

Two forward genetic screens for vein density mutants in sorghum converge on a cytochrome P450 gene in the brassinosteroid pathway

Govinda Rizal^{1,†}, Vivek Thakur^{1,†}, Jacqueline Dionora^{1,†}, Shanta Karki¹, Samart Wanchana¹, Kelvin Acebron¹, Nikki Larazo¹, Richard Garcia¹, Abigail Mabilangan¹, Florencia Montecillo¹, Florence Danila¹, Reychele Mogul¹, Paquito Pabloco¹, Hei Leung², Jane A. Langdale³, John Sheehy^{1,4}, Steven Kelly^{3,*} and William Paul Quick^{1,5}

¹C₄ Rice Center, International Rice Research Institute, DAPO Box 7777, Metro Manila, Philippines,

²Plant Breeding, Genetics, and Biotechnology Division, International Rice Research Institute, DAPO Box 7777, Metro Manila, Philippines,

³Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK,

⁴12 Barley Way, Marlow SL7 2UG, UK, and

⁵Department of Plant and Animal Sciences, University of Sheffield, Sheffield, UK

Received 1 April 2015; revised 30 July 2015; accepted 11 August 2015; published online 2 September 2015.

*For correspondence (e-mail steven.kelly@plants.ox.ac.uk).

†These authors contributed equally to this work.

SUMMARY

The specification of vascular patterning in plants has interested plant biologists for many years. In the last decade a new context has emerged for this interest. Specifically, recent proposals to engineer C₄ traits into C₃ plants such as rice require an understanding of how the distinctive venation pattern in the leaves of C₄ plants is determined. High vein density with Kranz anatomy, whereby photosynthetic cells are arranged in encircling layers around vascular bundles, is one of the major traits that differentiate C₄ species from C₃ species. To identify genetic factors that specify C₄ leaf anatomy, we generated ethyl methanesulfonate- and γ -ray-mutagenized populations of the C₄ species sorghum (*Sorghum bicolor*), and screened for lines with reduced vein density. Two mutations were identified that conferred low vein density. Both mutations segregated in backcrossed F₂ populations as homozygous recessive alleles. Bulk segregant analysis using next-generation sequencing revealed that, in both cases, the mutant phenotype was associated with mutations in the *CYP90D2* gene, which encodes an enzyme in the brassinosteroid biosynthesis pathway. Lack of complementation in allelism tests confirmed this result. These data indicate that the brassinosteroid pathway promotes high vein density in the sorghum leaf, and suggest that differences between C₄ and C₃ leaf anatomy may arise in part through differential activity of this pathway in the two leaf types.

Keywords: brassinosteroid, *Sorghum bicolor*, C₄ photosynthesis, gene identification, vascular patterning, kranz anatomy.

INTRODUCTION

Most plants fix atmospheric carbon dioxide (CO₂) into a three-carbon compound as the first product of photosynthesis, and are thus referred to as C₃ plants. In contrast, the first product of photosynthesis in C₄ plants is a four-carbon compound. Compared with C₃ photosynthesis, the more recently evolved C₄ pathway often confers increased productivity, especially in hot, arid conditions (Ludlow, 1985).

The evolution of C₄ photosynthesis from ancestral C₃ photosynthesis involved multiple changes to leaf anatomy and biochemistry (Monson, 1999). Despite these numerous

differences, C₄ photosynthesis evolved independently in over 60 different plant lineages (Sage *et al.*, 2011), and thus represents one of the most remarkable examples of convergent evolution in eukaryotic biology. The increased productivity of C₄ plants is due to the fact that they concentrate CO₂ around Rubisco, effectively eliminating the competitive (and ultimately energy-consuming) oxygenation reaction of the enzyme (Chollet and Ogren, 1975; Hatch and Osmond, 1976). In most C₄ species, the CO₂-concentrating mechanism is facilitated through spatial partitioning of the photosynthetic apparatus into distinct mesophyll (M) and bundle

sheath (BS) cells (Slack *et al.*, 1969), but in a small number of lineages, spatial partitioning occurs within individual cells (Voznesenskaya *et al.*, 2001).

