

RESEARCH ARTICLE

A sorghum (*Sorghum bicolor*) mutant with altered carbon isotope ratio

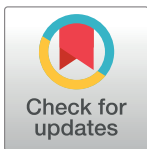
Govinda Rizal^{1,2a}, Shanta Karki^{1,2}, Vivek Thakur^{1,3}, Samart Wanchana¹, Hugo Alonso-Cantabrana⁴, Jacque Dionora¹, John E. Sheehy^{1,5,2b}, Robert Furbank⁴, Susanne von Caemmerer⁴, William Paul Quick^{1,5*}

1 C₄ Rice Center, IRRI, Los Banos, Laguna, the Philippines, **2** Government of Nepal, Ministry of Agricultural Development, Kathmandu, Nepal, **3** International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, India, **4** Australian Research Council Centre of Excellence for Translational Photosynthesis, Division of Plant Sciences, Research School of Biology, The Australian National University, Acton, ACT, Australia, **5** University of Sheffield, Sheffield, United Kingdom

^{2a} Current address: Beldangi 2, Extension E3- 25, Jhapa, Nepal

^{2b} Current address: Independent Researcher, Marlow, United Kingdom

* w.p.quick@irri.org



OPEN ACCESS

Citation: Rizal G, Karki S, Thakur V, Wanchana S, Alonso-Cantabrana H, Dionora J, et al. (2017) A sorghum (*Sorghum bicolor*) mutant with altered carbon isotope ratio. PLoS ONE 12(6): e0179567. <https://doi.org/10.1371/journal.pone.0179567>

Editor: Hector Candela, Universidad Miguel Hernández de Elche, SPAIN

Received: March 12, 2017

Accepted: May 31, 2017

Published: June 22, 2017

Copyright: © 2017 Rizal et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The research on C₄ rice is funded by Bill and Melinda Gates Foundation, UKAID (UK) and IRRI to Dr William Paul Quick, International Rice Research Institute (IRRI) and Dr Robert Furbank, Dr Susanne von Caemmerer, Australian National University (ANU).

Competing interests: The authors have declared that no competing interests exist.

Abstract

Recent efforts to engineer C₄ photosynthetic traits into C₃ plants such as rice demand an understanding of the genetic elements that enable C₄ plants to outperform C₃ plants. As a part of the C₄ Rice Consortium's efforts to identify genes needed to support C₄ photosynthesis, EMS mutagenized sorghum populations were generated and screened to identify genes that cause a loss of C₄ function. Stable carbon isotope ratio ($\delta^{13}\text{C}$) of leaf dry matter has been used to distinguish species with C₃ and C₄ photosynthetic pathways. Here, we report the identification of a sorghum (*Sorghum bicolor*) mutant with a low $\delta^{13}\text{C}$ characteristic. A mutant (named Mut33) with a pale phenotype and stunted growth was identified from an EMS treated sorghum M₂ population. The stable carbon isotope analysis of the mutants showed a decrease of ¹³C uptake capacity. The noise of random mutation was reduced by crossing the mutant and its wildtype (WT). The back-cross (BC₁F₁) progenies were like the WT parent in terms of ¹³C values and plant phenotypes. All the BC₁F₂ plants with low $\delta^{13}\text{C}$ died before they produced their 6th leaf. Gas exchange measurements of the low $\delta^{13}\text{C}$ sorghum mutants showed a higher CO₂ compensation point (25.24 $\mu\text{mol CO}_2 \cdot \text{mol}^{-1} \text{air}$) and the maximum rate of photosynthesis was less than 5 $\mu\text{mol} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$. To identify the genetic determinant of this trait, four DNA pools were isolated; two each from normal and low $\delta^{13}\text{C}$ BC₁F₂ mutant plants. These were sequenced using an Illumina platform. Comparison of allele frequency of the single nucleotide polymorphisms (SNPs) between the pools with contrasting phenotype showed that a locus in Chromosome 10 between 57,941,104 and 59,985,708 bps had an allele frequency of 1. There were 211 mutations and 37 genes in the locus, out of which mutations in 9 genes showed non-synonymous changes. This finding is expected to contribute to future research on the identification of the causal factor differentiating C₄ from C₃ species that can be used in the transformation of C₃ to C₄ plants.

Introduction

Stable carbon isotope ratios ($\delta^{13}\text{C}$) and carbon isotope discrimination ($\Delta^{13}\text{C}$) are used to distinguish C_4 and C_3 plants. Discrimination against ^{13}C during carbon fixation is greatly dependent on the photosynthetic type, mainly due to the characteristics of the enzyme catalyzing the first step in carbon fixation. Carbon isotope discrimination is defined as $\Delta = R_{\text{air}}/R_{\text{p}} - 1$ where R_{air} and R_{p} stand for the $^{13}\text{C}/^{12}\text{C}$ ratio in the air and the photosynthetic product, respectively [1,2]. In C_3 plants, this step is mediated by Rubisco, which has a strong preference for CO_2 containing the lighter isotope ^{12}C over the heavier and less abundant ^{13}C . In C_4 plants, inorganic carbon is initially fixed by Phosphoenolpyruvate carboxylase (PEPC) that does not discriminate between the two isotopes. The C_4 acid is then transferred to the bundle sheath cell where it is decarboxylated and the CO_2 accumulates to high concentrations. Leakages of CO_2 from the bundle sheath cells to the mesophyll cells are prevented by increased diffusive resistance of the bundle sheath cell walls, as such Rubisco is given no option but to fix both isotopes of carbon and hence the isotope discrimination is much lower in C_4 than in C_3 plant materials. As a result, the two photosynthetic types can be clearly distinguished by their signatures in carbon isotopic discrimination [3,4]. Interestingly, incomplete C_4 photosynthesis in intermediate C_3 - C_4 species can also be detected by its effect on $\Delta^{13}\text{C}$ [5,6]. Differential diffusion of $^{13}\text{CO}_2$ and $^{12}\text{CO}_2$ through stomata is another major component of overall discrimination [7]. In an ideal atmospheric condition, the ratio of ^{13}C to ^{12}C is roughly 1:99 [8]. The ratio of $^{13}\text{C}/^{12}\text{C}$ in the plant dry matter reflects the photosynthetic discrimination that occurred during its lifetime [9]. These variations in isotope ratios are integrated into the isotopic signature of leaf dry matter ($\delta^{13}\text{C}$) [3,10] which is usually referenced to the standard Pee Dee Belemnite (PDB) and defined as $\delta = R_{\text{p}}/R_{\text{PDB}} - 1$, where R_{p} and R_{PDB} stand for the $^{13}\text{C}/^{12}\text{C}$ ratio in leaf dry matter and the standard PDB, respectively [8]. The measurements of stable carbon isotope ratios ($\delta^{13}\text{C}$) and carbon isotope discrimination ($\Delta^{13}\text{C}$) are used to distinguish the photosynthetic efficiency of plants [3]. C_3 plants have $\delta^{13}\text{C}$ values between -23 and -35‰ which is lower than the $\delta^{13}\text{C}$ values of C_4 plants which are between -10 and -14‰ [10]. The genetic factors responsible for the differences in $\delta^{13}\text{C}$ between C_3 and C_4 plants have been the interest of scientists for a long time. In maize, one of the model C_4 species, the $\Delta^{13}\text{C}$ was proposed to be under polygenic control [11]. We hypothesized that induced mutations can revert a C_4 plant to a C_3 or an intermediate type and this reversion can be detected from its $\delta^{13}\text{C}$ signatures. The seeds of one of the accessions of the first sequenced sorghum (BTx623-Rooney) [12] were mutagenized using EMS and mutant seedlings screened to find those with altered $\delta^{13}\text{C}$ values. Several mutants, showing reduced (more C_3 -like) $\delta^{13}\text{C}$ were isolated for whole genome sequencing (WGS) [13] to map underlying mutations. The sequence analysis and the comparison between WT and mutants lead to the identification of several structural variances [14]. We were able to identify a potential genetic region controlling carbon isotope discrimination. These findings are useful for the identification of genetic factors driving the evolution of the C_4 photosynthetic pathway.

Results

Identification of sorghum mutant with low $\delta^{13}\text{C}$ values

The M_1 seedlings had green (92%), pale (4%) and albino (4%) phenotypes (Fig 1). From a population of one million M_1 seeds, 35,000 individual panicles were advanced to the M_2 generation. A total of 66 EMS mutants were selected based on their deviation from WT phenotypes and the $\delta^{13}\text{C}$ values of leaf dry matter of the mutants were measured. Four EMS mutants had a lower $\delta^{13}\text{C}$ than the WT (Fig 2). The four mutants, R.28-18184-01, R.28-18158-01, R.28-18161, and R.28-18012-01 with $\delta^{13}\text{C}$ values of -24.49, -16.76, -15.21 and -14.91 were labeled as Mut31,

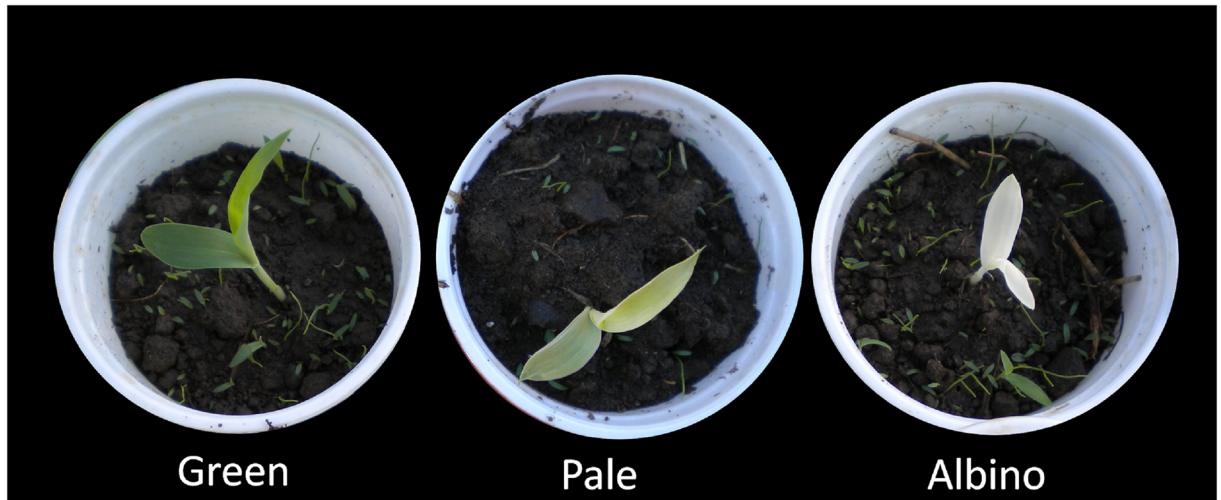


Fig 1. Representative images of notable phenotypes observed in EMS treated sorghum M₁ population. The percentage of green, pale and albino were 92, 4, and 4, respectively in the M₁ population.

