





Article

# Genetic Diversity of C<sub>4</sub> Photosynthesis Pathway Genes in *Sorghum bicolor* (L.)

Yongfu Tao <sup>1</sup> , Barbara George-Jaeggli <sup>1,2</sup> , Marie Bouteillé-Pallas <sup>1</sup>, Shuaishuai Tai <sup>3</sup> , Alan Cruickshank <sup>2</sup>, David Jordan <sup>1</sup>  and Emma Mace <sup>1,2,\*</sup>

- <sup>1</sup> Queensland Alliance for Agriculture and Food Innovation (QAAFI), The University of Queensland, Hermitage Research Facility, Warwick, QLD 4370, Australia; y.tao1@uq.edu.au (Y.T.); b.georgejaeggli@uq.edu.au (B.G.-J.); mariebouteille0384@gmail.com (M.B.-P.); david.jordan@uq.edu.au (D.J.)
- <sup>2</sup> Agri-Science Queensland, Department of Agriculture and Fisheries (DAF), Hermitage Research Facility, Warwick, QLD 4370, Australia; Alan.Cruickshank@daf.qld.gov.au
- <sup>3</sup> BGI Genomics, BGI-Shenzhen, Shenzhen 518083, China; shuaishuai.tai@bgi.com
- \* Correspondence: emma.mace@uq.edu.au

Received: 24 June 2020; Accepted: 13 July 2020; Published: 16 July 2020



**Abstract:** C<sub>4</sub> photosynthesis has evolved in over 60 different plant taxa and is an excellent example of convergent evolution. Plants using the C<sub>4</sub> photosynthetic pathway have an efficiency advantage, particularly in hot and dry environments. They account for 23% of global primary production and include some of our most productive cereals. While previous genetic studies comparing phylogenetically related C<sub>3</sub> and C<sub>4</sub> species have elucidated the genetic diversity underpinning the C<sub>4</sub> photosynthetic pathway, no previous studies have described the genetic diversity of the genes involved in this pathway within a C<sub>4</sub> crop species. Enhanced understanding of the allelic diversity and selection signatures of genes in this pathway may present opportunities to improve photosynthetic efficiency, and ultimately yield, by exploiting natural variation. Here, we present the first genetic diversity survey of 8 known C<sub>4</sub> gene families in an important C<sub>4</sub> crop, *Sorghum bicolor* (L.) Moench, using sequence data of 48 genotypes covering wild and domesticated sorghum accessions. Average nucleotide diversity of C<sub>4</sub> gene families varied more than 20-fold from the NADP-malate dehydrogenase (MDH) gene family ( $\theta\pi = 0.2 \times 10^{-3}$ ) to the pyruvate orthophosphate dikinase (PPDK) gene family ( $\theta\pi = 5.21 \times 10^{-3}$ ). Genetic diversity of C<sub>4</sub> genes was reduced by 22.43% in cultivated sorghum compared to wild and weedy sorghum, indicating that the group of wild and weedy sorghum may constitute an untapped reservoir for alleles related to the C<sub>4</sub> photosynthetic pathway. A SNP-level analysis identified purifying selection signals on C<sub>4</sub> PPDK and carbonic anhydrase (CA) genes, and balancing selection signals on C<sub>4</sub> PPDK-regulatory protein (RP) and phosphoenolpyruvate carboxylase (PEPC) genes. Allelic distribution of these C<sub>4</sub> genes was consistent with selection signals detected. A better understanding of the genetic diversity of C<sub>4</sub> pathway in sorghum paves the way for mining the natural allelic variation for the improvement of photosynthesis.

**Keywords:** sorghum; C<sub>4</sub> pathway; genetic diversity; SNPs; domestication

## 1. Introduction

C<sub>4</sub> photosynthesis has independently evolved in more than 60 different plant taxa [1]. The main driver for this convergent evolution is the tendency of Ribulose-1,5-bisphosphate carboxylase (Rubisco), which catalyzes the net fixation of carbon dioxide (CO<sub>2</sub>) to also catalyze an unfavorable oxygenation reaction. This reaction produces toxic phosphoglycolate which has to be converted to useful metabolites requiring substantial metabolic energy [2,3]. This wasteful use of CO<sub>2</sub> is termed photorespiration. Photorespiration becomes a major constraint of photosynthesis in situations where CO<sub>2</sub> to O<sub>2</sub> ratios

are low and temperatures are high. The evolution of C<sub>4</sub> photosynthesis coincided with declining atmospheric CO<sub>2</sub> concentrations [1,4] as a mechanism to avoid photorespiration by concentrating CO<sub>2</sub> in the vicinity of ribulose-1,5-bisphosphate carboxylase (Rubisco).

In the majority of C<sub>4</sub> plants, this is achieved via spatial separation of the initial CO<sub>2</sub> fixation and the Calvin–Benson–Bassham cycle in two different cell types, most often mesophyll cells and bundle sheath cells [5]. CO<sub>2</sub> concentration in C<sub>4</sub> bundle sheath cells is up to 10-fold higher than that found in C<sub>3</sub> mesophyll cells [6]. At higher temperatures, C<sub>4</sub> photosynthesis is not only more efficient compared with C<sub>3</sub> photosynthesis in terms of reducing energy losses from photorespiration, but due to the improved efficiency of this pathway, it renders plants more nitrogen- and water-use efficient [7,8]. C<sub>4</sub> plants are more productive than C<sub>3</sub> plants in areas with high light intensities, warm temperatures, and low rainfall, such as the sub-tropical and tropical areas around the globe.

Many of the major crops that originated from warm and dry regions of the world, such as maize, sorghum, millet, sugarcane, *miscanthus*, and switchgrass, use the C<sub>4</sub> pathway [9]. C<sub>4</sub> crops account for an estimated 23% of global primary production [10]. Improved photosynthetic capacity has been suggested as the next frontier in lifting crop productivity [11]. The C<sub>4</sub> photosynthesis pathway is a good starting point to improve photosynthetic capacity and resource efficiency in crop plants. Attempts are currently being undertaken to integrate characteristics of the C<sub>4</sub> pathway into C<sub>3</sub> crops [7,12–14].

However, possibly due to multiple independent evolutions of C<sub>4</sub> photosynthesis in different plant taxa [1], large variation also exists among C<sub>4</sub> species in terms of the biochemical pathway. It has long been known that three major biochemical subtypes—nicotinamide adenine dinucleotide phosphate-malic enzyme (NADP-ME), nicotinamide adenine dinucleotide-malic enzyme (NAD-ME) and phospho-enol-pyruvate carboxykinase (PCK)—exist among C<sub>4</sub> species [15]. More recently, it has been suggested that mixtures among them exist [16] and that the subtypes vary in their performance under different environmental conditions, e.g., low light [17]. Especially among the grasses, which all of the C<sub>4</sub> cereals belong to, differences in pathway and performance are likely to exist, as C<sub>4</sub> photosynthesis has evolved at least 25 times in this group of plants [18]. Exploring such variation may provide avenues to further improve C<sub>4</sub> photosynthetic efficiency [9].

Sorghum is an NADP-ME subtype C<sub>4</sub> crop well-known for its adaption to drought and high temperatures. It provides staple food for over 500 million people in the semi-arid tropics of Africa and Asia; in addition to being an important source of feed, fiber, and biofuel. Due to these characteristics, it is expected to play an increasingly important role in meeting the challenges of feeding the world's growing population under the threat of global warming. Substantial variation in photosynthesis and related traits has been revealed in sorghum [19–23], indicating existence of genetic variation of underlying genes. However, this variation has not yet been studied.