The two-cell-type C_4 species exhibit a characteristic leaf anatomy known as Kranz anatomy, which is defined by concentric wreaths of BS and M cells surrounding the leaf veins (Haberlandt, 1896; Brown, 1975). In grasses, these vascular wreaths are iterated across the breadth of the fully developed leaf such that each pair of veins is separated by one or two layers of BS cells and a small number of chlorenchymatous M cells (Hattersley and Watson, 1975; Dengler *et al.*, 1994). In C_4 panicoid grasses such as sorghum (*Sorghum bicolor*) and maize (*Zea mays*), there is only one BS cell layer and a single M cell layer, such that each pair of veins (V) is separated by only four cells (V-BS-M-M-BS-V). This anatomical organization is in stark contrast to that observed in C_3 panicoid leaves, where the veins may be separated by many more M cells. Our knowledge of the genetic factors that contribute to the specification of high vein density in C_4 leaves is limited. Within C_4 species, loss-of-function mutants in the *SCARECROW* (*SCR*) and *SHORTROOT* (*SHR*) genes of maize exhibit some reduction in vein density (Slewisinski *et al.*, 2012, 2014). The regulatory roles of *SCR* and *SHR* have also been inferred through transcriptomic and systems biology approaches (Wang *et al.*, 2013; Fouracre *et al.*, 2014). These latter studies propose a gene regulatory network for Kranz development that includes additional members of the *SCR/SHR* pathway. Although no other factors have been identified that regulate vein density in C_4 plants, experimental investigation of the C_3 plant *Arabidopsis thaliana* has implicated the brassinosteroid (BR) signaling pathway in the regulation of C_3 vascular density. Specifically, *Arabidopsis* mutants that over-express components of the BR signaling pathway exhibit an increased number of vascular bundles in stems, while those that are deficient in BR signaling have reduced numbers of vascular bundles in leaves (Ibañez *et al.*, 2009).

Here we used a forward-genetic approach to identify genes that modulate vein density in the C_4 monocot *Sorghum bicolor*. Two independent mutant populations were generated using ethyl methanesulfonate (EMS) or gamma (γ) irradiation, and plants were screened for altered vein spacing in the leaf. Two low-vein-density (LVD) mutants, one from each mutant population, were selected for further characterization. Using bulk segregant analysis and next-generation sequencing, we determined that, in both cases, the causal mutations affected the integrity of *CYP90D2*, a gene that encodes an enzyme involved in BR biosynthesis (Hong *et al.*, 2003). We further validated the causative mutations by showing that the mutations were allelic when the two lines were crossed. Our findings provide evidence that the BR signaling pathway plays a role in specifying venation patterns in C_4 leaves.

RESULTS

Identification of LVD mutants in sorghum

To generate mutant populations for forward-genetic screens in sorghum, two independent mutagenesis experiments were performed. In the first, 1 million seeds of line BTx623-Rooney (BTx623-R, WT-R) were treated with EMS, and in the second, 0.6 million seeds of line BTx623-Zhangguo (BTx623-Z, WT-Z) were treated with γ irradiation. Approximately 17 000 M_2 lines from each mutagenesis treatment were screened for altered leaf vein density, and two LVD mutants, mutant 8 (γ -M8) and mutant 30 (EMS-M30), were chosen from the γ and EMS populations, respectively, for further analysis (Figure 1a). The raw data, means and standard deviations for all measurements on all plants used in this study are provided in Table S1. The mean vein densities in leaves of γ -M8 and EMS-M30 were five and six veins per mm, respectively; representing a reduction of 28 and 33% in the number of veins per mm compared to parental non-mutagenized controls (Figure 1b–e and Table 1). The LVD phenotype of both mutants was accompanied by a substantial reduction in plant height (Figure 1a and Table 1) and a distal shift of leaf sheath tissue into the blade (Figure 1f). Quantitative analysis of leaf cross-sections showed that the reduction in vein density was due to an increase in both the number and size of M cells between veins (Figure 1g–l). On average, the number of cells between minor veins increased by one (Figure 1k), with each M cell containing chloroplasts (Figure S1) and exhibiting a larger cross-sectional area than those in the wild-type (WT) (Figure 1l). The size, shape, area and number of BS cells also showed a significant increase (Figure 1m,n). The size of epidermal cells is also increased (Table S1). As a corollary of the increase in cell size, both mutants exhibited thicker leaves (Figure 1h, j). While the light compensation point and quantum efficiency of the LVD mutants were equivalent to those of WT plants, both mutants exhibited decreased maximum CO_2 assimilation rate and carboxylation efficiency (Figure S2). The lower photosynthetic rate is consistent with the distal shift of leaf sheath tissue into the blade, as sheath tissue has a lower photosynthetic rate than blade (Pengelly *et al.*, 2011).

Generation of a segregating population for bulked segregant analysis

For bulked segregant analysis, homozygous mutant lines were crossed to their respective WT parental strains. The F_1 progeny of these crosses were selfed to produce the F_2 generation, and the vein density in the F_2 progeny was assessed. For both populations, the ratio of progeny exhibiting LVD versus WT vein density (WVD) was 1:3, indicative of a single recessive allele (Figure S3a,b). For both mutants, the variation in vein density in the F_2 lines



Figure 1. Phenotype of LVD mutants γ -M8 and EMS-M30.