<https://doi.org/10.1371/journal.pone.0179567.g001>

Mut32, Mut33, and Mut34, respectively (Fig 2). The $\delta^{13}\text{C}$ values of WT's grown in the same environment ranged between -12.89 and -12.55 (Fig 2). The M₂ progenies of the mutants were smaller, paler and slower growing than the WT (Fig 2B). However, only the Mut33 produced seeds.

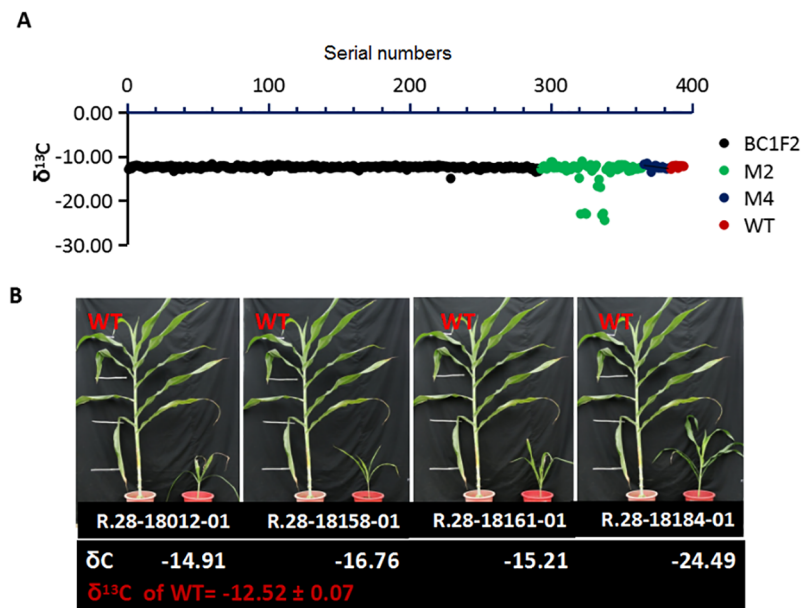


Fig 2. The carbon isotope ratio of the EMS mutants of different generations and wildtypes (WT). (A) The $\delta^{13}\text{C}$ values of selected wildtype and mutant sorghum plants accessions. The data of two generations of mutants M₂ and M₄ are shown. At M₂ generation the mutant was identified and the surviving M₄ corresponds to BC₁F₂ generation. The red dots represent WT values, the black dots represent mutants with normal $\delta^{13}\text{C}$, and the green dots represent the plants selected for further analysis. (B) A comparison of phenotypes and $\delta^{13}\text{C}$ values between low $\delta^{13}\text{C}$ mutants and WT. For convenience, the names of four mutants R.28-18184-01, R.28-18158-01, R.28-18161, and R.28-18012-01 were renamed as Mut31, Mut32, Mut33, and Mut34, respectively.

<https://doi.org/10.1371/journal.pone.0179567.g002>

Generation of Mut33 backcrossed population for sequencing

Mut33 was crossed to its WT and part of the panicle was self-pollinated to generate M_3 seeds. Six back-crossed F_1 (BC_1F_1) seeds from the cross between Mut33 and WT-Rooney and 18 self-pollinated (M_3) seeds were harvested. The BC_1F_1 plants grew normally and produced seeds at the same time as the WT. More than 3,000 seeds were harvested from a single plant. Both low $\delta^{13}C$ and normal $\delta^{13}C$ seedlings were identified in BC_1F_2 population. The M_3 seeds did not germinate in soil. Some M_3 seeds germinated in plant growth medium, but they could not survive beyond the sixth leaf stage. They were pale and the leaves desiccated starting from the tip of the leaves and died, even when grown under elevated CO_2 (10000ppm), controlled temperature, low light, and long day conditions. Even then they died before maturity. The BC_1F_2 seedlings were used for further analysis.

Characterization of BC_1F_2 progenies of Mut33

The $\delta^{13}C$ values in M_2 , M_3 , BC_1F_1 and BC_1F_2 population obtained from Mut33 were -15.21 ‰, -13.79 ‰ \pm 0.19, -12.95 ‰ and -13.22 ‰ \pm 0.72, respectively. The average (\pm SD) WT $\delta^{13}C$ values across the generations was -12.98 ‰ \pm 0.13. The BC_1F_1 had WT-like $\delta^{13}C$ values and phenotypes. In BC_1F_2 the range of $\delta^{13}C$ values broadened as indicated by large standard deviation (0.72) and ranged between (-15.61 and -12.06 ‰), showing a clear segregation of the $\delta^{13}C$ trait in the population. The samples were pooled according to their $\delta^{13}C$ values: SbPool1 (-14.319 to -13.5869), SbPool2 (-13.5762 to -13.3126), SbPool3 (-12.7924 to -11.1609) and SbPool4 (-12.6585 to -12.2538) (Table 1). Each pool had 30 samples and there were two pools from the first type of seedling. There were 115 seedlings with low $\delta^{13}C$ that died before producing sixth leaves, 31 seedlings with normal $\delta^{13}C$ which were slow growing and 51 normal seedlings with normal $\delta^{13}C$ values.

Chlorophyll content

A strong positive correlation was found between $\delta^{13}C$ and greenness in BC_1F_2 plants ($r = 0.77$; $n = 197$; using STAR Pearson's correlation analysis). This was explained by lower chlorophyll content in low $\delta^{13}C$ mutants. In free hand sections of the fresh leaves observed under a high-resolution microscope, the number of chloroplasts, seen as fluorescent red dots, was lower in the low $\delta^{13}C$ mutant plants. In the WT, the bundle sheath cells were full of chloroplasts and many chloroplasts were seen in mesophyll cells. In the BC_1F_2 plants, with normal $\delta^{13}C$, the number and distribution of chloroplasts were comparable to the WT. The chloroplasts number in the low $\delta^{13}C$ mutants had decreased substantially both in bundle sheath and mesophyll cells (Fig 3A–3C). The reduction in the number of chloroplasts in the BS cells as well as total number of chloroplasts in low $\delta^{13}C$ mutants, which correlated with reduced photosynthesis, could be the cause of early death in the mutants, as the absence of chloroplasts correlated with the significantly reduced rate of photosynthesis.

Stomatal density. The mutants with low $\delta^{13}C$ in BC_1F_2 had significantly higher stomatal density than the mutants with normal $\delta^{13}C$ and the WT ($P \leq 0.05$). The average density of leaf stomata in low $\delta^{13}C$, normal $\delta^{13}C$, and WT were 122 ± 0.52 , 60 ± 0.50 and 72 ± 2.45 stomata per mm^2 , respectively (Fig 3B). The stomatal density was not significantly different between the normal $\delta^{13}C$ and the WT.

Gas exchange measurement

We assumed that the first three leaves of a seedling are sustained by the nutrients stored in the seed and from the fourth leaf onward the photosynthesis supports complete autotropism. We

Table 1. Whole genome sequencing of pooled samples.

Sample	$\delta^{13}\text{C}$ values	Phenotypic traits	No. of paired-end reads (Millions)	Read length (bp)	No. of base pairs (Gbp)	Sorghum genome coverage
SbPool 1	-14.319 to -13.5869	Small plants, lethal	93.467	125	23.37	32X
SbPool 2	-13.5762 to -13.3126	Small plants, lethal	98.724	125	24.68	34X
SbPool 3	-12.7924 to -11.1609	Small plants, survived	135.059	125	33.76	46X
SbPool 4	-12.6585 to -12.2538	Tall plants, survived	129.253	125	32.31	44X

<https://doi.org/10.1371/journal.pone.0179567.t001>

observed the fourth leaf as the transition point for lethality in the BC_1F_2 plants with low $\delta^{13}\text{C}$. Thus, we took gas exchange measurements on the 3rd and 4th leaves. The A/C_i curves showed the P_{max} for the low $\delta^{13}\text{C}$ BC_1F_2 (green), normal $\delta^{13}\text{C}$ BC_1F_2 (blue) and WT (red) were 4.23, 18.89 and 25.45 $\mu\text{molCO}_2 \text{ m}^{-2}\text{s}^{-1}$ for the third leaf (square marker) and 4.09, 39.4, 43.74 $\mu\text{molCO}_2 \text{ m}^{-2}\text{s}^{-1}$ for the fourth leaf (circle marker), respectively (Fig 4A). The CO_2 compensation point (CP) had increased to 25.24 and 27.68 $\mu\text{molCO}_2 \text{ mol}_{\text{air}}^{-1}$ for the third and the fourth leaf of low $\delta^{13}\text{C}$ mutants, respectively. The CP for the normal $\delta^{13}\text{C}$ BC_1F_2 and WT (Fig 4A) was within the range of C_4 species (0 to 12 ppm) [15].