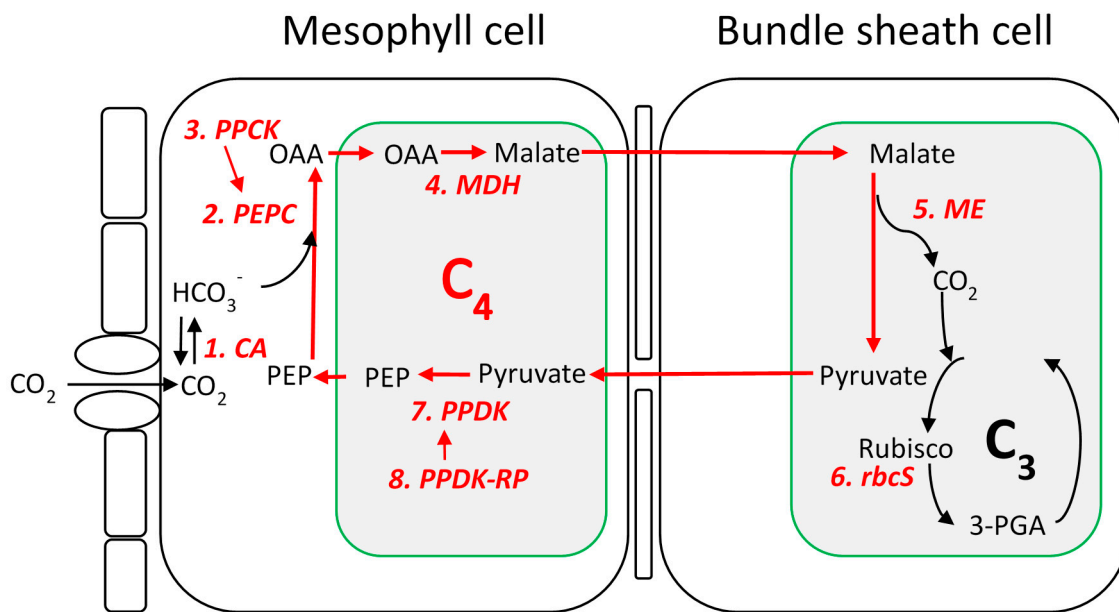
The recent assembly of whole-genome sequences for a wide range of wild and cultivated sorghum species [24–26] provides an excellent opportunity to explore genetic diversity of genes related to the C<sub>4</sub> photosynthetic pathway. Several high-throughput comparative transcriptomics and evolutionary studies using C<sub>3</sub> and C<sub>4</sub> phylogenetically related species and cell-specific gene expression have elucidated the key genes and regulatory networks that underpin the C<sub>4</sub> photosynthetic pathway [5,27–37]. In the present study, we explored the genetic variation in genes that have previously been identified as core C<sub>4</sub> genes, mined their allelic diversity and investigated signatures of selection during domestication in sorghum.

## 2. Materials and Methods

### *Identification of C<sub>4</sub> Gene Families*

This study focuses on 8 key proteins in the NADP-ME photosynthetic pathway in sorghum (Figure 1). A total of 9 genes encoding these proteins with expression and evolutionary evidence supporting their involvement in NADP-ME pathway (hereafter, referred as C<sub>4</sub> genes), and their non-C<sub>4</sub> isoforms in sorghum were extracted from two previous studies [38,39] (Table 1). These non-C<sub>4</sub> isoforms

are homologous of  $C_4$  genes but there was no evidence supporting their involvement in the NADP-ME photosynthetic pathway. Homology between these sorghum  $C_4$  genes and their non- $C_4$  isoforms was further verified via a local blast strategy. Protein sequences of these 9 core  $C_4$  genes were extracted from the sorghum reference genome V3.1 and were blasted against the reference genome. Blast hits of each gene were filtered using the criteria: E-value  $< -10$ , sequence identity  $> 60\%$ , and alignment length  $> 80\%$ . All hits of the same gene satisfying the criteria were plotted based on  $-\log(E\text{-value})$ ; only hits of top  $-\log(E\text{-value})$  class were considered if clear differentiation among them was visualized, otherwise all hits were used.



**Figure 1.** Diagram of the nicotinamide adenine dinucleotide phosphate-malic enzyme (NADP-ME) biosynthetic pathway of  $C_4$  photosynthesis (adapted from [40]). In the mesophyll cells,  $CO_2$  is converted to  $HCO_3^-$  catalyzed by carbonic anhydrase (CA) and fixed into the four-carbon acid, oxaloacetate (OAA), by phosphoenolpyruvate carboxylase (PEPC). Phosphorylation of PEPC is carried out by PEPC kinase (PPCK). The OAA generated by PEPC is then reduced to malate by the NADP-malate dehydrogenase (NADP-MDH) or trans-aminated to aspartate. The resultant  $C_4$  acids, malate and aspartate, are transported to the bundle sheath and then decarboxylated in the vicinity of Rubisco to release  $CO_2$  and pyruvate. Pyruvate is transported back to mesophyll cells to regenerate PEP by pyruvate orthophosphate dikinase (PPDK), while  $CO_2$  enters the Calvin–Benson–Bassham cycle and is fixed by ribulose-1,5-bisphosphate carboxylase (Rubisco). Activation and inactivation of PPDK is catalyzed by PPDK regulatory protein (PPDK-RP).

**Table 1.** Single nucleotide polymorphism (SNP) information and selection signals across 27 genes from C<sub>4</sub> gene families.

Gene ID	Enzyme	GL	CDSL	NoS	NoSiC	NoNS	NoSS	UPSGL	UBSGL	NoSUPS	NoNSUPS	NoSUBS	NoNSUBS
Sobic.002G230100	CA	4823	1014	115	14	4	10	No	No	0	0	1	0
<b>Sobic.003G234200</b>	CA	10440	1371	475	33	7	26	No	No	1	1	0	0
<b>Sobic.003G234400</b>	CA	4749	615	138	13	3	10	No	No	0	0	0	0
Sobic.003G234500	CA	2986	609	173	11	5	6	No	No	0	0	0	0
Sobic.003G234600	CA	4750	771	210	18	10	8	No	No	0	0	0	0
Sobic.007G166200	NADP-MDH	3354	1308	53	11	6	5	No	No	0	0	0	0
<b>Sobic.007G166300</b>	NADP-MDH	3816	1290	108	12	3	9	No	No	0	0	0	0
Sobic.003G036000	NADP-ME	6107	1941	111	11	4	7	No	No	0	0	0	0
<b>Sobic.003G036200</b>	NADP-ME	5447	1911	141	12	3	9	No	No	0	0	0	0
Sobic.003G280900	NADP-ME	5691	1782	175	22	13	9	No	No	1	1	0	0
Sobic.003G292400	NADP-ME	4527	1782	95	22	8	14	No	No	10	2	0	0
Sobic.009G069600	NADP-ME	3624	1713	118	34	10	24	No	No	3	1	0	0
Sobic.002G167000	PEPC	5632	2904	41	11	6	5	No	No	0	0	0	0
Sobic.003G100600	PEPC	8881	3117	371	43	9	34	No	No	0	0	21	2
Sobic.003G301800	PEPC	7610	2901	138	19	3	17	No	No	0	0	0	0
Sobic.004G106900	PEPC	6977	2883	146	34	5	29	No	No	0	0	7	0
Sobic.007G106500	PEPC	5616	2895	64	12	8	4	No	No	1	1	0	0
<b>Sobic.010G160700</b>	PEPC	6647	3087	193	28	9	19	No	No	0	0	2	0
Sobic.004G219900	PPCK	1612	924	40	9	1	8	No	No	0	0	2	0
<b>Sobic.004G338000</b>	PPCK	1749	855	37	9	4	4	No	No	0	0	0	0
Sobic.006G148300	PPCK	1997	900	64	4	1	3	No	No	0	0	0	0
Sobic.001G326900	PPDK	8494	2730	321	46	18	28	No	Yes	0	0	24	5
<b>Sobic.009G132900</b>	PPDK	12748	2847	441	16	0	16	No	No	3	0	0	0
<b>Sobic.002G324400</b>	PPDK-RP	2507	1290	79	22	8	14	No	No	0	0	3	0
Sobic.002G324500	PPDK-RP	3072	1260	69	20	5	15	No	No	4	0	0	0
Sobic.002G324700	PPDK-RP	4662	1587	222	28	19	9	No	No	1	1	2	2
<b>Sobic.005G042000</b>	RbcS	1556	510	45	7	4	3	No	No	0	0	0	0

Gene ID is according to sorghum reference genome V3.1. Gene IDs in bold indicate their C<sub>4</sub> genes. Enzyme: Encoded enzyme. GL: Gene length. CDSL: Length of coding sequence (CDS). NoS: Total number of SNPs identified across the gene. NoSiC: Number of SNPs identified in CDS. NoNS: Number of non-synonymous SNPs. NoSS: Number of synonymous SNPs. UPSGL: Under purifying selection based on gene level analysis. UBSGL: Under balancing selection based on gene level analysis. NoSUPS: Number of SNPs under purifying selection. NoNSUPS: Number of non-synonymous SNPs under purifying selection. NoSUBS: Number of SNPs under balancing selection. NoNSUBS: Number of non-synonymous SNPs under balancing selection.

### 3. Plant Material and Genomic Data

Sequence data of the identified  $C_4$  genes were extracted from 48 accessions of *Sorghum bicolor* with high mapping depth (~22× per accession, ranging from 16 to 45×) reported in previous studies [24–26]. These 48 accessions represent all major cultivated sorghum races and some wild progenitors (Table S1).