(a) Whole-plant phenotype for mutants and corresponding WT progenitor lines. Photographs were taken 60 days after sowing.

(b–e) Images of the leaf surface of (b) WT-Z, (c) WT-R, (d) γ -M8 and (e) EMS-M30.

(f) Images of plants, with arrows indicating proximal shift of leaf sheath tissue into the blade.

(g–j) Transverse leaf sections of (g) WT-Z, (h) γ -M8, (i) WT-R and (j) EMS-M30.

(k) Number of M cells between veins.

(l) Area of the M cells.

(m) Number of bundle sheath cells (BSC) around veins.

(n) Area of the bundle sheath cells (BSC).

Asterisks indicate significant differences from the respective wild-types (tested using Student's *t*-test) at a confidence level of 95%. Error bars indicate standard deviation. The number of samples (*n*) is 10 in all cases.

Table 1 Vein density and plant heights of γ -M8 and EMS-M30 mutants

Plant line (generation, parent)	Number of veins per mm (means $\pm$ SD)	Plant height 30 days after sowing (cm)
γ -M8 (M ₂ , BTx623-Z)	5.30 $\pm$ 0.11	64
BTx623-Z	7.12 $\pm$ 0.33	111
EMS-M30 (M ₂ , BTx623-R)	6.15 $\pm$ 0.41	73
BTx623-R	9.45 $\pm$ 0.44	231

was such that the range of values observed for the LVD and WVD populations marginally overlapped (Figure S3a, b). To ensure that the LVD and WVD individuals in each F₂ population were correctly segregated prior to sequencing, a stringent probabilistic classification method was applied (see Experimental procedures). For γ -M8, pooled DNA samples were generated from 30 LVD and 30 WVD F₂ individuals, and samples from 28 individuals of each type were pooled for EMS-M30 (Tables S2 and S3).

A single nucleotide change in EMS-M30 results in a premature stop codon in the *CYP90D2* gene

To identify the chromosomal region linked to the causative EMS-M30 mutation, paired-end read libraries were generated from the pooled DNA samples using Illumina protocols (Table S4). Reads were mapped to the appropriate reference genome (see Experimental procedures), and then the distribution of likely EMS mutations (i.e. C→T conversions) in the genome was analyzed. Using this approach, the frequency of mutations in the LVD pooled sample should be approximately 0.5 in genomic regions that are not linked to the causative mutation, and should tend towards 1 in regions that are linked (Abe *et al.*, 2012). A plot of the mutation frequencies along each chromosome clearly identified a set of variants linked to the mutant allele within the first 10 Mb region of chromosome 3 (Figure 2a and Figure S4). All SNPs within this

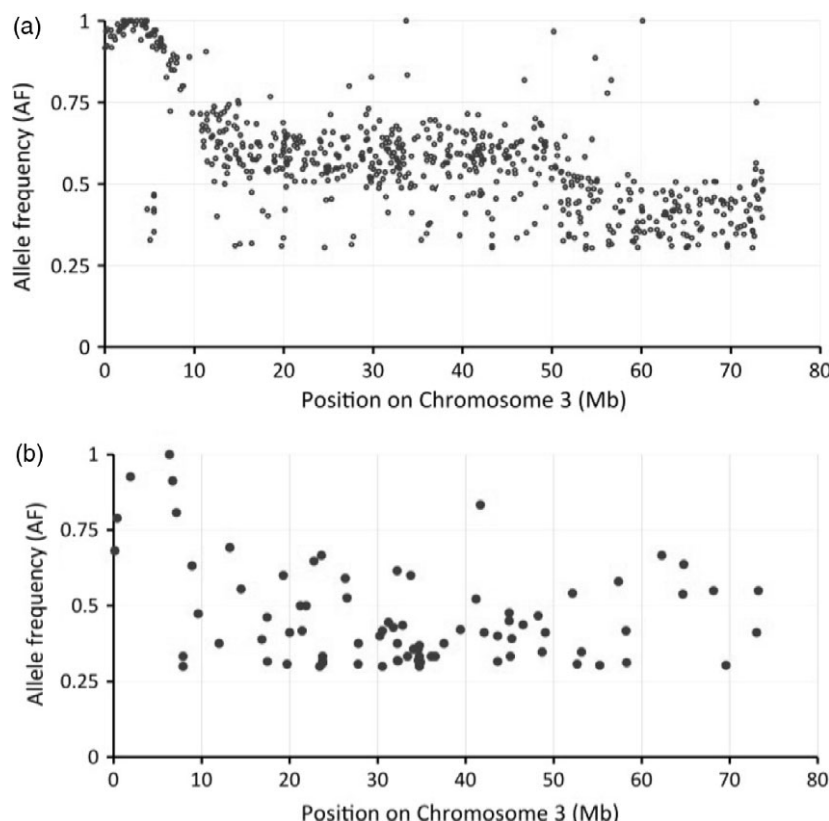


Figure 2. Mutant allele frequencies in chromosome 3.