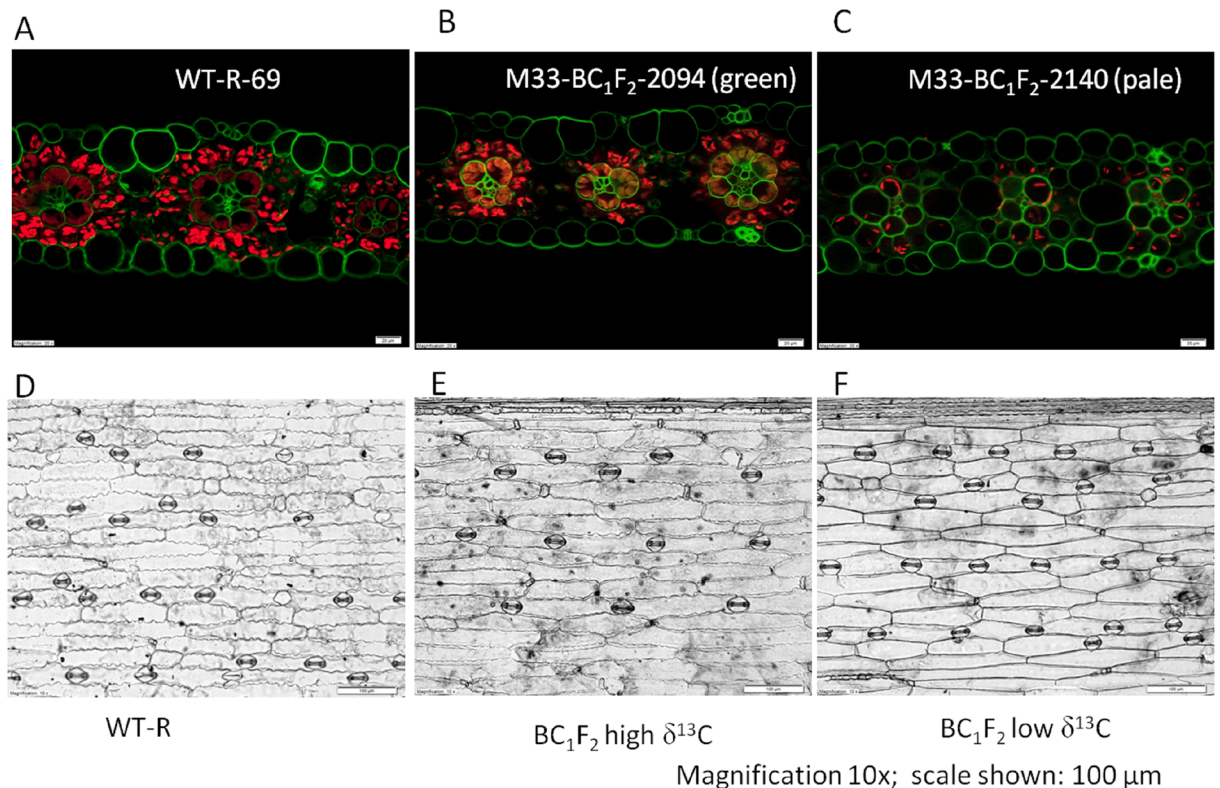


Fig 3. Free hand sections of fresh leaves of (A) WT, (B) normal $\delta^{13}\text{C}$ BC_1F_2 and (C) low $\delta^{13}\text{C}$ BC_1F_2 seedlings. The red dots show the presence of chlorophyll in the leaf section. (D, E, and F) The stomatal densities of the seedlings are in the same order as above. The low $\delta^{13}\text{C}$ BC_1F_2 mutants had more stomata per mm^2 than the WT/normal. Scale: in A, B, and C the magnification is 20x and the scale shown is 20 μm ; in D, E and F magnification is 10x and the scale shown is 100 μm .

<https://doi.org/10.1371/journal.pone.0179567.g003>

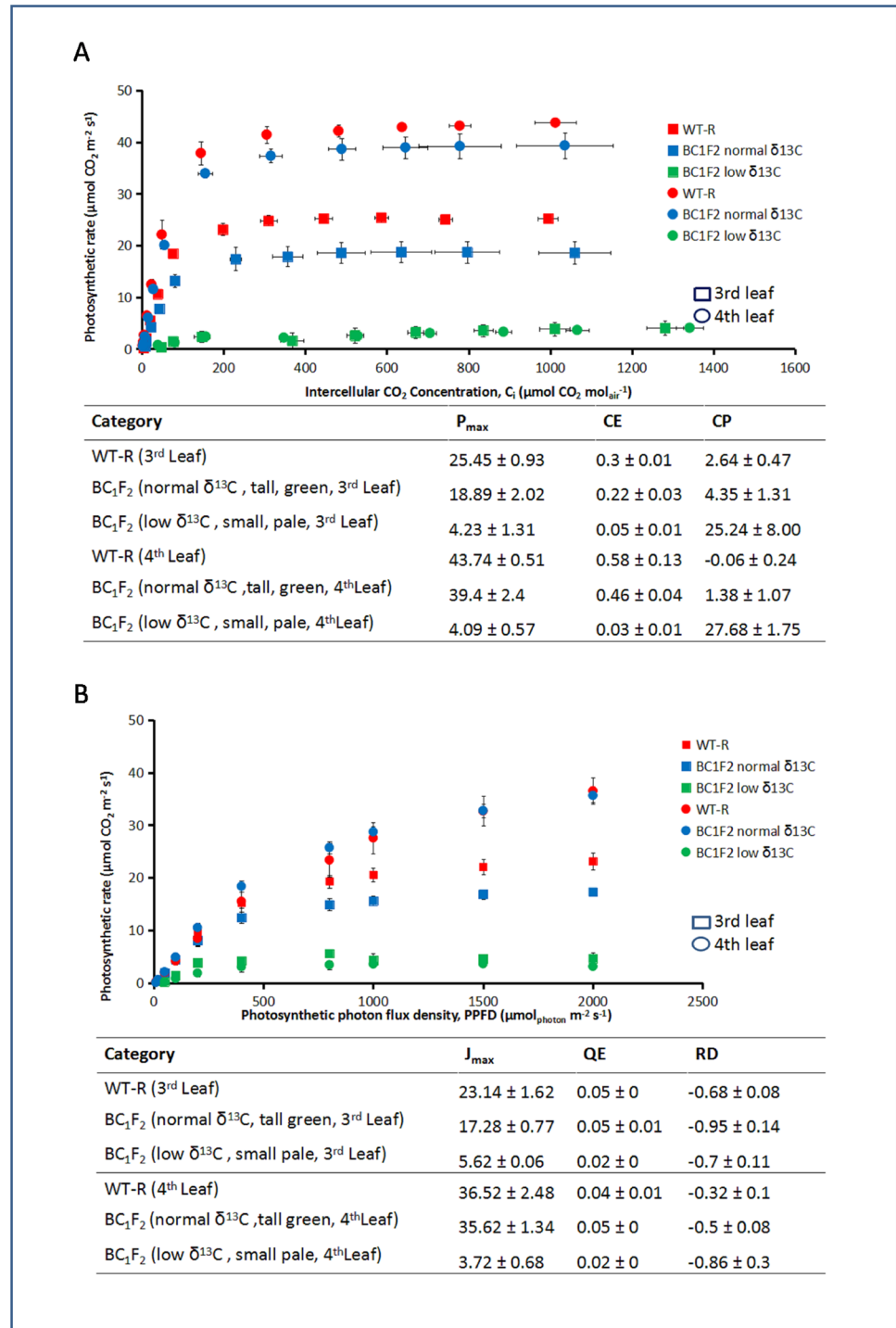


Fig 4. Measurement of photosynthetic parameters of BC₁F₂ mutants and wild type. (A) The photosynthesis rate per internal CO₂ concentration response curves (ACi) for the low $\delta^{13}C$ BC₁F₂ (green square), normal $\delta^{13}C$ BC₁F₂ (blue square) and wildtype (WT) (red square). The maximum rate of photosynthesis (P_{max}), carboxylation efficiency (CE) and CO₂ compensation points (CP) are shown in the table. However, the CO₂ compensation point is high for low $\delta^{13}C$ mutants. The squares show data of 3rd leaf and circles for the 4th leaf. The green, blue, and red colors represent low $\delta^{13}C$ BC₁F₂, normal $\delta^{13}C$ BC₁F₂, and WT, respectively. (B) The light response curve (LRC) for the low $\delta^{13}C$ BC₁F₂ (green square) normal $\delta^{13}C$ BC₁F₂ (blue square) and wildtype (red square). The maximum rate of photosynthesis (J_{max}), quantum efficiency (QE) and light compensation point (RD) are shown in the table. The squares show data of 3rd leaf

and circles of the 4th leaf. The green, blue and red marks represent low $\delta^{13}\text{C}$ BC₁F₂, normal $\delta^{13}\text{C}$ BC₁F₂, and WT, respectively.

<https://doi.org/10.1371/journal.pone.0179567.g004>

The J_{max} calculated from the light response curve for the low $\delta^{13}\text{C}$ BC₁F₂ (green), normal $\delta^{13}\text{C}$ BC₁F₂ (blue) and WT (red) were 5.62, 17.28, and 23.14 $\mu\text{molCO}_2\text{m}^{-2}\text{s}^{-1}$ for the third leaf (square) and 3.72, 35.62, and 36.52 $\mu\text{molCO}_2\text{m}^{-2}\text{s}^{-1}$ for the fourth leaf (circle), respectively (Fig 4B). There were no significant alterations in quantum efficiency and light compensation points.

One fourth of the BC₁F₂ seedlings died before the sixth leaf stage. Those lethal plants had a high CP and low P_{max} .

Whole genome sequencing and discovery of candidate genes

Four pools of DNA samples were sequenced at 32–46 X coverage (Table 1). The overall quality of sequence reads was good as filtering/trimming of low-quality reads or bases lost less than one percent of reads (Table 2). Since the reads were filtered and aligned against the WT parental genome (i.e., BTx623-Rooney), a very high percent of reads (~95%) aligned successfully to the reference genome (Table 2). Variants were discovered jointly in four samples followed by their filtering using multiple criteria (see Methods). The number of EMS-induced SNPs in the four samples ranged between 22,000 and 25,000 (S1 Fig).

Identification of causal locus

To identify the region of the genome associated with the mutant phenotype using a bulk segregant strategy, the frequency of mutant/alternate alleles (AF) was plotted along the genome [15] for all pooled samples. The density of AF was uniform across all chromosomes except at the end of the tenth chromosome (S2 Fig). Clearly, the end of the q-arm of chromosome 10 had a region where the mutant allele frequency in mutant pool (SbPool 1) rose to 1 and then declined (Fig 5, or S2 Fig), whereas in non-mutant pool (SbPool 4) the allele frequency in the same region was much lower than the average value of 0.5, which agreed with the expected allele frequency in the non-mutant pool of 0.33. The SNPs were extracted from this region of chromosome 10 (S2 Fig) and were annotated to identify those that can potentially affect the protein sequence. Out of 211 SNPs, ~70 SNPs overlapped with the gene boundaries (including core promoters and/or UTRs but excluding introns); only 9 of them caused a change in the amino acids (S3 Table). A database search for information on the function of these genes revealed that only four had curated information available: two of them (Sobic.010G239700 and Sobic.010G241900) have protein kinase activity, the third one (Sobic.010G249000) was a

Table 2. Sequence data analysis and variant discovery.