### 4. Gene-Level Population Genetic Analyses

Population genetic parameters including nucleotide diversity ( $\theta\pi$ ) [41], Tajima's D [42], and Watterson's Estimator (hW) [43] were directly calculated for each of the 27 genes using the Bio::PopGen::Statistics module.  $F_{ST}$  [44], which measures population differentiation, was also calculated for each of the 27 genes using the Bio::PopGen::PopStats module [26]. The Bio::PopGen::IO module was used to read input file, which was prepared using an in-house Perl script for calculation of these population genetic parameters.

The criteria used in Mace et al. (2013) were employed to identify genes under purifying selection and balancing selection, respectively. Criteria for purifying selection included: (1)  $\theta\pi$  and hW < 5% of the empirical distribution in the cultivated group, (2)  $F_{ST}$  between the group of cultivated sorghum and the group of wild and weedy sorghum > 95% of the population pairwise distribution, (3) Tajima's D < 0. Criteria for balancing selection included: (1)  $\theta\pi$  and hW > 25% of the empirical distribution in the cultivated group, (2)  $F_{ST}$  between the group of cultivated sorghum and the group of wild and weedy sorghum < 90% of the population pairwise distribution, (3) Tajima's D > 5% of the empirical distribution.

### 5. SNP-Level Identification of Selection Signature

Population genetics parameters including  $\theta\pi$ , Tajima's D, and  $F_{ST}$  between the group of cultivated sorghum and the group of wild and weedy sorghum were computed for these 27 genes using CDS sequence in PopGenome, a population genomics package implemented in the R environment (<http://cran.r-project.org/>) [45]. Specifically, commands diversity.stats, F\_ST.stats, and neutrality.stats were called to calculate  $\theta\pi$ ,  $F_{ST}$ , and Tajima's D for each single nucleotide polymorphism (SNP), respectively, with a slide window of 1-bp and 1-bp step size. Functional annotation of each SNP was conducted using get.codons command. Fold decrease of  $\theta\pi$  in the cultivated sorghum group compared to the group of wild and weedy sorghum was calculated to represent reduction of diversity (RoD). The following criteria were adopted to identify sites with signature of purifying selection: (1) A RoD greater than the average of neutral genes; (2)  $F_{ST} > 0$ ; (3) Tajima's D < 0. The following criteria were adopted to identify sites with signature of balancing selection: (1) An increase in diversity (IoD) in the cultivated group and the group of wild and weedy comparison; (2)  $F_{ST} > 0$ ; (3) Tajima's D > 0.

### 6. Phylogenetic and Haplotype Analysis

A phylogenetic tree was constructed based on CDS of all 27 genes from  $C_4$  gene families using the neighbor-joining method with default settings (bootstrapped 100 times; support threshold, 50%) in Geneious 8.1.2 (<https://www.geneious.com/>, Biomatters Ltd., Auckland, New Zealand). Analysis of haplotype network was conducted using a combination of the R package ape [46] and pegas [47]. All 48 sorghum accessions were classified into four groups: Cultivated, wild and weedy, *Guinea margaritifera* and *S. propinquum* (Table S2).

## 7. Results

#### *Nucleotide Diversity of Core $C_4$ Gene Families in Sorghum*

Based on 9 genes corresponding to 8 core  $C_4$  enzymes in sorghum, 18 homologous genes were identified across the sorghum genome. In total, 5 CA genes, 2 NADP-MDH genes, 5 NADP-ME genes, 6 PEPC genes, 3 PPKK genes, 2 PPKK genes, 3 PPKK-RP genes, and 1 rbcS gene were identified

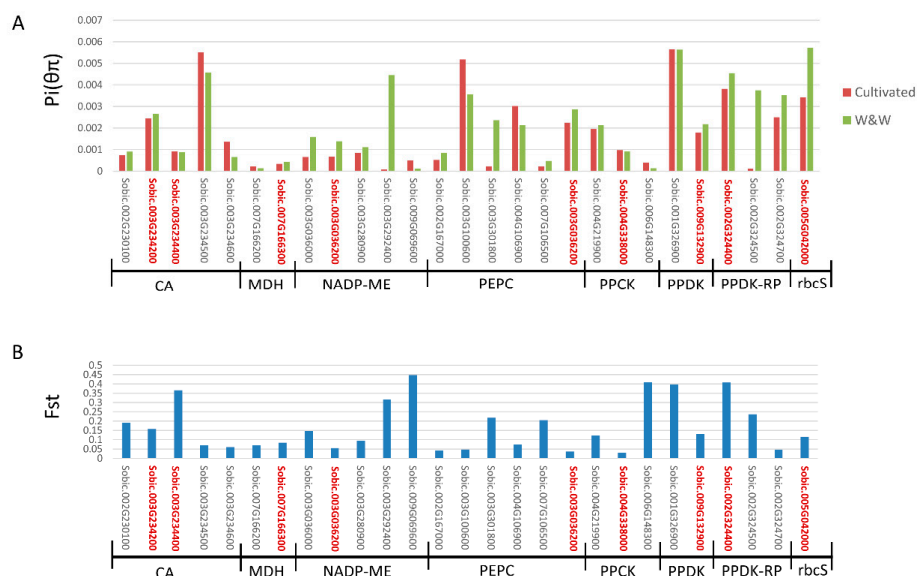
(Table 1). Nucleotide diversity ( $\theta\pi$ ) of these 27 genes was investigated using sequence data of 48 genotypes covering wild and weedy, and cultivated sorghum (Mace et al., 2013). A total number of 4183 single nucleotide polymorphisms (SNPs) were identified in these 27 genes with 521 SNPs located in coding sequence (CDS) regions (Table 1). These  $C_4$  gene families displayed an average overall nucleotide diversity of  $\theta\pi = 2.09 \times 10^{-3}$ , which is comparable to that of 130 housekeeping genes ( $\theta\pi = 1.97 \times 10^{-3}$ , Mace et al., 2013) (*t*-test, *p*-value > 0.05). Nucleotide diversity varied dramatically among the  $C_4$  gene families, with the NADP-MDH genes displaying the lowest levels of diversity across all genotypes (average  $\theta\pi = 0.25 \times 10^{-3}$ ), followed by NADP-ME genes ( $\theta\pi = 0.93 \times 10^{-3}$ ), PPCK genes ( $\theta\pi = 1.20 \times 10^{-3}$ ), PEPC genes ( $\theta\pi = 2.11 \times 10^{-3}$ ), CA ( $\theta\pi = 2.26 \times 10^{-3}$ ), and PPDK-RP ( $\theta\pi = 2.96 \times 10^{-3}$ ), while PPDK genes showed the highest level of diversity ( $\theta\pi = 5.21 \times 10^{-3}$ ) (Table 2, Figure 2A). The only gene encoding ribulose biphosphate carboxylase/oxygenase small-subunit (*rbcS*), *Sobic.005G042000*, had relatively high genetic diversity among  $C_4$  gene families with  $\theta\pi = 4.32 \times 10^{-3}$  across all 48 genotypes,  $5.72 \times 10^{-3}$  in the wild and weedy group, and  $3.03 \times 10^{-3}$  in the cultivated group.

**Table 2.** Genetic diversity ( $\theta\pi$ ) and fixation index ( $F_{ST}$ ) of 27 genes from  $C_4$  gene families.

GeneID	Enzyme	$\theta\pi$ -All	$\theta\pi$ -Cultivated	$\theta\pi$ -W&W	$F_{ST}$
Sobic.002G230100	CA	0.80	0.74	0.90	0.19
<b>Sobic.003G234200</b>	CA	2.65	2.46	2.66	0.16
<b>Sobic.003G234400</b>	CA	1.01	0.91	0.88	0.37
Sobic.003G234500	CA	5.55	5.51	4.56	0.07
Sobic.003G234600	CA	1.27	1.35	0.65	0.06
Sobic.007G166200	NADP-MDH	0.18	0.21	0.13	0.07
<b>Sobic.007G166300</b>	NADP-MDH	0.33	0.33	0.42	0.08
Sobic.003G036000	NADP-ME	0.88	0.65	1.59	0.15
<b>Sobic.003G036200</b>	NADP-ME	0.89	0.67	1.39	0.06
Sobic.003G280900	NADP-ME	0.93	0.85	1.11	0.09
Sobic.003G292400	NADP-ME	1.43	0.08	4.44	0.32
Sobic.009G069600	NADP-ME	0.52	0.49	0.10	0.45
Sobic.002G167000	PEPC	0.58	0.51	0.85	0.04
Sobic.003G100600	PEPC	5.36	5.18	3.56	0.05
Sobic.003G301800	PEPC	0.64	0.22	2.37	0.22
Sobic.004G106900	PEPC	3.18	3.02	2.14	0.07
Sobic.007G106500	PEPC	0.44	0.22	0.47	0.21
<b>Sobic.010G160700</b>	PEPC	2.49	2.25	2.86	0.04
Sobic.004G219900	PPCK	2.08	1.94	2.12	0.12
<b>Sobic.004G338000</b>	PPCK	1.03	0.96	0.91	0.03
Sobic.006G148300	PPCK	0.48	0.39	0.13	0.41
Sobic.001G326900	PPDK	8.34	5.64	5.64	0.40
<b>Sobic.009G132900</b>	PPDK	2.07	1.79	2.19	0.13
<b>Sobic.002G324400</b>	PPDK-RP	5.04	3.82	4.55	0.41
Sobic.002G324500	PPDK-RP	1.27	0.10	3.75	0.24
Sobic.002G324700	PPDK-RP	2.58	2.50	3.51	0.05
<b>Sobic.005G042000</b>	<i>rbcS</i>	4.32	3.41	5.72	0.12

