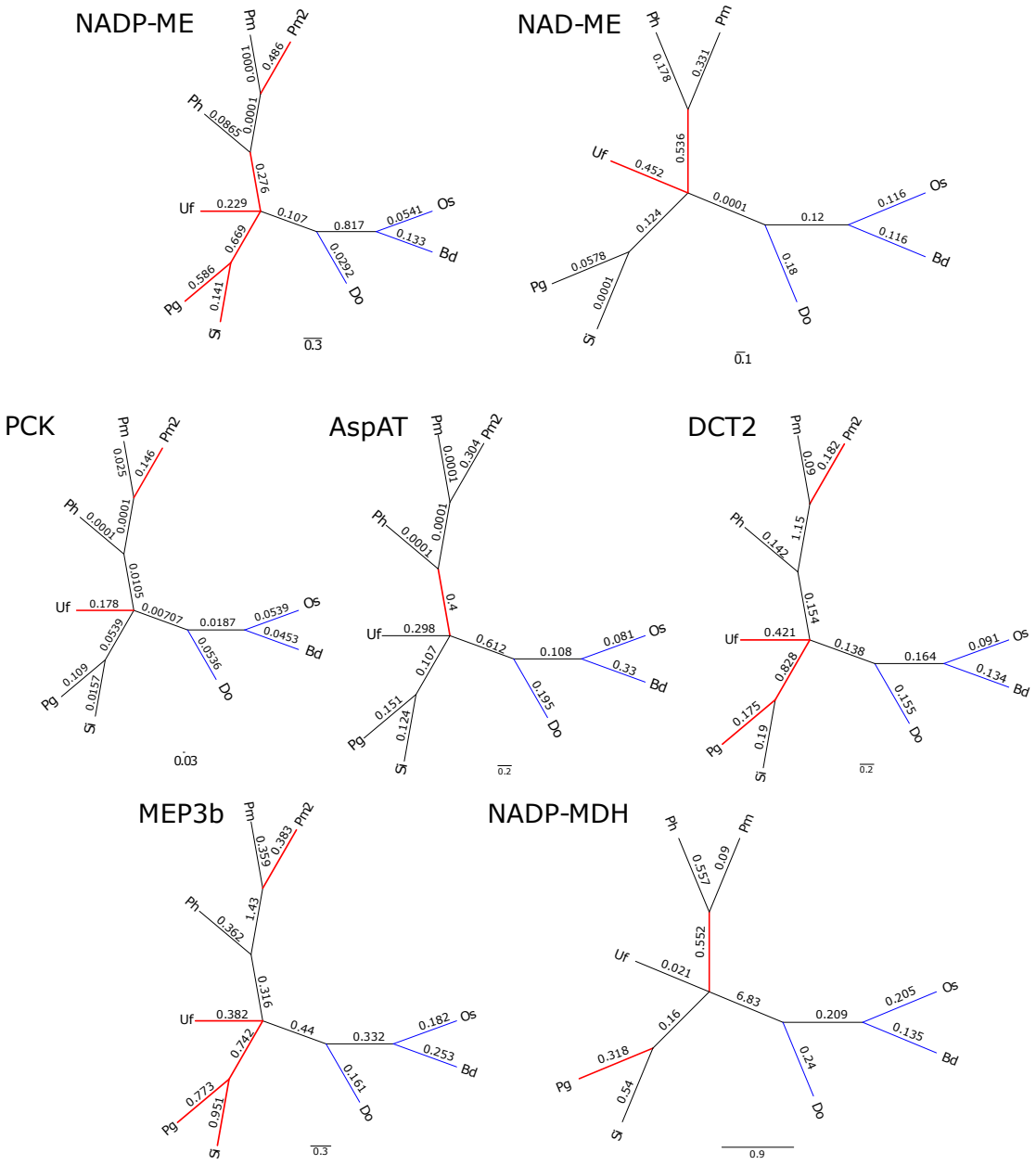

Figure S2: **Unrooted phylogenetic trees of C_4 photosynthesis subtype specific enzymes.** Branch lengths are equal. Thick red branches represent branches evolving significantly faster than background C_3 branches in blue. Abbreviations: Os = *Oryza sativa*, Bd = *Brachypodium distachyon*, Do = *Dichanthelium oligosanthes*, Si = *Setaria italica*, Pg = *Pennisetum glaucum*, Uf = *Urochloa fusca*, Ph = *Panicum hallii*, Pm = *Panicum miliaceum*.