Sample	Raw unpaired reads (in Millions)	Unpaired reads passed filtering (in millions)	% passed	No. of clean reads (in millions)	No. of mapped reads (in millions)	Percent alignment	No. of variants
SbPool 1	186.93	185.69	99.33	185.69	176.68	95.14	10694
SbPool 2	197.44	196.16	99.34	196.16	189.74	96.72	10910
SbPool 3	270.11	267.65	99.08	267.65	247.54	92.48	10694
SbPool 4	258.5	256.15	99.09	256.15	252.04	98.39	10694

<https://doi.org/10.1371/journal.pone.0179567.t002>

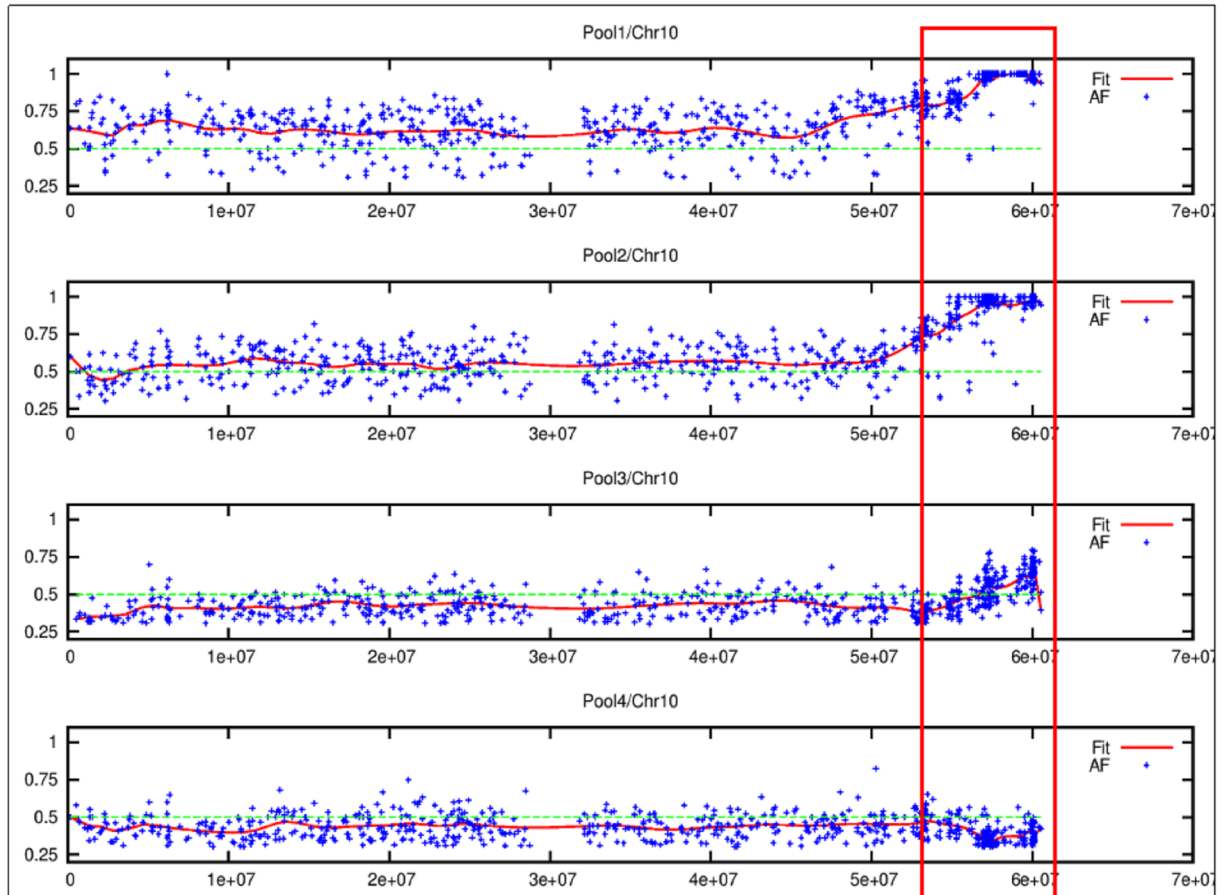


Fig 5. The focused image of q arm of chromosome 10 where the alternate allele frequency was high (≈ 1) in SbPool 1 and SbPool 2, the pools with low $\delta^{13}\text{C}$ and fatal samples about 0.5 in SbPool 3, the pool with normal $\delta^{13}\text{C}$ but slow growing samples and about 0.33 in SbPool 4, the pool with normal $\delta^{13}\text{C}$ and phenotype.

<https://doi.org/10.1371/journal.pone.0179567.g005>

disease resistance gene, while the fourth one (Sobic.010G250100) showed homology with a transcription factor involved in chloroplast targeted protein import (Table 3). Among the uncharacterized genes, the cellular location of one of the genes (Sobic.010G266000) was predicted to be related to the plastid and its ortholog in maize shows abundance exclusively in leaf tissues (S3 Fig). There are two candidates that show a very clear association with the

Table 3. The SNPs, from the locus of the genome associated with the mutant phenotype, causing amino acid changes.

SN	Gene ID	Position (BTx623)	Amino acid changes	Annotation
1	Sobic.010G239700	57965771	Ala>Thr	cysteine-rich RLK (RECEPTOR-like protein kinase) 10
2	Sobic.010G239900	57978990	Cys->Tyr	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
3	Sobic.010G241900	58127862	Leu>Phe	Leucine-rich receptor-like protein kinase family protein
4	Sobic.010G249000	58665427	Asp>Asn	NB-ARC domain containing
5	Sobic.010G250100	58730675	Pro>Leu	CIL, CIA2 like (Chloroplast import apparatus)
6	Sobic.010G263800	59789763	Gly>Asp	uncharacterized protein
7	Sobic.010G266000	59985708	Pro>Ser	uncharacterized protein
8	Sobic.010G264000	59800420	Val>Ile	expressed protein
9	Sobic.010G265600	59962313	Val>Ile	Protein of unknown function (DUF581)

<https://doi.org/10.1371/journal.pone.0179567.t003>

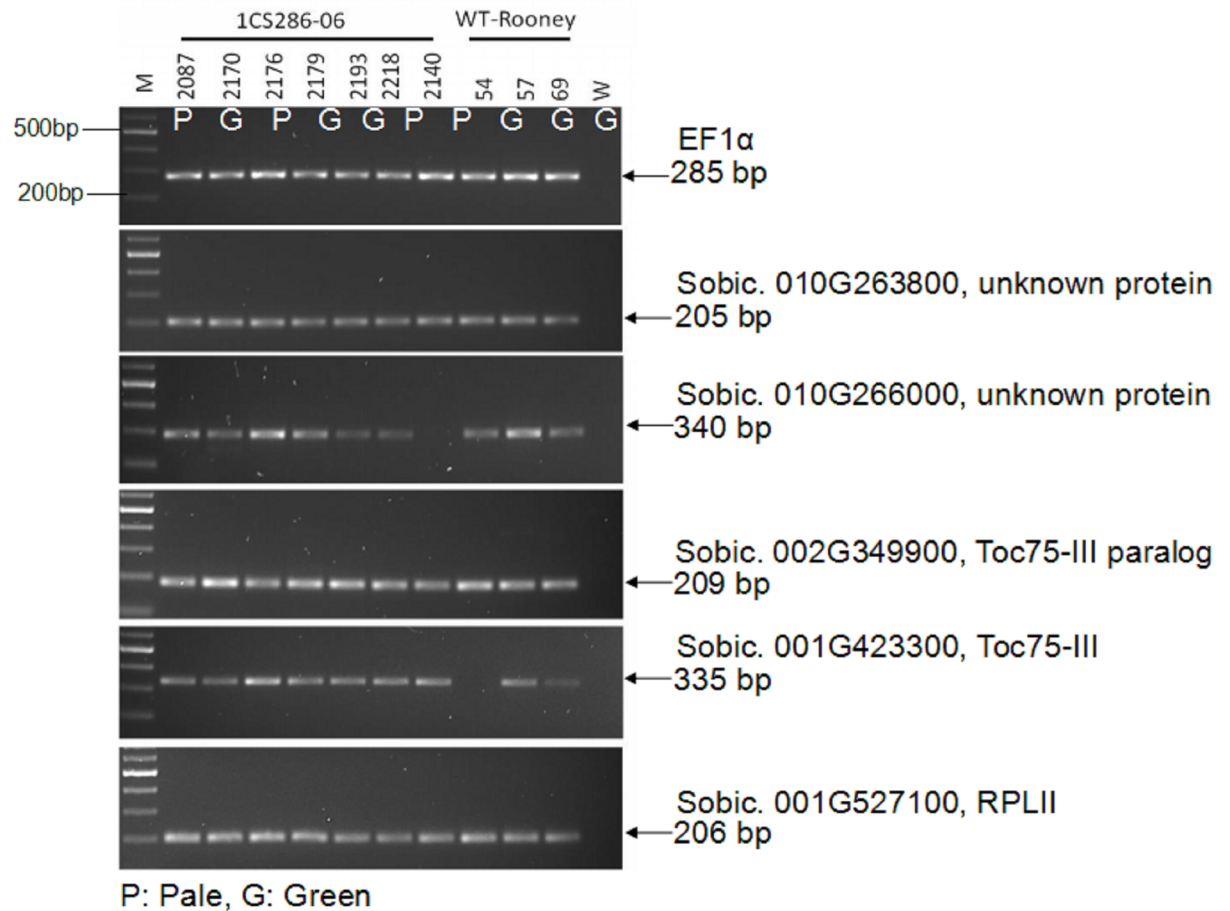


Fig 6. Semi-quantitative PCR analysis of candidate genes of Mut33 BC₁F₂ seedlings. The bands represent expression of Sobic.010G263800, Sobic.010G266000, and the three downstream genes of CIA2, Sobic.002G349900 (TOC75-III paralog), Sobic.001G423300 (TOC75-III), and Sobic.001G527100 (RPL11) in BC₁F₂ (1CS286-06) and wildtype (WT-Rooney) samples. There are no changes in gene expression levels of TOC75-III, and RPL11. EF1α is the housekeeping gene.