Gene ID is according to sorghum reference genome V3.1. Gene IDs in bold indicate the  $C_4$  gene versions. Enzyme: Encoded enzyme.  $\theta\pi$ -All: Nucleotide diversity across all 48 genotypes.  $\theta\pi$ -Cultivated: Nucleotide diversity across cultivated genotypes.  $\theta\pi$ -W&W: Nucleotide diversity across wild and weedy genotypes. All  $\theta\pi$  values are in unites of per kb.  $F_{ST}$ : Fixation index between cultivated genotypes and wild and weedy genotypes.





**Figure 2.** Genetic diversity and fixation index ( $F_{ST}$ ) of  $C_4$  gene families between cultivated sorghum and the wild and weedy group. **(A)** Genetic diversity ( $\pi$ ) for each of the  $C_4$  gene families. Gene IDs in red indicate core  $C_4$  genes. Red bars represent the  $\pi$  of cultivated sorghum, while dark blue bars represent the  $\pi$  of wild and weedy. **(B)**  $F_{ST}$  between cultivated and wild and weedy of each of  $C_4$  gene families. Gene IDs in red indicate core  $C_4$  genes.

Mixed trends were found when comparing  $C_4$  genes with non- $C_4$  isoforms in each gene family with the average overall genetic diversity of  $C_4$  genes being comparable to that of their non- $C_4$  counterpart (Table 2). The  $C_4$  PPDk-RP gene (Sobic.007G166300) and  $C_4$  NADP-MDH gene (Sobic.002G324400) had an overall  $\theta\pi$  which was 161.76% and 79.85% higher than their non- $C_4$  isoforms, respectively, whereas the  $\theta\pi$  of the  $C_4$  PPDk gene (Sobic.009G132900) was 75.16% lower than that of the non- $C_4$  PPDk isoform. Nucleotide diversity of  $C_4$  genes in the other gene families was within the range of variation of their non- $C_4$  isoforms.

Genetic diversity across  $C_4$  gene families was significantly reduced during sorghum domestication (paired  $t$ -test,  $p$ -value < 0.05). Averaged across all  $C_4$  gene families genetic diversity was reduced by 22.44% in the domesticated compared with the wild and weedy group and when just the 9 core  $C_4$  genes were considered, the reduction was 22.98%. However, the reduction of genetic diversity during domestication in  $C_4$  genes was not significantly different from that in housekeeping genes (Table S2) ( $t$ -test,  $p$ -value > 0.05). Among the 27 genes, Sobic.003G292400, a non- $C_4$  NADP-ME isoform, exhibited the most severe reduction in genetic diversity, with a reduction of 98.23%. The  $C_4$  version of that gene, the NADP-ME gene (Sobic.003G036200), showed the greatest loss of genetic diversity (51.89%) among the  $C_4$  genes, with an  $F_{ST}$  between the cultivated and wild and weedy groups of 0.06 (Figure 2B). In contrast, another non- $C_4$  isoform of NADP-ME (Sobic.009G069600), a non- $C_4$  isoform of PPCK (Sobic.006G148300), and a non- $C_4$  CA isoform (Sobic.003G234600) showed a more than 2-fold increase in genetic diversity in the cultivated group.

## 8. Identification of Selection Signals during Domestication across the 27 Genes

The selection signature of these  $C_4$  gene families was firstly investigated at the gene level. Based on thresholds of genome-wide rankings described in Mace et al. (2013), only one gene (Sobic.001G326900, non- $C_4$  PPDk isoform) was identified as being under balancing selection, which maintains diversity of selected genes, during sorghum domestication, while no gene was identified as being under purifying selection, which reduces diversity of selected genes (Table 1). Subsequent to this, a higher resolution detection of selection signature was conducted at the SNP level using the CDS of the 27 genes. Among 521 SNPs across 27 CDS, 176 were non-synonymous.

The number of non-synonymous SNPs within genes varied from 19 in the non-C<sub>4</sub> PPDK-RP isoform (Sobic.002G324700) to 0 in the C<sub>4</sub> PPDK (Sobic.009G132900). The C<sub>4</sub> PEPC gene (Sobic.010G160700) had the highest number of non-synonymous SNPs (9) among the 9 C<sub>4</sub> genes (Table 1). In contrast to the gene-level analysis, SNP-level analysis identified 24 SNPs across 8 genes under purifying selection, including 7 non-synonymous SNPs in 6 genes (Table S3). Genes with SNPs under purifying selection included two C<sub>4</sub> isoforms, PPDK (Sobic.009G132900) and CA (Sobic.003G234200), three of 4 non-C<sub>4</sub> NADP-ME (Sobic.003G280900, Sobic.003G292400, Sobic.009G069600), both two non-C<sub>4</sub> PPDK-RP (Sobic.002G324500, Sobic.002G324700), and a non-C<sub>4</sub> PEPC gene (Sobic.007G106500). Among the 2 C<sub>4</sub> genes with SNPs under selection, Sobic.009G132900 had 3 synonymous SNPs under purifying selection, while Sobic.003G234200 had a non-synonymous SNP under purifying selection.

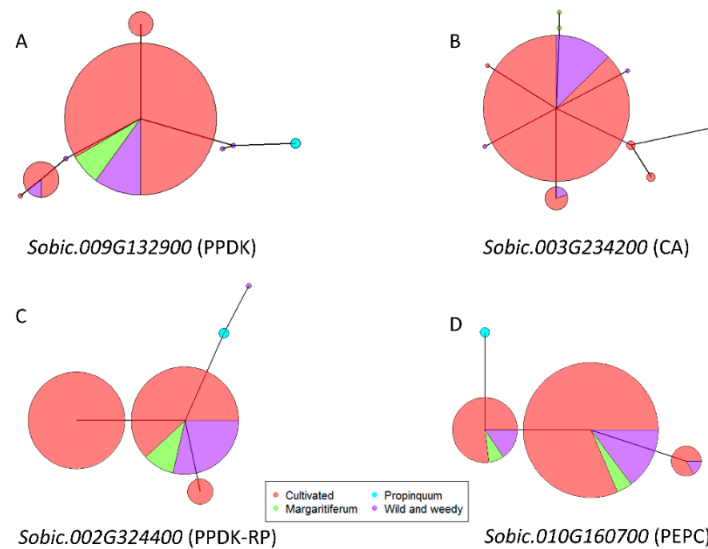
A total of 60 SNPs across 8 genes were identified as being under balancing selection, 7 of which were non-synonymous SNPs distributed across 2 genes (Table S4). The non-C<sub>4</sub> PPDK (Sobic.001G326900) had 24 SNPs under balancing selection including 5 non-synonymous SNPs, and additionally had an overall gene-level signature of balancing selection based on the previous analysis. Two C<sub>4</sub> isoforms, PPDK-RP (Sobic.002G324400) and PEPC (Sobic.010G160700), were identified with 3 and 2 SNPs under balancing selection, respectively, although none of them were non-synonymous SNPs. Two non-C<sub>4</sub> PEPC (Sobic.003G100600, Sobic.004G106900) were identified with SNPs under balancing selection, with Sobic.003G100600 having 21 SNPs including 2 non-synonymous SNPs exhibiting signatures of balancing selection. The other 2 genes with SNPs under balancing selection were a non-C<sub>4</sub> CA isoform, Sobic.002G230100, and a non-C<sub>4</sub> PPCK isoform, Sobic.004G219900.