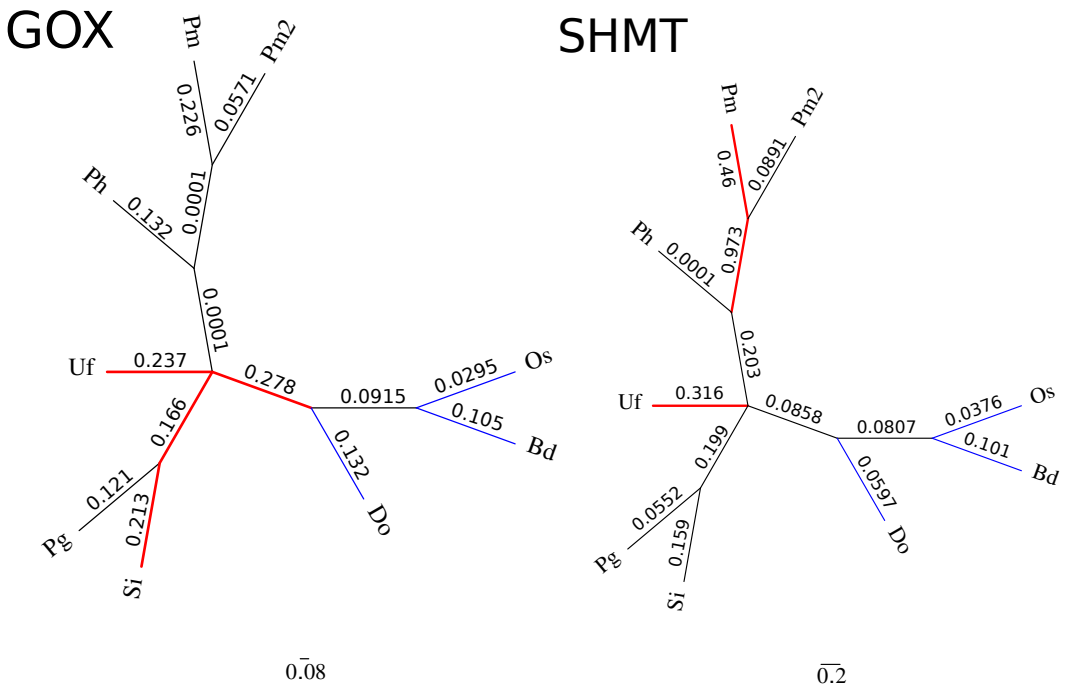

Figure S3: **Unrooted phylogenetic trees of C_4 photosynthesis photorespiratory enzymes.** Branch lengths are equal. Thick red branches represent branches evolving significantly faster than background C_3 branches in blue. Abbreviations: Os = *Oryza sativa*, Bd = *Brachypodium distachyon*, Do = *Dichanthelium oligosanthes*, Si = *Setaria italica*, Pg = *Pennisetum glaucum*, Uf = *Urochloa fusca*, Ph = *Panicum hallii*, Pm = *Panicum miliaceum*.

References

Akhani, H., Ghobadnejhad, M. and Hashemi, S. (2003). Ecology, biogeography and pollen morphology of bienertia cycloptera bunge ex boiss.(chenopodiaceae), an enigmatic c4 plant without kranz anatomy. *Plant Biology* **5**, 167–178.

Bennetzen, J. L., Schmutz, J., Wang, H., Percifield, R., Hawkins, J., Pontaroli, A. C., Estep, M., Feng, L., Vaughn, J. N., Grimwood, J. et al. (2012). Reference genome sequence of the model plant setaria. *Nature biotechnology* **30**, 555.

Bull, T. (1969). Photosynthetic efficiencies and photorespiration in calvin cycle and c4-dicarboxylic acid plants 1. *Crop Science* **9**, 726–729.

Christin, P.-A., Petitpierre, B., Salamin, N., Büchi, L. and Besnard, G. (2008). Evolution of c4 phospho enol pyruvate carboxykinase in grasses, from genotype to phenotype. *Molecular Biology and Evolution* **26**, 357–365.