<https://doi.org/10.1371/journal.pone.0179567.g006>

chloroplast. A transcription factor (TF) predicted to be involved in chloroplastic protein import (Sobic.010G250100) [16] was found to have very high transcript abundance in the maize primordial tissues and almost nil in the foliar immature and expanded stages (S4 Fig). Expression profile of all the candidates showed a variation in expression across different tissues (S5 Fig).

Gene expression analysis of the major candidate genes

Candidate genes were tested for transcript expression using RNA from the third leaf stage. Results from semi-quantitative PCR showed that transcripts of Sobic 010G263800 and Sobic 010G266000, both annotated as uncharacterized proteins were amplified at similar levels in WT and the pale mutants (Fig 6). Sobic 010G239700 showed faint amplification in the two samples while Sobic 010G24900 could not be amplified (S6 Fig). The most promising candidate gene namely CIA2 like (chloroplast import apparatus) Sobic 010G250100 showed faint amplification only in three samples, which could be due to the RNA from the leaf tissue as this gene was found to express in the primordial stage in maize (Fig 6). Therefore, three CIA2 pathway genes Toc75-III (Sobic 001G423300), Toc75-III paralog (Sobic 002G349900) and RPL11

(Sobic 001G527100) were analyzed. All three showed similar levels of transcript abundance in the mutants and WT (Fig 6, S6 Fig). Other genes in the CIA2 pathway that are related to chloroplast formation and development could be tested to gain more insight of the causal gene.

Discussion

The mutants were first selected through visual inspection of deviation from WT phenotypes. These mutants were slower growing and paler compared to the WT, similar characteristics were observed in mutants with defective CIA2, one of the transcription factors involved in chloroplast protein import [17]. The sorghum mutants had high mortality and low $\delta^{13}\text{C}$. The differences in degree of stable carbon isotope discrimination between C_3 and C_4 plants can be used to screen mutants with the loss of function of the C_4 pathway. In maize, another C_4 species, the $\Delta^{13}\text{C}$ was demonstrated to be under polygenic control [11]. There are no reports on the genetic factors controlling ^{13}C composition in sorghum. An inverse relationship between grain yield and carbon isotope discrimination probably due either to the porosity of BS cells affecting the light use efficiency or the differences in CO_2 assimilation rate and stomatal conductance due to variable transpiration efficiency, was previously reported [18]. The sorghum mutants identified in this study had a higher stomatal density (Fig 3F), fewer chloroplasts (Fig 3C), lower carboxylation efficiency and higher CO_2 compensation point compared to the WT (Fig 5A). The range and average $\delta^{13}\text{C}$ are -23 to -35 ‰ and -26 ‰ in C_3 plants and from -10 to -14 ‰ and -13 ‰ in C_4 plants, respectively [10]. We found that even small changes in carbon isotope ratios were correlated with lethal effects in the plants. Is the lethality in the mutants controlled by the same factors controlling the ^{13}C composition or some major genes directly or pleiotropically affecting the ^{13}C composition?

The CO_2 response curve indicated that the photosynthetic capacities of these mutants were impaired compared to the WT. The phenotype was recovered in the BC_1F_1 generation. The photosynthetic impairment was severe to lethal in one-fourth of the seedlings in the BC_1F_2 generation. Unlike in maize [19] where ^{13}C amount is controlled by polygenes, the alteration of $\delta^{13}\text{C}$ in sorghum mutants is either a single recessive gene or mutation in a vital gene or in a pathway that affected the assimilation of ^{13}C and caused premature senescence. The available mutant populations are important materials to study the genetic factors controlling the indiscriminate uptake of ^{13}C isotope by C_4 plants, which is absent in C_3 plants. The alignment of DNA sequences from BC_1F_2 mutants with low and normal $\delta^{13}\text{C}$ against the WT genome sequence showed a consistently high alternate frequency (AF) in the tail (q) end of chromosome 10. Based on the consistently high AF, we propose the factor responsible for $\delta^{13}\text{C}$ variation between WT and the mutants is in the locus between the 57,941,104 and 59,985,708 bp. Mutations affecting 37 genes were found in that locus (S2 Table) of which mutations in 9 genes had caused non-synonymous changes (Table 3). The functional annotation of Sobic.010G239700 was cysteine-rich repeat receptor-like protein kinase (*CRR-RLK*) [20–23]. The *CRR-RLKs* are receptor-like kinases in plants with roles in signal transduction in response to extracellular stimuli and stimulation of downstream pathways. The *CRR-RLKs* are involved in growth regulation, development, and physiological responses [24]. In *Arabidopsis*, they are activated in response to UV rays and pathogen-response [24, 25]. The Sobic.010G239900 has roles in oxidoreductase activity [20, 21]. Oxido-reductase activity uses ferrous iron as a cofactor to catalyze 2-oxoglutarate into succinate [20]. The Fe (II) 2OG dioxygenase domain enzymes in plants catalyze the production of plant hormones, such as ethylene, gibberellins, anthocyanidins and pigments such as flavones [26]. Sobic.010G239900 is one of the Leucine-rich repeat receptor-like protein kinases (*LRR-RLK*) [27]. The *LRR-RLKs* are transmembrane receptor-like kinases in plants. They regulate various developmental and defense-related

processes such as cell proliferation, stem cell maintenance, and hormone perception [28–30]. They play an important role in pathways of brassinosteroid signaling in the wound-responsive signaling pathway in Solanaceous plants, nodule development in leguminous plants and pathogen-recognition in *Arabidopsis* [28, 30]. The *LRR-RLKs* are an integral component of membranes consisting of gene products and complexes having peptides embedded in the hydrophobic region of the plasma membrane. They perform transmembrane signaling receptor activity that transmits a signal across the membrane via kinase activity or phosphorylation of amino acid residues in a protein such as threonine and serine signaling a wound-response [20]. The Sobic.010G241900 has a functional annotation of Leucine-rich repeat-containing protein [27, 31]. The gene has roles in defense responses to foreign bodies or injuries to reduce damage [27, 31, 32]. The gene Sobic.010G250100 is similar to putative stress resistance-related protein [30]. One of the proteins is related to CHLOROPLAST IMPORT APPARATUS 2 (CIA2) e.g. At5g57180 [33]. The mutation of this gene results in a pale phenotype that is defective in the general chloroplast protein import pathway. Its subcellular location is in the chloroplast and nucleus and is expressed in young leaves and flower buds [33]. The gene CIA2 is crucial for development of photosynthetic apparatus but it is not known if it also has a pleiotropic effect on the carbon concentrating mechanism. Sobic.010G250100 is responsible for specific up-regulation of the translocon genes TOC33 and TOC75 in leaves. It is also involved in the general chloroplast protein import pathway regulation, including protein import and protein translation efficiency [33, 34]. The biological process includes protein targeting to chloroplasts [33] and regulation of transcription [34]. The gene Sobic.010G263800 encodes a protein that belongs to uncharacterized protein family (UPF0183 protein) [20, 23, 27, 31]. The gene is involved in the response to symbiotic fungus [35]. The Sobic.010G266000 (GRMZM2G373420) has no functional annotation (<http://phytozome.jgi.doe.gov/pz/portal.html>). The gene Sobic.010G264000 (GRMZM2G015818) encodes a protein of unknown function (DUF1668). The hypothetical proteins found in this family are expressed in *Oryza sativa* and are of unknown function [36]. The gene Sobic.010G265600 (GRMZM2G009080) or “DUF581” is a Zf-FCS type zinc finger. Zinc fingers are a ubiquitous class of protein domains with considerable variation in structure and function. FCS-like zinc finger proteins have small motifs with multiple finger-like protrusions that make tandem contacts with their target molecule with which they interact. Some of its functions are (1) modular building blocks for the construction of larger protein domains that recognize and bind to specific DNA sequences [37], (2) DNA-binding transcription factors [17] and (3) recognition of RNA and other protein [38]. Transcript abundance of the eight candidate genes were examined which showed only marginal differences. Finer analysis and study of the effect of individual genes could reveal the genetic factors responsible for the differentiation of $\delta^{13}\text{C}$ in C_3 and C_4 plants. The causal gene is most likely to be one of the genes mentioned above (Table 3) unless the causal mutation lie outside the coding sequence of a gene, or in a sequence not annotated as a gene (such as a micro RNA sequence or an incorrectly annotated gene). But given the mutant phenotype characterized by significantly reduced number of chloroplasts, and one of the candidate genes being a transcription factor annotated to regulate chloroplast protein import during chloroplast development (CIA2 like), and this being transcriptionally active in the developmental stages of leaf known for plastid biogenesis and differentiation, we propose this gene to be most likely causal gene.

If carbon comes from the seed then it will have the C_4 signature of the parent which was not the case. If carbon comes from recently fixed CO_2 , it reflects the type of pathway (C_3 or C_4) which was more C_3 like in the mutants. Therefore, we assume that something has gone wrong in the mutant with the C_4 pathway. It is the current photosynthetic pathway that is providing the sugar for growth with a more C_3 -like signature. Moreover, the screen based on ^{13}C

was effective for the identification of photosynthetic mutants and was a useful parameter to cross-check successful crossings.

In conclusion, our assumption is that the C_4 pathway has been compromised and the cost of CO_2 concentration has been increased leading to a negative carbon balance and hence is lethal. So, whilst development of C_4 may be the cause, the end result is loss of C_4 function. Further experiments are needed to pin down the causal trait to the gene level. This finding is expected to help future research on the genetic control of carbon isotope discrimination in plants.

Methods

Plant materials

Sorghum (*Sorghum bicolor* L. [Moench]) BTx623 seeds were obtained from Professor William L. Rooney of the Sorghum Breeding and Genetics Division (Texas A & M University, College Station, Texas, USA). Hence, the accession was named BTx623-Rooney or WT-R [12]. Cultivation for seed multiplication was done in the fields in IRRI, Los Baños (14° 11' N, 121° 15' E) in the Philippines. Mature seeds of sorghum BTx623-Rooney treated with 0.28% EMS to generate a mutant population. The M_1 seeds were grown in 250 ml plastic cups filled with soil fertilized with 30 kg per hectare of nitrogen, phosphorus and potassium (NPK) combined at a ratio of 3:1:1. The cups were laid out in 1 m² plot(s). Three weeks old seedlings were transplanted into soil with 10 cm x 75 cm plant spacing. Seedlings were fertilized with a basal dose of NPK at the rate of 30 kg ha⁻¹. Urea was added at a rate of 97 kg ha⁻¹ at 21 and 35 days after transplantation (DAT). A sprinkler irrigation system was used. Around 100 DAT, the M_2 seeds were harvested, air-dried for four days at 40°C to bring down the moisture content between 8 and 10%. Each panicle was treated as a line. The seeds were stored in a 4°C cold room. For the cultivation of M_2 generation, we used lines that produced more than 50 seeds. From a population of one million M_1 seeds, 35,000 individual panicles were advanced to the M_2 generation.