## 9. Allelic Variation of Core C<sub>4</sub> Genes under Selection in Sorghum

A phylogenetic tree was constructed using the CDS of these 27 genes to depict the genetic relationship of 48 accessions (Figure S1). The inter- and intra-species distribution of private haplotypes of each gene is detailed in Table S5, with the majority (~90%) of the genes with private inter-species haplotypes from *S. propinquum*, e.g., 4 unique haplotypes were observed for the C<sub>4</sub> isoform of PEPC, with the 2 *S. propinquum* accessions sharing a single private haplotype. To investigate allelic variation of 4 core C<sub>4</sub> genes with SNPs under selection in sorghum, haplotype networks were constructed using CDS SNPs. Based on 16 SNPs within the CDS of the PPDK gene (Sobic.009G132900), 8 haplotypes were identified. Five haplotypes were identified in the wild and weedy genotypes, with 3 being private haplotypes and two of them being maintained in cultivated sorghum; two new haplotypes arose in cultivated sorghum after domestication (Figure 3A). Ten haplotypes of one CA gene (Sobic.003G234200) were revealed using 33 SNPs, with 4 distinct haplotypes being characterized by the wild and weedy genotypes. Two of the wild and weedy haplotypes were maintained in cultivated sorghum during domestication, with three new haplotypes arising after domestication (Figure 3B). The loss of wild and weedy haplotypes in cultivated sorghum in these two genes was consistent with the finding that they were under purifying selection.

The PPDK-RP gene (Sobic.002G324400) had 22 SNPs in the CDS, based on which 5 haplotypes were identified. Two haplotypes were characterized by the wild and weedy genotypes, with the main wild haplotype maintained and further diversifying into two new haplotypes in the cultivated group (Figure 3C). Based on 28 SNPs in the CDS of the C<sub>4</sub> PEPC gene (Sobic.010G160700), 4 haplotypes were identified. Wild and weedy genotypes encompassed 3 haplotypes and all of them were maintained in cultivated sorghum (Figure 3D). *S. propinquum* had unique haplotypes across all 4 genes, while the *Sorghum bicolor* race *guinea margaritifera* shared haplotypes with the wild and weedy genotypes in most cases, indicating a closer relationship with the wild and weedy group.





**Figure 3.** Haplotype network of 4 core  $C_4$  gene with selection signal based on individual SNP analysis. (A) The PPDK gene (Sobic.009G132900) with signal of purifying selection; (B) one of the CA genes (Sobic.003G234200) with signal of purifying selection; (C) the PPDK-RP gene (Sobic.002G324400) with signal of balancing selection; (D) the PEPC gene (Sobic.010G160700) with signal of balancing selection. Group classification of sorghum accessions used as detailed in Table S1. Color-coding as follows; cultivated sorghum (red), wild and weedy genotypes (purple), *Sorghum propinquum* (blue), and *Sorghum guinea margaritifera* (green). The size of the circles in the haplotype networks is proportionate to the number of accessions with that haplotype. The branch length represents the genetic distance between two haplotypes.

## 10. Discussion

The evolution of  $C_4$  photosynthesis has been studied extensively at the cross-species level with signals of adaptive evolution identified on key genes in the  $C_4$  pathway [28,34,48–50]. As the evolution of  $C_4$  photosynthesis is driven by environments characterized by low  $CO_2$  availability, such as hot and dry environments in which  $CO_2$  uptake is limited by stomatal closure, it is likely that within-species adaptive variation also exists. However, to our knowledge, studies of within-species allele diversity and signatures of selection on key genes in the  $C_4$  pathway have not previously been undertaken.

Knowledge of existing natural variation and levels of genetic diversity is a pre-requisite for the optimization of  $C_4$  photosynthesis. In this study, we performed the first investigation of the genetic diversity of  $C_4$  gene families within a  $C_4$  species using a collection of 48 sorghum lines. We focused on 9  $C_4$  genes due to their reported key roles in  $C_4$  photosynthesis. Our collection of sorghum represents all major cultivated sorghum races, landraces, and wild progenitors, and captures a good proportion of genetic diversity within sorghum. Substantial variation of nucleotide diversity was observed among these 8  $C_4$  gene families in sorghum, with the NADP-MDH gene family showing the least diversity and the PPDK gene family showing the greatest diversity. Nine core  $C_4$  genes also exhibited varying degrees of genetic diversity, ranging from  $\theta\pi$  values of  $5.04 \times 10^{-3}$  and  $4.32 \times 10^{-3}$  in PPDK-RP and *rbcS* to  $\theta\pi$  values of  $0.33 \times 10^{-3}$  and  $0.67 \times 10^{-3}$  in NADP-MDH and NADP-ME. However, despite such low levels of diversity, non-synonymous SNPs were identified in both NADP-MDH and NADP-ME (Table 1).  $C_4$  PPDK was the only gene which did not contain a non-synonymous SNP, despite its fairly large size (gene size, 12748bp; CDS, 2847bp), indicating the function of this gene is highly conserved.

Cultivated sorghum was domesticated more than five thousand years ago in Africa [51–53]. This artificial selection process has morphologically and physiologically reshaped sorghum to better suit human needs, and also resulted in substantial reduction of genetic diversity genome wide in cultivated sorghum compared with wild and weedy types [26,54,55]. In this study, reduction of genetic diversity during sorghum domestication was also observed in the  $C_4$  gene families, indicating

that wild sorghum, as a repository for genetic diversity, might harbor alleles useful for improving C<sub>4</sub> photosynthesis.

However, the overall reduction in diversity of C<sub>4</sub> gene families was not significantly different from the genome-wide average, indicating that this gene family has not been under particularly strong selection pressure. Similarly, none of the 9 core C<sub>4</sub> genes showed a domestication signal at the gene level. The absence of large sequence variation at the gene level is also consistent with previous evolutionary studies suggesting that relatively minor changes to pre-existing regulatory networks and the use of pre-existing cis-elements were often sufficient to recruit genes into the C<sub>4</sub> pathway [56,57]. The C<sub>4</sub> isoform of the NADP-ME gene found in maize and sorghum is one such gene that has been found to be activated for C<sub>4</sub> photosynthesis via subtle changes to its promoter, while the rest of the gene is highly conserved [33]. This is consistent with the low diversity in this gene family observed in our study.

A further high-resolution investigation of domestication signature at the SNP level revealed 2 C<sub>4</sub> genes, PPDK (Sobic.009G132900) and CA (Sobic.003G234200), with SNPs under purifying selection, while the other 2 C<sub>4</sub> genes, PPDK-RP (Sobic.002G324400) and PEPC (Sobic.010G160700), were identified with SNPs under balancing selection. Previous studies have demonstrated that SNP-level analysis using less stringent criteria is superior for capturing soft selection signals compared with genome-wide ranking [54,58]. However, the higher sensitivity may come with a cost of a greater chance of false positives, and therefore requires cautious interpretation. The contrasting selection signals on genes from the same pathway within taxa found in this study was also reported previously in signal transduction pathways [59] and the starch biosynthesis pathway [60].

The C<sub>4</sub> isoforms of PPDK and PEPC were also found to show signals of positive selection in a previous cross-species evolutionary study using orthologous groups from closely related C<sub>3</sub> and C<sub>4</sub> grass species including sorghum [28]. PPDK and PPDK-RP regulate the regeneration of PEP and as such have a direct effect on CO<sub>2</sub> assimilation rate [61], especially under cool temperatures [62,63]. However, it is thought that only minor changes to the enzyme properties of PPDK were sufficient to recruit it into the C<sub>4</sub> pathway and its residues and regions involved in catalyzes are highly conserved in C<sub>4</sub> species [64], possibly validating the fact that only soft selection signals via SNP-level were found for the C<sub>4</sub> isoform of the PPDK gene in our study.

PEPC is also regarded as a potential limiting step in the assimilation of CO<sub>2</sub>, and variation of its affinity for CO<sub>2</sub>/HCO<sub>3</sub><sup>-</sup> amongst species has been documented [65–67]. CA is also critical to C<sub>4</sub> photosynthesis as it catalyzes the first step of the C<sub>4</sub> pathway, converting CO<sub>2</sub> to HCO<sub>3</sub><sup>-</sup> [68]. It was reported in the C<sub>4</sub> dicot *Flaveria bidentis*, where antisense plants with <10% of wild-type CA activity required high CO<sub>2</sub> for growth and showed reduced CO<sub>2</sub> assimilation rates [69,70]. Recent experiments showed CA and PEPC will be more limiting when stomates are partially closed, e.g., under water limitation [71].

The signal of soft purifying selection on PPDK and CA may suggest the C<sub>4</sub> pathway was indirectly improved during sorghum domestication. Without photosynthetic rate being a direct selection target in breeding programs, a steady increase in leaf photosynthetic rate over time of cultivar release has been shown in other cereals, e.g., in Australian bread wheat [72]. The balancing selection signal on C<sub>4</sub> PPDK-RP and PEPC may reflect adaptation to diverse environments, as both PPDK-RP and PEPC are associated with abiotic stress [73,74]. Interestingly, within the PPDK-RP and PPDK gene families, the non-C<sub>4</sub> genes all showed selection signals contrasting with their C<sub>4</sub> counterparts with both two non-C<sub>4</sub> PPDK-RP (Sobic.002G324500, Sobic.002G324700) containing SNPs under purifying selection and the non-C<sub>4</sub> PPDK (Sobic.001G326900) containing SNPs under balancing selection.