Christin, P.-A., Salamin, N., Savolainen, V., Duvall, M. R. and Besnard, G. (2007). C4 photosynthesis evolved in grasses via parallel adaptive genetic changes. *Current Biology* **17**, 1241–1247.

de Oliveira Dal'Molin, C. G., Orellana, C., Gebbie, L., Steen, J., Hodson, M. P., Chrysanthopoulos, P., Plan, M. R., McQualter, R., Palfreyman, R. W. and Nielsen, L. K. (2016). Metabolic reconstruction of setaria italica: a systems biology approach for integrating tissue-specific omics and pathway analysis of bioenergy grasses. *Frontiers in plant science* **7**.

Edwards, E. (2019). Evolutionary trajectories, accessibility, and other metaphors: the case of c4 and cam photosynthesis. *The New phytologist* .

Edwards, E., Aliscioni, S., Bell, H. et al. (2011). New grass phylogeny resolves deep evolutionary relationships and discovers c4 origins. *New Phytologist* **193**, 304–312.

Ehleringer, J. R., Cerling, T. E. and Helliker, B. R. (1997). C4 photosynthesis, atmospheric co2, and climate. *Oecologia* **112**, 285–299.

Furbank, R. T. (2011). Evolution of the c4 photosynthetic mechanism: are there really three c4 acid decarboxylation types? *Journal of experimental botany* **62**, 3103–3108.

Gillon, J. and Yakir, D. (2001). Influence of carbonic anhydrase activity in terrestrial vegetation on the 18o content of atmospheric co2. *Science* **291**, 2584–2587.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q. et al. (2011). Trinity: reconstructing a full-length transcriptome without a genome from rna-seq data. *Nature biotechnology* **29**, 644.

Harris, R. S. (2007). *Improved pairwise alignment of genomic DNA*. The Pennsylvania State University.

Hatch, M., Kagawa, T. and Craig, S. (1975). Subdivision of c4-pathway species based on differing c4 acid decarboxylating systems and ultrastructural features. *Functional Plant Biology* **2**, 111–128.

Huang, P., Studer, A. J., Schnable, J. C., Kellogg, E. A. and Brutnell, T. P. (2016). Cross species selection scans identify components of c4 photosynthesis in the grasses. *Journal of Experimental Botany* p. erw256.

Initiative, I. B. et al. (2010). Genome sequencing and analysis of the model grass brachypodium distachyon. *Nature* **463**, 763.

John, C. R., Smith-Unna, R. D., Woodfield, H., Covshoff, S. and Hibberd, J. M. (2014). Evolutionary convergence of cell-specific gene expression in independent lineages of c4 grasses. *Plant physiology* **165**, 62–75.

Kanai, R. and Edwards, G. E. (1999). The biochemistry of c4 photosynthesis. *C4 plant biology* pp. 49–87.

Kawahara, Y., de la Bastide, M., Hamilton, J. P., Kanamori, H., McCombie, W. R., Ouyang, S., Schwartz, D. C., Tanaka, T., Wu, J., Zhou, S. et al. (2013). Improvement of the oryza sativa nipponbare reference genome using next generation sequence and optical map data. *Rice* **6**, 4.

Keys, A. (1986). Rubisco: its role in photorespiration. *Phil. Trans. R. Soc. Lond. B* **313**, 325–336.

Kumar, S., Stecher, G., Suleski, M. and Hedges, S. B. (2017). Timetree: a resource for timelines, timetrees, and divergence times. *Molecular biology and evolution* **34**, 1812–1819.

Lassmann, T. and Sonnhammer, E. L. (2005). Kalign–an accurate and fast multiple sequence alignment algorithm. *BMC bioinformatics* **6**, 298.

Lovell, J. T., Jenkins, J., Lowry, D. B., Mamidi, S., Sreedasyam, A., Weng, X., Barry, K., Bonnette, J., Campitelli, B., Daum, C. et al. (2018). The genomic landscape of molecular responses to natural drought stress in panicum hallii. *Nature communications* **9**, 5213.

Lyons, E. and Freeling, M. (2008). How to usefully compare homologous plant genes and chromosomes as dna sequences. *The Plant Journal* **53**, 661–673.

Mallmann, J., Heckmann, D., Bräutigam, A., Lercher, M. J., Weber, A. P., Westhoff, P. and Gowik, U. (2014). The role of photorespiration during the evolution of c4 photosynthesis in the genus flaveria. *Elife* **3**, e02478.