More than 35,000 M_2 lines were sown. For each M_2 line, 24 seeds were grown. Each seed was sown in seedling trays (Ronaash[®] Slim, Rannoch) containing 50 cells of 100 ml capacity. Each cell was filled with fertilized soil (0.025 g NPK kg⁻¹ soil combined at a ratio of 3:1:1). The seedling trays were grown in field conditions. Seven days after sowing (DAS) percentage germination and visual phenotypes were recorded and seedlings with the desired phenotypes were selected. The selected plants were transplanted into 8 L pots and fertilized with 0.025 g NPK kg⁻¹ soil combined at a ratio of 3:1:1. Pots were irrigated daily. Around 100 DAT, the M_3 seeds were harvested, air-dried for four days at 40°C and stored at 4°C cold room until used.

Phenotyping

A week after sowing, the number of seeds that germinated, the frequency of visual phenotypes such as albino, pale, crinkly and slow growing plants were recorded to analyze the occurrence of such mutants. Disadvantaged (pale, crinkly and slow growing) mutants were sampled for stable ¹³C isotope analysis. The rate of photosynthesis was also measured for some pale and slow growing mutants using LI-6400 portable photosynthesis systems. After the measurements and analysis, the seedlings were grown in a high CO_2 chamber at 10,000 ppm for rescue and 25 ± 2°C for 12 hours during the night and 29 ± 2°C during the day.

Dry matter $\delta^{13}C$ analysis

After measurement of gas exchange, sections of the same leaf were sampled for dry matter and microscopic analysis. Tissue from the center of the leaf blade was harvested, dried overnight in an oven at 80°C, and ground to a fine powder. Dry matter carbon isotope composition

measurements were performed as previously described [39]. One to two mg samples were flash-combusted in a CE1110 CHN-S analyzer (Carlo Erba, UK) and the CO₂ isotopic composition determined by mass spectrometry in a Fisons Isochrom CF-IRMS (Continuous-flow Isotope Ratio Mass Spectrometer; Isoprime, UK). The δ¹³C signatures are presented as isotopic ratios (per mill, ‰) relative to the isotopic standard Pee Dee Belemnite.

Gas exchange measurements

Plants selected for gas exchange measurement were watered in the morning and kept in the area of measurement for at least 90 minutes for acclimatization of the plant prior to the actual measurement. The third and the fourth leaves were clipped by the head of the infrared gas analyzer (IRGA) of the LI-6400XT portable photosynthesis system (LICOR Biosciences, Lincoln, NE, USA). The leaf temperature was maintained at 30°C as per the temperature of the measurement room, with a light intensity of 2,000 μmol m⁻² s⁻¹. The mean atmospheric pressure at the measurement site (38 m above the sea level) was 94.8 kPa. A constant airflow of 400 μmol s⁻¹ was maintained and the leaf-to-air vapor pressure deficit was between 1.0 and 1.5 kPa. The leaves were acclimated in the cuvette for about 30 minutes.

The response curves of the rate of CO₂ assimilation (A , μmolCO₂ m⁻²s⁻¹) to the changes in intercellular CO₂ concentration (C_i , μmolCO₂mol⁻¹) were acquired by increasing CO₂ concentration in the cuvette (C_a) from 0 to 2000 μmol CO₂ mol⁻¹ at a photosynthetic photon flux density (PPFD) of 2000 μmol m⁻²s⁻¹. The carboxylation efficiency (CE) was calculated from the initial slope ($C_i < 100$) of the A - C_i curves; and CO₂ compensation point (CP, μmolCO₂ per mol_{air}) was taken from the point of intercept between the A - C_i curve and the X-axis [40].

Similarly, the light-response curves were obtained by plotting photosynthetic rate against PPFD from 2000 to 0 μmol_{quanta}·m⁻²·s⁻¹ at a constant C_a of 400 μmolCO₂mol⁻¹_{air}. The quantum efficiency (QE) was calculated from the initial slope of the light-response curves using the first four points of PPFD that were less than 100 μmol photons m⁻² s⁻¹ [41]. The light compensation point was taken from the point of intercept between the light response curve and the X-axis. All other conditions were as in the AC_i measurement.

Chlorophyll fluorescence microscope imaging

The third leaf, when it was the youngest fully expanded leaf of the seedling, was sampled. Three cm long leaf strips from the middle part of the leaf blade were cut, fixed in 2.5% glutaraldehyde, placed in scintillation vials, vacuum infiltrated (20 psi) for 20 minutes and stored at 4°C overnight. Using a sharp razor blade, very fine sections of leaf were cut and stained with 0.05% calcofluor-white. Leaf sections were mounted in glycerol, viewed and imaged using a fluorescence microscope (Olympus DSU (Disc Spinning Unit) Confocal System, Japan). The total chlorophyll content or “greenness” was measured using Chlorophyll meter SPAD. The data from the SPAD reading are comparable to the greenness of the leaves [42].

Stomatal density

For each plant type (low δ¹³C, normal δ¹³C, and WT) four representative plants were sampled. For each sample, the stomatal densities in 10 sections were recorded and their mean and standard deviation calculated. A section of the third leaf that was sampled for chlorophyll fluorescence imaging was used to study the density of stomata. The abaxial and adaxial epidermis of the leaf were gently cleaned using a damp paper towel, carefully smeared with nail varnish in the mid-area between the central vein and the leaf edge and left for 20 min until dry. The thin film (approximately 5 mm×10 mm) was peeled off the leaf surface, mounted on a glass slide, and viewed using a light microscope (Olympus BX63, Japan).

Crossing and advancement of generation

The sorghum mutant was crossed to its WT (BTx623-Rooney) to obtain both cross- and self-pollinated seeds following a segregation crossing strategy [43]. The BC₁F₂ was used for segregation analysis of $\delta^{13}\text{C}$ and for whole genome re-sequencing.

Leaf sampling, DNA extraction, quality control, and sequencing

Young and tender leaves were transferred into liquid nitrogen. One leaf each from five seedlings were pooled and then ground to a fine powder for DNA isolation. Total DNA was extracted following the CTAB method [44] and checked for quality by running on a 1.5% agarose gel. The concentration of DNA was measured using a nanodrop (ND-8000, Thermo Scientific). Equal concentrations of DNA from six individual sub-pools were pooled to obtain 25 μg DNA. Four pools were prepared namely SbPool 1 to SbPool 4, such that each pool had equal amount of DNA from 30 seedlings. The SbPool 1 and SbPool 2 contained DNA from samples of slow growing pale mutant seedlings with low $\delta^{13}\text{C}$. The SbPool 3 DNA was obtained from small plants with normal $\delta^{13}\text{C}$, and the SbPool 4 DNA was obtained from normal BC₁F₂ seedlings with normal $\delta^{13}\text{C}$ (S1 Fig). The DNA pools were sequenced using HiSeq2500 PE125 sequencing strategy (BGI Tech Solutions Co., Shenzhen, China). The WT samples were also sequenced for comparison of sequences against the mutants. The raw read data for this project have been submitted to the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) under BioProject ID PRJNA384699. The SRA accession numbers are SRX2768409, SRX2771011, SRX2771012 and SRX2771013. The sequence data of wildtype sorghum was submitted previously with SRA accession number SRX973468.

Construction of individual genome for wildtype (BTx623-Rooney)

An alignment of WT reads onto Sorghum reference genome (BTx623) showed the WT sequences varied from the reference genome in a number of positions. Therefore, construction of an individual genome for the WT [12] was preferred and submitted to SRA accession number SRX973468. In short, the steps involved filtering reads for base quality using FASTX toolkit (parameters used: base quality ≥ 20 , minimum length after 3' trimming ≥ 30 , and fraction of read length with high-quality bases ≥ 0.85), followed by their alignment with default parameters except that only paired-end reads were used for alignment. Reads of size ~ 100 bp were aligned using BWA-MEM v0.6.9 (<http://bio-bwa.sourceforge.net/>) and that ~ 50 bp were aligned using Bowtie2 [45]. The WT genome was reconstructed from the alignment by Pilon [46], a tool used for assembly improvement. To further improve the quality of the individual WT genome, Pilon was run iteratively eight times so that the detected variants stabilized.

Processing of short sequence reads and variant calling

The quality of sequencing data was initially evaluated using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), followed by read trimming/ filtering for base or read quality using Trimmomatic v0.32; [47]. The parameters used were leading = 10, trailing = 10, sliding window = 5:15, and minimum length = 50. The filtered reads were aligned to BTx623-Rooney genome by BWA-MEM (v0.7; <http://bio-bwa.sourceforge.net/>) with default parameters. The alignment was improved by (1) fixing mate information by samtoolsv1.2 [48], (2) realignment around InDels, (3) base quality re-calibration using a set of very high confidence variants, and (4) mark PCR duplicates by Picard v1.96 (<http://broadinstitute.github.io/picard/>). The steps 2 and 3 were carried out using GATK v3.3-0 [49]. Variant calling was done for individual samples using HaplotypeCaller, a part of GATK followed by joint genotyping of all four samples using

GenotypeGVCF, a part of GATK. The variants were filtered using in-house PERL scripts for (1) those not induced by EMS (other than G->A or C->T), although other types of SNPs may be induced; but it was not considered in this analysis mainly to improve the quality of SNP discovery (2) read depth outside the range of 10–150, (3) allele frequency less than 0.3, (4) genotype quality less than 30, and (5) any WT allele. Despite these stringent filters, few loci in the genome had usually high density of variants, mainly arising from the repetitive regions. Such variants were filtered (only from that region of chromosome 10 where the association was observed) by masking the repetitive regions. The sequence fragments in hard masked BTx623 reference genome were extracted and mapped back to BTx623-Rooney genome. The variants in the mapped region were used for further analysis.