After domestication, sorghum was introduced from tropical to temperate areas, and adapted to divergent local environments. New mutations also arose during this diversification process, and played an important role in local adaptation. In the haplotype analysis, these haplotypes unique to cultivated sorghum are likely to be young alleles arising after domestication, while haplotypes unique to the wild progenitor indicate that some haplotypes were lost during domestication of sorghum. Nevertheless, the

loss of wild haplotypes of  $C_4$  genes in cultivated sorghum does not mean these haplotypes are inferior in terms of photosynthetic efficiency, as photosynthesis was not specifically targeted during sorghum domestication [11]. On the contrary, bringing these wild haplotypes back to breeding programs after evaluation of their functions may enrich breeders' toolkits to manipulate photosynthetic efficiency, ultimately contributing to yield improvements.

$C_4$  photosynthesis has been well studied over the past 50 years and key components of this complex pathway have been identified following the advent of transgenic and sequencing technologies [9]. Understanding the genetic diversity of the key enzymes of the  $C_4$  pathway is an important step towards mining the natural allelic variation for the improvement of photosynthesis.

Further investigation of these allelic variation to link them with agronomical traits will provide new targets for sorghum improvement [75].

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2073-4425/11/7/806/s1>. Table S1: List of re-sequenced sorghum accessions and their racial and geographic origins. Table S2: List of housekeeping genes and their genetic diversity. Table S3 List of SNPs under purifying selection. Table S4: List of SNPs under balancing selection. Table S5: Inter- and intra-species distribution of private alleles across 27 genes from  $C_4$  gene families. Figure S1: Phylogenetic tree of 48 sorghum accessions based on CDS of 27 genes from  $C_4$  gene families.

**Author Contributions:** D.J. and E.M. conceived the original idea. Y.T., M.B.-P., S.T., A.C., and E.M. analyzed the data. Y.T. and B.G.-J. write the manuscript. All authors discussed the results and contributed to the final manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded partially by the Australian Government through the Australian Research Council Centre of Excellence for Translational Photosynthesis (grant number CE140100015) and State Key Laboratory of Agricultural Genomics, China (grant number 2011DQ782025).

**Acknowledgments:** We thank Susanne von Caemmerer and Robert Furbank for their valuable comments and suggestions towards the improvement of this manuscript.

**Conflicts of Interest:** The authors declare that there is no conflict of interest.

## Abbreviations

CA, carbonic anhydrase; PEPC, phosphoenolpyruvate carboxylase; PPCK, phosphoenolpyruvate carboxylase kinase; NADP-MDH, NADP-malate dehydrogenase; NADP-ME, NADP-malic enzyme; PPDK, pyruvate orthophosphate dikinase; PPDK-RP, PPDK regulatory protein; RbcS, ribulose biphosphate carboxylase/oxygenase small-subunit.

## References

1. Sage, R.F.; Sage, T.L.; Kocacinar, F. Photorespiration and the evolution of  $C_4$  photosynthesis. *Annu. Rev. Plant Biol.* **2012**, *63*, 19–47. [[CrossRef](#)] [[PubMed](#)]
2. Sage, R.F. The evolution of  $C_4$  photosynthesis. *New Phytol.* **2004**, *161*, 341–370. [[CrossRef](#)]
3. Zelitch, I.; Schultes, N.P.; Peterson, R.B.; Brown, P.; Brutnell, T.P. High glycolate oxidase activity is required for survival of maize in normal air. *Plant Physiol.* **2009**, *149*, 195–204. [[CrossRef](#)] [[PubMed](#)]
4. Edwards, E.J.; Smith, S.A. Phylogenetic analyses reveal the shady history of  $C_4$  grasses. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 2532–2537. [[CrossRef](#)]
5. Hibberd, J.M.; Covshoff, S. The regulation of gene expression required for  $C_4$  photosynthesis. *Annu. Rev. Plant Biol.* **2010**, *61*, 181–207. [[CrossRef](#)]
6. von Caemmerer, S.; Furbank, R.T. The  $C_4$  pathway: An efficient  $CO_2$  pump. *Photosynth. Res.* **2003**, *77*, 191. [[CrossRef](#)] [[PubMed](#)]
7. Hibberd, J.M.; Sheehy, J.E.; Langdale, J.A. Using  $C_4$  photosynthesis to increase the yield of rice—Rationale and feasibility. *Curr. Opin. Plant Biol.* **2008**, *11*, 228–231. [[CrossRef](#)] [[PubMed](#)]
8. Langdale, J.A.  $C_4$  cycles: Past, present, and future research on  $C_4$  photosynthesis. *Plant Cell* **2011**, *23*, 3879–3892. [[CrossRef](#)] [[PubMed](#)]
9. von Caemmerer, S.; Furbank, R.T. Strategies for improving  $C_4$  photosynthesis. *Curr. Opin. Plant Biol.* **2016**, *31*, 125–134. [[CrossRef](#)]
10. Still, C.J.; Berry, J.A.; Collatz, G.J.; DeFries, R.S. Global distribution of  $C_3$  and  $C_4$  vegetation: Carbon cycle implications. *Glob. Biogeochem. Cy.* **2003**, *17*, 1006. [[CrossRef](#)]