Morrone, O., Aagesen, L., Scataglini, M. A., Salariato, D. L., Denham, S. S., Chemisquy, M. A., Sede, S. M., Giussani, L. M., Kellogg, E. A. and Zuloaga, F. O. (2012). Phylogeny of the paniceae (poaceae: Panicoideae): integrating plastid dna sequences and morphology into a new classification. *Cladistics* **28**, 333–356.

Offermann, S., Friso, G., Doroshenk, K. A., Sun, Q., Sharpe, R. M., Okita, T. W., Wimmer, D., Edwards, G. E. and van Wijk, K. J. (2015). Developmental and subcellular organization of single-cell c4 photosynthesis in bienertia sinuspersici determined by large-scale proteomics and cdna assembly from 454 dna sequencing. *Journal of proteome research* **14**, 2090–2108.

Sage, R. (2001). Environmental and evolutionary preconditions for the origin and diversification of the c4 photosynthetic syndrome. *Plant Biology* **3**, 202–213.

Sage, R. F., Christin, P.-A. and Edwards, E. J. (2011). The c4 plant lineages of planet earth. *Journal of Experimental Botany* **62**, 3155–3169.

Studer, A. J., Schnable, J. C., Weissmann, S., Kolbe, A. R., McKain, M. R., Shao, Y., Cousins, A. B., Kellogg, E. A. and Brutnell, T. P. (2016). The draft genome of the c 3 panicoid grass species dichanthelium oligosanthes. *Genome biology* **17**, 223.

Tang, H., Lyons, E., Pedersen, B., Schnable, J. C., Paterson, A. H. and Freeling, M. (2011). Screening synteny blocks in pairwise genome comparisons through integer programming. *BMC bioinformatics* **12**, 102.

Tolbert, N. (1997). The c2 oxidative photosynthetic carbon cycle. *Annual review of plant biology* **48**, 1–25.

Varshney, R. K., Shi, C., Thudi, M., Mariac, C., Wallace, J., Qi, P., Zhang, H., Zhao, Y., Wang, X., Rathore, A. et al. (2017). Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. *Nature biotechnology* **35**, 969.

Vicentini, A., Barber, J. C., Aliscioni, S. S., Giussani, L. M. and Kellogg, E. A. (2008). The age of the grasses and clusters of origins of c4 photosynthesis. *Global Change Biology* **14**, 2963–2977.

Walker, R. P., Acheson, R. M., Técsi, L. I. and Leegood, R. C. (1997). Phosphoenolpyruvate carboxykinase in c4 plants: its role and regulation. *Functional Plant Biology* **24**, 459–468.

Wand, S. J., Midgley, G., Jones, M. H., Curtis, P. S. et al. (1999). Responses of wild c4 and c3 grass (poaceae) species to elevated atmospheric co2 concentration: a meta-analytic test of current theories and perceptions. *Global Change Biology* **5**, 723–741.

Wang, L., Peterson, R. B. and Brutnell, T. P. (2011). Regulatory mechanisms underlying c4 photosynthesis. *New Phytologist* **190**, 9–20.

Washburn, J. D., Kothapalli, S. S., Brose, J. M., Covshoff, S., Hibberd, J. M., Conant, G. C. and Pires, J. C. (2017). Ancestral reconstruction and c3 bundle sheath transcript abundance in the paniceae grasses indicate the foundations for all three biochemical c4 subtypes were likely present in the most recent ancestor. *bioRxiv* p. 162644.

Washburn, J. D., Schnable, J. C., Davidse, G. and Pires, J. C. (2015). Phylogeny and photosynthesis of the grass tribe paniceae. *American journal of botany* **102**, 1493–1505.

Yang, Z. (2007). Paml 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution* **24**, 1586–1591.

Zhang, Y., Ding, Z., Ma, F., Chauhan, R. D., Allen, D. K., Brutnell, T. P., Wang, W., Peng, M. and Li, P. (2015). Transcriptional response to petiole heat girdling in cassava. *Scientific reports* **5**, 8414.

Zou, C., Li, L., Miki, D., Li, D., Tang, Q., Xiao, L., Rajput, S., Deng, P., Peng, L., Jia, W. et al. (2019). The genome of broomcorn millet. *Nature Communications* **10**, 436.