Causal gene discovery. For the discovery of causal gene, the region of the genome linked to the phenotype in all mutant lines in the pool was searched as described in MutMap [15]. For that, the alternate allele frequency (AF), which is the ratio of the number of reads supporting the mutant allele to the total number of reads aligned, was plotted for each of the ten chromosomes. The SNPs from the region of the genome with AF = 1 in mutant pools were extracted (to avoid any putative causal SNP to be missed out, the filters described previously for SNP calling were relaxed for SNPs of this region and each case was examined manually) and were annotated using in-house Perl scripts to find out if any of them caused amino acid change or introduced a stop codon. The genes with either of the two changes were chosen as candidate genes and were subjected to further tests to find the causal gene. The information on the nine candidate genes was collected from the databases such as Phytozome 10.3, Gramene, Quick GO etc., and published literature on sorghum or other species [50, 51].

Leaf sampling, RNA extraction, and quality control

The fully expanded third leaf was sampled from 10:00 to 10:30 am and immediately frozen in liquid nitrogen. The leaf was ground in liquid nitrogen and total RNA was extracted using TRIZOL reagent following the manufacturer's instructions (Invitrogen, USA). RNA integrity was checked by running 1 µg of RNA on a 2% agarose gel in 1X TAE running buffer. Total RNA was treated with RQ1 RNase free DNase (Promega, USA) followed by phenol-chloroform purification. One microgram of the purified RNA was used as the template to synthesize cDNA through reverse transcription using a first strand cDNA synthesis kit (Roche Diagnostics, Germany). The concentration of cDNA was normalized to 100 ng/µl and used for PCR amplification using the primers specific to the candidate genes.

Gene expression analysis by semi-quantitative PCR

Semi-quantitative PCR with SYBR Green I Master mix (Roche Diagnostics, Germany) in a final reaction volume of 20 µl was performed with primers specific to the gene of interest (S1 Table). The EF1α was used as an internal control. Relative transcript abundance quantification was quantified based on band intensity using the Image J [<https://imagej.nih.gov/ij/>].

Supporting information

S1 Fig. A plot of alternate allele frequency (AF) in mutant pool along all chromosomes.

The X-axis shows the position in Mb units in respective chromosome whereas Y-axis shows the AF values.

(TIF)

S2 Fig. A focus on the target region of alternate allele frequency (AF) plot in the q arm of chromosome ten of the mutant. The X-axis shows the position in Mb units in respective

chromosome whereas Y-axis shows the AF values.
(TIF)

S3 Fig. An expression profile of Sobic.010G266000 an ortholog related with chloroplast based on the maize atlas [50].
(TIF)

S4 Fig. Profile of expression of Sobic.010G250100 in maize primordial tissues [51] based on maize atlas. Tissue name: FP = foliar primordial plastochron 1; FP34 = foliar primordial plastochron 3 or 4; FP5 = foliar primordial plastochron 5; FI = foliar immature; FE = foliar expanded.
(TIF)

S5 Fig. The expression profile of all the candidate genes based on maize atlas.
(TIF)

S6 Fig. Semi-quantitative PCR analysis of candidate genes of mutant 33 BC₁F₂ seedlings. The Sobic.010G239700, Sobic.010G249000, Sobic.010G250100, Sobic.010G264000, and Sobic.010G241900 showed no apparent changes in the transcript expression compared to the wild type. EF1 α is the housekeeping gene.
(TIF)

S1 Table. List of SNPs from a region of chromosome 10 with AF = 1 in the mutant pool with low $\delta^{13}\text{C}$ (SbPool 1) whereas AF < = 0.6 in WT pool with normal $\delta^{13}\text{C}$ (SbPool 4). The annotations of the genes are according to phytozome v.02.
(XLSX)

S2 Table. The detail of 37 SNPs in the region with high alternate frequency.
(XLSX)

S3 Table. List of primers specific to the gene of interest.
(XLSX)

Acknowledgments

We acknowledge the technical support of the following members of C₄ Rice Center at IRRI Richard Garcia, Nikki Larazo, Reychelle Mogul for the cultivation of sorghum and management of samples, Abraham Darius Llave for assisting bioinformatics team, Michael Alcasid for crossing, Ronald Tapia and Menard de la Rosa for sampling, Joana Mari Salonga and Irma Canicosa for gas exchange measurements, Abigail Mabilangan for microscopy, Kelvin Acebron for staff management, and Melvin for shipping materials across continents. We also acknowledge Hilary Stuart-Williams, Stable Isotopes Lab, Research School of Biology, Australian National University, Canberra, Australia for technical assistance with the dry matter carbon isotope composition analysis.

Author Contributions

Conceptualization: John E. Sheehy, Robert Furbank, Susanne von Caemmerer, William Paul Quick.

Data curation: Govinda Rizal, Shanta Karki, Vivek Thakur, Samart Wanchana, Hugo Alonso-Cantabrana.

- Formal analysis:** Govinda Rizal, Shanta Karki, Vivek Thakur, Samart Wanchana, Hugo Alonso-Cantabrana, Susanne von Caemmerer.
- Funding acquisition:** John E. Sheehy, Robert Furbank, Susanne von Caemmerer, William Paul Quick.
- Investigation:** Govinda Rizal, Shanta Karki, Vivek Thakur, Samart Wanchana, Hugo Alonso-Cantabrana, Jacque Dionora.
- Methodology:** Govinda Rizal, Shanta Karki, Vivek Thakur, Hugo Alonso-Cantabrana, Jacque Dionora.
- Project administration:** John E. Sheehy, Robert Furbank, Susanne von Caemmerer, William Paul Quick.
- Resources:** Govinda Rizal, Shanta Karki, Vivek Thakur, Hugo Alonso-Cantabrana, Jacque Dionora, John E. Sheehy, Robert Furbank, Susanne von Caemmerer, William Paul Quick.
- Software:** Govinda Rizal, Shanta Karki, Vivek Thakur, Samart Wanchana.
- Supervision:** Robert Furbank, Susanne von Caemmerer, William Paul Quick.
- Validation:** Govinda Rizal, Shanta Karki, Vivek Thakur, Hugo Alonso-Cantabrana, Jacque Dionora.
- Visualization:** Govinda Rizal, Shanta Karki, Vivek Thakur, Hugo Alonso-Cantabrana.
- Writing – original draft:** Govinda Rizal, Shanta Karki, Vivek Thakur, Hugo Alonso-Cantabrana.
- Writing – review & editing:** Govinda Rizal, Shanta Karki, Vivek Thakur, Hugo Alonso-Cantabrana, Jacque Dionora, Robert Furbank, Susanne von Caemmerer, William Paul Quick.

References

1. Farquhar GD, O'Leary MH, Berry JA. On the relationship between carbon isotope discrimination and the intercellular carbon dioxide concentration in leaves. *Funct Plant Biol.* 1982 April 01; 9(2):121–37.
2. Farquhar GD, Ehleringer IJR, Hubick KT. Carbon isotope discrimination and photosynthesis. *Ann Rev Plant Physiol Plant Mol Biol.* 1989 Jan; 40:503–37.
3. O'Leary MH. Carbon isotope fractionation in plants. *Phytochem.* 1981; 20(4):553–67.
4. von Caemmerer S, Ghannoum O, Pengelly JLL, Cousins AB. Carbon isotope discrimination as a tool to explore C₄ photosynthesis. *J Exp Bot.* 2014 April; 06: 65(13):3459–30. PMID: [24711615](https://doi.org/10.1093/jxb/eru127). <https://doi.org/10.1093/jxb/eru127>.
5. Alonso-Cantabrana H, von Caemmerer S. Carbon isotope discrimination as a diagnostic tool for C₄ photosynthesis in C₃-C₄ intermediate species. *J Exp Bot.* 2016 May; 67(10):3109–21. PMID: [26862154](https://doi.org/10.1093/jxb/erv555), <https://doi.org/10.1093/jxb/erv555>
6. von Caemmerer S. 1992. Carbon isotope discrimination in C₃-C₄ intermediates. *Plant Cell Env.* 1992 Dec; 15(9):1063–72. <https://doi.org/10.1111/j.1365-3040.1992.tb01656.x>
7. Farquhar GD, Richards R. Isotopic composition of plant carbon correlates with water-use efficiency of wheat genotypes. *Funct Plant Biol.* 1984; 11(6):539–52.
8. Farquhar GD, Ehleringer IJR, Hubick KT. Carbon isotope discrimination and photosynthesis. *Ann Rev Plant Physiol Plant Mol Biol.* 1989 Jan; 40:503–37.
9. Henderson SA, von Caemmerer S, Farquhar GD. Short-term measurements of carbon isotope discrimination in several C₄ species. *Funct Plant Biol.* 1992 January; 19(3):263–85.
10. Cerling TE, Wang Y, Quade J. Expansion of C₄ ecosystems as an indicator of global ecological change in the late Miocene. *Nature.* 1993 January; 361:344–5. <https://doi.org/10.1038/361344a0>
11. Gresset S, Westermeier P, Rademacher S, Ouzunova M, Presterl T, Westhoff P, Schon CC. Stable carbon isotope discrimination is under genetic control in the C₄ species maize with several genomic regions influencing trait expression. *Plant Physiol.* 2013 Nov 26; 164(1):131–43. <http://dx.doi.org/10.1104/pp.113.224816>.