11. Long, S.P.; Zhu, X.G.; Naidu, S.L.; Ort, D.R. Can improvement in photosynthesis increase crop yields? *Plant Cell Environ.* **2006**, *29*, 315–330. [[CrossRef](#)] [[PubMed](#)]
12. Zhu, X.G.; Shan, L.L.; Wang, Y.; Quick, W.P. C<sub>4</sub> Rice—An ideal arena for systems biology research. *J. Integr. Plant Biol.* **2010**, *52*, 762–770. [[CrossRef](#)] [[PubMed](#)]
13. von Caemmerer, S.; Quick, W.P.; Furbank, R.T. The development of C<sub>4</sub> rice: Current progress and future challenges. *Science* **2012**, *336*, 1671–1672. [[CrossRef](#)] [[PubMed](#)]
14. Covshoff, S.; Szecowka, M.; Hughes, T.E.; Smith-Unna, R.; Kelly, S.; Bailey, K.J.; Sage, T.L.; Pachebat, J.A.; Leegood, R.; Hibberd, J.M. C<sub>4</sub> Photosynthesis in the Rice Paddy: Insights from the Noxious Weed *Echinochloa glabrescens*. *Plant Physiol.* **2016**, *170*, 57–73. [[CrossRef](#)] [[PubMed](#)]
15. Hatch, M.D.; Kagawa, T.; Craig, S. Subdivision of C<sub>4</sub>-pathway species based on differing C<sub>4</sub> acid decarboxylating systems and ultrastructural features. *Funct. Plant Biol.* **1975**, *2*, 111–128. [[CrossRef](#)]
16. Bräutigam, A.; Schliesky, S.; Külahoglu, C.; Osborne, C.P.; Weber, A.P.M. Towards an integrative model of C<sub>4</sub> photosynthetic subtypes: Insights from comparative transcriptome analysis of NAD-ME, NADP-ME, and PEP-CK C<sub>4</sub> species. *J. Exp. Bot.* **2014**, *65*, 3579–3593. [[CrossRef](#)]
17. Sonawane, B.V.; Sharwood, R.E.; Whitney, S.; Ghannoum, O. Shade compromises the photosynthetic efficiency of NADP-ME less than PEP-CK and NAD-ME C<sub>4</sub> Grasses. *J. Exp. Bot.* **2018**, *69*, 3053–3068. [[CrossRef](#)] [[PubMed](#)]
18. Grass Phylogeny Working Group II. New grass phylogeny resolves deep evolutionary relationships and discovers C<sub>4</sub> origins. *New Phytol.* **2012**, *193*, 304–312. [[CrossRef](#)]
19. Kidambi, S.P.; Krieg, D.R.; Rosenow, D.T. Genetic variation for gas-exchange rates in grain sorghum. *Plant Physiol.* **1990**, *92*, 1211–1214. [[CrossRef](#)]
20. Peng, S.B.; Krieg, D.R. Gas-exchange traits and their relationship to water-use efficiency of grain sorghum. *Crop Sci.* **1992**, *32*, 386–391. [[CrossRef](#)]
21. Henderson, S.; von Caemmerer, S.; Farquhar, G.D.; Wade, L.J.; Hammer, G. Correlation between carbon isotope discrimination and transpiration efficiency in lines of the C<sub>4</sub> species *Sorghum bicolor* in the glasshouse and the field. *Aust. J. Plant Physiol.* **1998**, *25*, 111–123. [[CrossRef](#)]
22. Balota, M.; Payne, W.A.; Rooney, W.; Rosenow, D. Gas exchange and transpiration ratio in sorghum. *Crop Sci.* **2008**, *48*, 2361–2371. [[CrossRef](#)]
23. Fernandez, M.G.S.; Strand, K.; Hamblin, M.T.; Westgate, M.; Heaton, E.; Kresovich, S. Genetic analysis and phenotypic characterization of leaf photosynthetic capacity in a sorghum (*Sorghum spp.*) diversity panel. *Genet. Resour. Crop Ev.* **2015**, *62*, 939–950. [[CrossRef](#)]
24. Zheng, L.Y.; Guo, X.S.; He, B.; Sun, L.J.; Peng, Y.; Dong, S.S.; Liu, T.F.; Jiang, S.Y.; Ramachandran, S.; Liu, C.M.; et al. Genome-wide patterns of genetic variation in sweet and grain sorghum (*Sorghum bicolor*). *Genome Biol.* **2011**, *12*, R114. [[CrossRef](#)] [[PubMed](#)]
25. Paterson, A.H.; Bowers, J.E.; Bruggmann, R.; Dubchak, I.; Grimwood, J.; Gundlach, H.; Haberler, G.; Hellsten, U.; Mitros, T.; Poliakov, A.; et al. The *Sorghum bicolor* genome and the diversification of grasses. *Nature* **2009**, *457*, 551–556. [[CrossRef](#)] [[PubMed](#)]
26. Mace, E.S.; Tai, S.S.; Gilding, E.K.; Li, Y.H.; Prentis, P.J.; Bian, L.L.; Campbell, B.C.; Hu, W.S.; Innes, D.J.; Han, X.L.; et al. Whole-genome sequencing reveals untapped genetic potential in Africa’s indigenous cereal crop sorghum. *Nat. Commun.* **2013**, *4*, 2320. [[CrossRef](#)]
27. Fankhauser, N.; Aubry, S. Post-transcriptional regulation of photosynthetic genes is a key driver of C<sub>4</sub> leaf ontogeny. *J. Exp. Bot.* **2017**, *68*, 137–146. [[CrossRef](#)]
28. Huang, P.; Studer, A.J.; Schnable, J.C.; Kellogg, E.A.; Brutnell, T.P. Cross species selection scans identify components of C<sub>4</sub> photosynthesis in the grasses. *J. Exp. Bot.* **2017**, *68*, 127–135. [[CrossRef](#)]
29. Burgess, S.J.; Hibberd, J.M. Insights into C<sub>4</sub> metabolism from comparative deep sequencing. *Curr. Opin. Plant Biol.* **2015**, *25*, 138–144. [[CrossRef](#)]
30. Reeves, G.; Grangé-Guermente, M.J.; Hibberd, J.M. Regulatory gateways for cell-specific gene expression in C<sub>4</sub> leaves with Kranz anatomy. *J. Exp. Bot.* **2017**, *68*, 107–116. [[CrossRef](#)]
31. Christin, P.-A.; Osborne, C.P.; Chatelet, D.S.; Columbus, J.T.; Besnard, G.; Hodkinson, T.R.; Garrison, L.M.; Vorontsova, M.S.; Edwards, E.J. Anatomical enablers and the evolution of C<sub>4</sub> photosynthesis in grasses. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 1381–1386. [[CrossRef](#)]
32. Moreno-Villena, J.J.; Dunning, L.T.; Osborne, C.P.; Christin, P.-A. Highly expressed genes are preferentially co-opted for C<sub>4</sub> photosynthesis. *Mol. Biol. Evol.* **2018**, *35*, 94–106. [[CrossRef](#)]

33. Borba, A.R.; Serra, T.S.; Gorska, A.; Gouveia, P.; Cordeiro, A.M.; Reyna-Llorens, I.; Knerova, J.; Barros, P.M.; Abreu, I.A.; Oliveira, M.M.O.; et al. Synergistic binding of bHLH transcription factors to the promoter of the maize *NADP-ME* gene used in C<sub>4</sub> photosynthesis is based on an ancient code found in the ancestral C<sub>3</sub> state. *Mol. Biol. Evol.* **2018**, *35*, 1690–1705. [[CrossRef](#)]
34. Christin, P.-A.; Petitpierre, B.; Salamin, N.; Büchi, L.; Besnard, G. Evolution of C<sub>4</sub> phosphoenolpyruvate carboxykinase in grasses, from genotype to phenotype. *Mol. Biol. Evol.* **2008**, *26*, 357–365. [[CrossRef](#)]
35. Christin, P.-A.; Salamin, N.; Muasya, A.M.; Roalson, E.H.; Russier, F.; Besnard, G. Evolutionary switch and genetic convergence on *rbcL* following the evolution of C<sub>4</sub> Photosynthesis. *Mol. Biol. Evol.* **2008**, *25*, 2361–2368. [[CrossRef](#)]
36. Gowik, U.; Bräutigam, A.; Weber, K.L.; Weber, A.P.; Westhoff, P. Evolution of C<sub>4</sub> photosynthesis in the genus *Flaveria*: How many and which genes does it take to make C<sub>4</sub>? *Plant Cell* **2011**, *23*, 2087–2105. [[CrossRef](#)]
37. Mallmann, J.; Heckmann, D.; Bräutigam, A.; Lercher, M.J.; Weber, A.P.; Westhoff, P.; Gowik, U. The role of photorespiration during the evolution of C<sub>4</sub> photosynthesis in the genus *Flaveria*. *eLife* **2014**, *3*, e02478. [[CrossRef](#)]
38. Williams, B.P.; Aubry, S.; Hibberd, J.M. Molecular evolution of genes recruited into C<sub>4</sub> photosynthesis. *Trends Plant Sci.* **2012**, *17*, 213–220. [[CrossRef](#)]
39. Wang, X.; Gowik, U.; Tang, H.; Bowers, J.E.; Westhoff, P.; Paterson, A.H. Comparative genomic analysis of C<sub>4</sub> photosynthetic pathway evolution in grasses. *Genome Biol.* **2009**, *10*, R68. [[CrossRef](#)]
40. Ermakova, M.; Danila, F.R.; Furbank, R.T.; von Caemmerer, S. On the road to C<sub>4</sub> rice: Advances and perspectives. *Plant J.* **2020**, *101*, 940–950. [[CrossRef](#)]
41. Nei, M.; Li, W.-H. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA* **1979**, *76*, 5269–5273. [[CrossRef](#)] [[PubMed](#)]
42. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **1989**, *123*, 585–595. [[PubMed](#)]
43. Watterson, G.A. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **1975**, *7*, 256–276. [[CrossRef](#)]
44. Hudson, R.; Boos, D.D.; Kaplan, N. A statistical test for detecting geographic subdivision. *Mol. Biol. Evol.* **1992**, *9*, 138–151. [[PubMed](#)]
45. Pfeifer, B.; Wittelsbürger, U.; Onsin, S.E.R.; Lercher, M.J. PopGenome: An efficient Swiss army knife for population genomic analyses in R. *Mol. Biol. Evol.* **2014**, *31*, 1929–1936. [[CrossRef](#)]
46. Paradis, E.; Claude, J.; Strimmer, K.J.B. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **2004**, *20*, 289–290. [[CrossRef](#)]
47. Paradis, E. Pegas: An R package for population genetics with an integrated–modular approach. *Bioinformatics* **2010**, *26*, 419–420. [[CrossRef](#)]
48. Ehleringer, J.R.; Sage, R.F.; Flanagan, L.B.; Pearcy, R.W. Climate change and the evolution of C<sub>4</sub> photosynthesis. *Trends Ecol. Evol.* **1991**, *6*, 95–99. [[CrossRef](#)]
49. Christin, P.A.; Salamin, N.; Savolainen, V.; Duvall, M.R.; Besnard, G. C<sub>4</sub> Photosynthesis evolved in grasses via parallel adaptive genetic changes. *Curr. Biol. CB* **2007**, *17*, 1241–1247. [[CrossRef](#)]
50. Christin, P.A.; Samaritani, E.; Petitpierre, B.; Salamin, N.; Besnard, G. Evolutionary insights on C<sub>4</sub> photosynthetic subtypes in grasses from genomics and phylogenetics. *Genome Biol. Evol.* **2009**, *1*, 221–230. [[CrossRef](#)]
51. Clark, J.D.; Stemler, A. Early domesticated sorghum from central Sudan. *Nature* **1975**, *254*, 588–591. [[CrossRef](#)]
52. Mann, J.A.; Kimber, C.T.; Miller, F.R. *The Origin and Early Cultivation of Sorghums in Africa*; Bulletin 1454; Texas Agricultural Experiment Station: College Station, TX, USA, 1983.
53. Wendorf, F.; Close, A.E.; Schild, R.; Wasylikowa, K.; Housley, R.A.; Harlan, J.R.; Królik, H. Saharan exploitation of plants 8,000 years BP. *Nature* **1992**, *359*, 721–724. [[CrossRef](#)]
54. Tao, Y.; Mace, E.S.; Tai, S.; Cruickshank, A.; Campbell, B.C.; Zhao, X.; Van Oosterom, E.J.; Godwin, I.D.; Botella, J.R.; Jordan, D.R. Whole-genome analysis of candidate genes associated with seed size and weight in *Sorghum bicolor* reveals signatures of artificial selection and insights into parallel domestication in cereal crops. *Front. Plant Sci.* **2017**, *8*, 1237. [[CrossRef](#)]
55. Tao, Y.; Zhao, X.; Mace, E.; Henry, R.; Jordan, D. Exploring and exploiting pan-genomics for crop improvement. *Mol. Plant* **2019**, *12*, 156–169. [[CrossRef](#)]



56. Kumpers, B.M.C.; Burgess, S.J.; Reyna-Llorens, I.; Smith-Unna, R.; Bournsnell, C.; Hibberd, J.M. Shared characteristics underpinning C<sub>4</sub> leaf maturation derived from analysis of multiple C<sub>3</sub> and C<sub>4</sub> species of *Flaveria*. *J. Exp. Bot.* **2017**, *68*, 177–189. [[CrossRef](#)]
57. Külahoglu, C.; Denton, A.K.; Sommer, M.; Maß, J.; Schliesky, S.; Wrobel, T.J.; Berckmans, B.; Gongora-Castillo, E.; Buell, C.R.; Simon, R.; et al. Comparative transcriptome atlases reveal altered gene expression modules between two *Cleomaceae* C<sub>3</sub> and C<sub>4</sub> plant species. *Plant Cell* **2014**, *26*, 3243–3260. [[CrossRef](#)]
58. Massel, K.; Campbell, B.C.; Mace, E.S.; Tai, S.; Tao, Y.; Worland, B.G.; Jordan, D.R.; Botella, J.R.; Godwin, I.D. Whole genome sequencing reveals potential new targets for improving nitrogen uptake and utilization in *Sorghum bicolor*. *Front. Plant Sci.* **2016**, *7*, 1544. [[CrossRef](#)]
59. Riley, R.M.; Jin, W.; Gibson, G. Contrasting selection pressures on components of the Ras-mediated signal transduction pathway in *Drosophila*. *Mol. Ecol.* **2003**, *12*, 1315–1323. [[CrossRef](#)]
60. Campbell, B.C.; Gilding, E.K.; Mace, E.S.; Tai, S.; Tao, Y.; Prentis, P.J.; Thomelin, P.; Jordan, D.R.; Godwin, I.D. Domestication and the storage starch biosynthesis pathway: Signatures of selection from a whole sorghum genome sequencing strategy. *Plant Biotechnol. J.* **2016**, *14*, 2240–2253. [[CrossRef](#)]
61. Wang, Y.-M.; Xu, W.-G.; Hu, L.; Zhang, L.; Li, Y.; Du, X.-H. Expression of maize gene encoding C<sub>4</sub>-pyruvate orthophosphate dikinase (PPDK) and C<sub>4</sub>-phosphoenolpyruvate carboxylase (PEPC) in transgenic *Arabidopsis*. *Plant Mol. Biol. Rep.* **2012**, *30*, 1367–1374. [[CrossRef](#)]
62. Wang, D.; Portis, A.R.; Moose, S.P.; Long, S.P. Cool C<sub>4</sub> photosynthesis: Pyruvate P<sub>i</sub> dikinase expression and activity corresponds to the exceptional cold tolerance of carbon assimilation in *Miscanthus × giganteus*. *Plant Physiol.* **2008**, *148*, 557–567. [[CrossRef](#)]
63. Naidu, S.L.; Moose, S.P.; AL-Shoaibi, A.K.; Raines, C.A.; Long, S.P. Cold Tolerance of C<sub>4</sub> photosynthesis in *Miscanthus × giganteus*: Adaptation in amounts and sequence of C<sub>4</sub> photosynthetic enzymes. *Plant Physiol.* **2003**, *132*, 1688–1697. [[CrossRef](#)]
64. Chastain, C.J.; Failing, C.J.; Manandhar, L.; Zimmerman, M.A.; Lakner, M.M.; Nguyen, T.H.T. Functional evolution of C<sub>4</sub> pyruvate, orthophosphate dikinase. *J. Exp. Bot.* **2011**, *62*, 3083–3091. [[CrossRef](#)]
65. Bauwe, H.; Chollet, R. Kinetic properties of phosphoenolpyruvate carboxylase from C<sub>3</sub>, C<sub>4</sub>, and C<sub>3</sub>-C<sub>4</sub> intermediate species of *Flaveria* (*Asteraceae*). *Plant Physiol.* **1986**, *82*, 695–699. [[CrossRef](#)]
66. von Caemmerer, S.; Edwards, G.E.; Koteyeva, N.; Cousins, A.B. Single cell C<sub>4</sub> photosynthesis in aquatic and terrestrial plants: A gas exchange perspective. *Aquat. Bot.* **2014**, *118*, 71–80. [[CrossRef](#)]
67. Boyd, R.A.; Gandin, A.; Cousins, A.B. Temperature responses of C<sub>4</sub> photosynthesis: Biochemical analysis of rubisco, phosphoenolpyruvate carboxylase, and carbonic anhydrase in *Setaria viridis*. *Plant Physiol.* **2015**, *169*, 1850–1861. [[CrossRef](#)] [[PubMed](#)]
68. Hatch, M.D.; Burnell, J.N. Carbonic anhydrase activity in leaves and its role in the first step of C<sub>4</sub> photosynthesis. *Plant Physiol.* **1990**, *93*, 825–828. [[CrossRef](#)]
69. von Caemmerer, S.; Quinn, V.; Hancock, N.; Price, G.D.; Furbank, R.T.; Ludwig, M. Carbonic anhydrase and C<sub>4</sub> photosynthesis: A transgenic analysis. *Plant Cell Environ.* **2004**, *27*, 697–703. [[CrossRef](#)]
70. Cousins, A.B.; Badger, M.R.; von Caemmerer, S. Carbonic anhydrase and its influence on carbon isotope discrimination during C<sub>4</sub> photosynthesis. Insights from antisense RNA in *Flaveria bidentis*. *Plant Physiol.* **2006**, *141*, 232–242. [[CrossRef](#)]
71. Osborn, H.L.; Alonso-Cantabrana, H.; Sharwood, R.E.; Covshoff, S.; Evans, J.R.; Furbank, R.T.; von Caemmerer, S. Effects of reduced carbonic anhydrase activity on CO<sub>2</sub> assimilation rates in *Setaria viridis*: A transgenic analysis. *J. Exp. Bot.* **2016**, *68*, 299–310. [[CrossRef](#)]
72. Watanabe, N.; Evans, J.R.; Chow, W.S. Changes in the Photosynthetic Properties of Australian wheat cultivars over the last century. *Aust. J. Plant Physiol.* **1994**, *21*, 169–183. [[CrossRef](#)]
73. Liu, X.; Li, X.; Zhang, C.; Dai, C.; Zhou, J.; Ren, C.; Zhang, J. Phosphoenolpyruvate carboxylase regulation in C<sub>4</sub>-PEPC-expressing transgenic rice during early responses to drought stress. *Physiol. Plant.* **2017**, *159*, 178–200. [[CrossRef](#)] [[PubMed](#)]



74. Jeanneau, M.; Gerentes, D.; Foueillassar, X.; Zivy, M.; Vidal, J.; Toppan, A.; Perez, P. Improvement of drought tolerance in maize: Towards the functional validation of the *Zm-Asr1* gene and increase of water use efficiency by over-expressing  $C_4$ -PEPC. *Biochimie* **2002**, *84*, 1127–1135. [[CrossRef](#)]
75. Tao, Y.; Zhao, X.; Wang, X.; Hathorn, A.; Hunt, C.; Cruickshank, A.W.; van Oosterom, E.J.; Godwin, I.D.; Mace, E.S.; Jordan, D.R. Large-scale GWAS in sorghum reveals common genetic control of grain size among cereals. *Plant Biotechnol. J.* **2020**, *18*, 1093–1105. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).