12. Rizal G, Thakur V, Dionora J, Karki S, Wanchana S, Acebron K, et al. Two forward genetic screens for vein density mutants in sorghum converge on a Cytochrome P450 gene in the brassinosteroid pathway. *Plant J*. 2015 Oct; 84(2):257–66. PMID: 26333774. <https://doi.org/10.1111/tpj.13007>
13. Mochida K, Shinozaki K. Advances in omics and bioinformatics tools for systems analyses of plant functions. *Plant and Cell Physiol*. 2011 Dec; 52(12):2017–2038. <https://doi.org/10.1093/pcp/pcr153> PMID: 22156726
14. Abe A, Kosugi S, Yoshida K, Natsume S, Takagi H, Kanzaki H, et al. Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat Biotechnol*. 2012 Jan 22; 30(2):174–8. PMID: 22267009. <https://doi.org/10.1038/nbt.2095>
15. Ku SB, Edwards GE. Oxygen inhibition of photosynthesis in the C₄ species *Amaranthus graecizans* L. *Planta*. 1980 Jan; 147(4):227–82.
16. Wang P, Kelly S, Fouracre JP, Langdale JA. Genome-wide transcript analysis of early maize leaf development reveals gene cohorts associated with the differentiation of C₄ Kranz anatomy. *Plant J*. 2013 Aug; 75(4):656–70. PMID: 23647263, <https://doi.org/10.1111/tpj.12229>
17. Hall TM. Multiple modes of RNA recognition by zinc finger proteins. *Curr Opin Struct Biol*. 2005; 15(3):367–73. PMID: 15963892. <https://doi.org/10.1016/j.sbi.2005.04.004>
18. Hubick KT, Hammer GL, Farquhar GD, Wade LJ, von Caemmerer S, Henderson SA. Carbon isotope discrimination varies genetically in C₄ species. *Plant Physiol*. 1990 Feb; 92(2):534–7. PMID: 16667310,
19. Gresset S, Westermeier P, Rademacher S, Ouzunova M, Presterl T, Westhoff P, Schon CC. Stable carbon isotope discrimination is under genetic control in the C₄ species maize with several genomic regions influencing trait expression. *Plant Physiol*. 2013 Nov 26; 164(1):131–43. <http://dx.doi.org/10.1104/pp.113.224816>.
20. Ashburner M, Ball CA, Blake JA, Bostein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. *Nature Genet*. 2000 May; 25(1):25–9. <https://doi.org/10.1038/75556> PMID: 10802651
21. Thomas PD, Campbell MJ, Kejariwal A. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res*. 2003 Sep; 13(9):2129–41. PMID: 12952881, <https://doi.org/10.1101/gr.772403>
22. Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res*. 2013 Jan; 41(Database issue):D377–86. <https://doi.org/10.1093/nar/gks1118> PMID: 23193289
23. Mi H, Poudel S, Muruganujan A, Casagrande JT, Thomas PD. PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic Acids Res*. 2016 Jan; 44(Database issue): D336–42. <https://doi.org/10.1093/nar/gkv1194> PMID: 26578592
24. Burdiak P, Rusaczonok A, Witon D, Gow D, Karpinski S. Cysteine-rich receptor-like kinase CRK5 as a regulator of growth, development, and ultraviolet radiation responses in *Arabidopsis thaliana*. *J Exp Bot*. 2015 Jun; 66(11):3325–37. PMID: 25969551, <https://doi.org/10.1093/jxb/erv143>
25. Chen Z. A superfamily of proteins with novel Cysteine-rich repeats. *Plant Physiol*. 2001 Jun; 126(2):473–6. PMID: 11402176,
26. Mitchell AJ, Dunham NP, Bergman JA, Wang B, Zhu Q, Chang W, Liu X, Boal AK. Structure-guided reprogramming of a hydroxylase to halogenate its small molecule substrate. *Biochem*. 2016 Dec 28; 56(3): 441–44. <https://doi.org/10.1021/acs.biochem.6b01173> PMID: 28029241
27. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytosome: a comparative platform for green plant genomics. *Nucl Acids Res*. 2012 Jan; 40(Database issue): D1178–86. PMID: 22110026, <https://doi.org/10.1093/nar/gkr944>
28. Torii KU, Mitsukawa N, Oosumi T, Matsuura Y, Yokoyama R, Whittier RF, Komeda Y. The *Arabidopsis ERECTA* gene encodes a putative receptor protein kinase with extracellular leucine-rich repeats. *Plant Cell*. 1996 Apr; 8(4):735–46. PMID: 8624444, <https://doi.org/10.1105/tpc.8.4.735>
29. Diévarit A, Gilbert N, Droc G, Attard A, Gourgues M, Guiderdoni E, Périn C. Leucine-rich repeat receptor kinases are sporadically distributed in eukaryotic genomes. *BMC Evol Biol*. 2011 Dec 20; 11:367. PMID: 22185365, <https://doi.org/10.1186/1471-2148-11-367>
30. Roux M, Schwessinger B, Albrecht C, Chinchilla D, Jones A, Holton N, Malinovsky FG, Tör M, de Vries S, Zipfel C. The *Arabidopsis* leucine-rich repeat receptor-like kinases BAK1/SERK3 and BKK1/SERK4 are required for innate immunity to hemibiotrophic and biotrophic pathogens. *Plant Cell*. 2011 Jun; 23(6):2440–55. <http://dx.doi.org/10.1105/tpc.111.084301>. PMID: 21693696
31. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, et al. The B73 maize genome: Complexity, diversity, and dynamics. *Science*. 2009 Nov 20; 326(5956):1112–5. PMID: 19965430. <https://doi.org/10.1126/science.1178534>

32. Collins NC, Webb CA, Seah S, Ellis JG, Hulbert SH, Pryor A. The isolation and mapping of disease resistance gene analogs in maize. *Mol Plant Microbe Interact.* 1998 Oct; 11(10):968–78. PMID: [9768514](https://pubmed.ncbi.nlm.nih.gov/9768514/). <https://doi.org/10.1094/MPMI.1998.11.10.968>
33. Sun CW, Chen LJ, Lin LC, Li HM. Leaf-specific upregulation of chloroplast translocon genes by a CCT motif-containing protein, CIA2. *Plant Cell.* 2001 Sep; 13(9):2053–61. PMID: [11549763](https://pubmed.ncbi.nlm.nih.gov/11549763/)
34. Sun CW, Huang YC, Chang HY. CIA2 coordinately up-regulates protein import and synthesis in leaf chloroplasts. *Plant Physiol.* 2009 Jun; 150(2):879–88. PMID: [19386807](https://pubmed.ncbi.nlm.nih.gov/19386807/), <https://doi.org/10.1104/pp.109.137240>
35. Contreras-Cornejo HA, López-Bucio JS, Méndez-Bravo A, Macías-Rodríguez L, Ramos-Vega M, Guevara-García AA, López-Bucio J. Mitogen-activated protein kinase 6 and ethylene and auxin signaling pathways are involved in Arabidopsis root-system architecture alterations by *Trichoderma atroviride*. *Mol Plant Microbe Interact.* 2015 Jun; 28(6):701–10. PMID: [26067203](https://pubmed.ncbi.nlm.nih.gov/26067203/). <https://doi.org/10.1094/MPMI-01-15-0005-R>
36. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. *Nucl Acids Res.* 2016 Jan 04; 44 (Database issue):D279–85. PMID: [26673716](https://pubmed.ncbi.nlm.nih.gov/26673716/), <https://doi.org/10.1093/nar/gkv1344>
37. Klug A. Zinc finger peptides for the regulation of gene expression. *J Mol Biol.* 1999 Oct; 293(2):215–18. PMID: [10529348](https://pubmed.ncbi.nlm.nih.gov/10529348/). <https://doi.org/10.1006/jmbi.1999.3007>
38. Gamsjaeger R, Liew CK, Loughlin FE, Crossley M, Mackay JP. Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends Biochem Sci.* 2007 Feb; 32(2):63–0. PMID: [17210253](https://pubmed.ncbi.nlm.nih.gov/17210253/). <https://doi.org/10.1016/j.tibs.2006.12.007>
39. Pengelly JLL, Sirault XRR, Tazoe Y, Evans JR, Furbank RT, von Caemmerer S. Growth of the C₄ dicot *Flaveria bidentis*: photosynthetic acclimation to low light through shifts in leaf anatomy and biochemistry. *J Exp Bot.* 2010 Sept; 61(14):4109–22. <https://doi.org/10.1093/jxb/erq226> PMID: [20693408](https://pubmed.ncbi.nlm.nih.gov/20693408/)
40. Schultz HR, Kiefer W, Gruppe W. Photosynthetic duration, carboxylation efficiency and stomatal limitation of sun and shade leaves of different ages in field-grown grapevine (*Vitis vinifera* L.). *Vitis* 1996; 35(4):169–176.
41. Zeinalov Y, Maslenkova L. Estimation of the quantum efficiency of photosynthesis. I. Theoretical ground and experimental approaches. *Bulg J Plant Physiol.* 1999; 25(1–2): 26–38.
42. Ling Q, Huang W, Jarvis P. Use of a SPAD-502 meter to measure leaf chlorophyll concentration in *Arabidopsis thaliana*. *Photosynth Res.* 2011 Feb; 107(2):209–14. doi: [10.1007/s11120-010-9606-0](https://doi.org/10.1007/s11120-010-9606-0). Epub 2010 Dec 28. PMID: [21188527](https://pubmed.ncbi.nlm.nih.gov/21188527/)
43. Rizal G, Karki S, Alcasid M, Montecillo F, Acebron K, Larazo N, et al. Shortening the breeding cycle of sorghum, a model crop for research. *Crop Sci.* 2014 Mar-Apr; 54:520–29. <https://doi.org/10.2135/cropsci2013.07.0471>
44. Murray MG, Thompson WF. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 1980 Oct 10; 8(19): 4321–25.
45. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012 Mar 04; 9(4):357–9. doi: [10.1038/nmeth.1923](https://doi.org/10.1038/nmeth.1923). PMID: [22388286](https://pubmed.ncbi.nlm.nih.gov/22388286/).
46. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE.* 2014 Nov 19; 9(11):e112963. PMID: [25409509](https://pubmed.ncbi.nlm.nih.gov/25409509/), <https://doi.org/10.1371/journal.pone.0112963>
47. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014 Aug 1; 30(15):2114–20. PMID: [24695404](https://pubmed.ncbi.nlm.nih.gov/24695404/), <https://doi.org/10.1093/bioinformatics/btu170>
48. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment / map format and SAMtools. *Bioinformatics.* 2009 Aug 15; 25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352> PMID: [19505943](https://pubmed.ncbi.nlm.nih.gov/19505943/)
49. DePristo MA, Banks E, Poplin RE, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011 May; 43(5):491–8. PMID: [21478889](https://pubmed.ncbi.nlm.nih.gov/21478889/), <https://doi.org/10.1038/ng.806>
50. Downs GS, Bi YM, Colasanti J, Wu W, Chen X, Zhu T, Rothstein SJ, Lukens LN. A developmental transcriptional network for maize defines coexpression modules. *Plant Physiol.* 2013 Apr; 161(4):1830–43. PMID: [23388120](https://pubmed.ncbi.nlm.nih.gov/23388120/), <https://doi.org/10.1104/pp.112.213231>
51. Wang P, Kelly S, Fouracre JP, Langdale JA. Genome-wide transcript analysis of early maize leaf development reveals gene cohorts associated with the differentiation of C4 Kranz anatomy. *Plant J.* 2013 May; 75:656–70. <https://doi.org/10.1111/tpj.12229> PMID: [23647263](https://pubmed.ncbi.nlm.nih.gov/23647263/)