(a) Plot of mutant (alternate) allele frequency of chromosome 3 variants in the LVD pool of EMS-M30.

(b) Allele frequency plot of chromosome 3 variants in the LVD pool of γ -M8.

region were therefore annotated, and investigated to discover potential causative mutations. A total of 41 SNPs were observed within this window, of which three caused changes to predicted coding sequences and two occurred near coding sequences, i.e. were potentially located in UTRs or regulatory regions (Table S5). All five of these coding sequence-associated mutations were independently validated by PCR and Sanger sequencing (Table S6). Only one of the three coding sequence mutations resulted in a change in the encoded protein. This mutation led to a premature stop codon in the gene *CYP90D2* (Sobic.003G030600), and was located near the center of the peak in the allele frequency plot.

A chromosome inversion in γ -M8 disrupts the *CYP90D2* gene

To identify the chromosomal region linked to the causative γ -M8 mutation, Illumina sequencing and read mapping were performed as for EMS-M30. In this case, however, when the genome-wide distribution of mutations was analyzed, a wider range of mutation types was found (Figure S5). As observed for EMS-M30, in the LVD pool only the first 10 Mb region of chromosome 3 contained mutations ($n = 7$) that had allele frequencies typical of a region linked to a causal mutation (Figure 2b and Table S7). Analysis of these seven SNPs showed that five

occurred in intergenic regions, one was located in an intron, and one was located in a 3' UTR (Table S7). None of the variants were predicted to cause changes to encoded protein sequences. However, in addition to these SNPs, a structural mutation was identified that comprised an approximately 3.3 Mb inversion of chromosome 3 and an inter-chromosomal translocation of approximately 3 kb from chromosome 1 to chromosome 3 (Figure 3a,b). These structural changes were predicted to disrupt three genes, namely *CYP90D2* (Sobic.003G030600), Sobic.001G154100 and Sobic.003G070400 (Table S8). The terminal exon and 3' UTR were separated from the rest of the *CYP90D2* gene, a small portion of the 5' UTR of Sobic.003G070400 was lost, and Sobic.001G154100 was split into two equally sized partial gene fragments (Figure 3a,b). To validate these structural changes, primer sets were designed to flank and amplify the predicted breakpoints (labeled 1, 3, 5 and 6 in Figure 3a,b) and the corresponding non-mutagenized reference genome (labeled 2 and 4). PCR amplification of genomic DNA from parental WT, parental γ -M8 and F_2 progeny of a WT $\times$ γ -M8 cross confirmed the translocation and inversion events (Figure 3c) in the parental γ -M8 line. Notably, however, the deletion on chromosome 1 that was caused by the translocation event (captured by primer pair 1, Figure 3c) was repaired in the F_2 individual selected for this analysis.

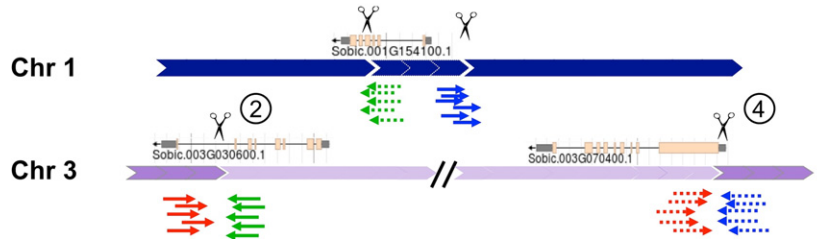
Figure 3. Structural variants in LVD pool of γ -M8 mutants.

(a) Diagram showing the location in the genome of the chromosomal rearrangements supported by the read data. Three sets of paired reads (red, blue and green) support the location of these rearrangements.

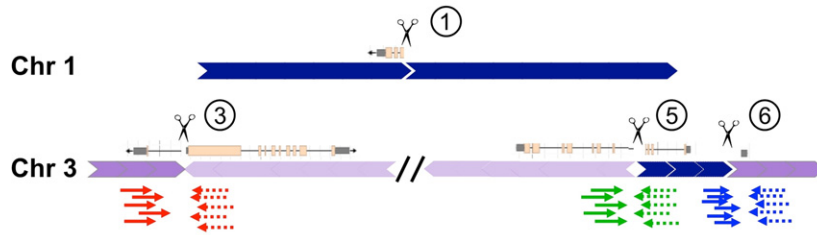
(b) The reconstructed mutant genome, inferred from this read data. The genes in the vicinity of the chromosomal rearrangements are shown with peach blocks indicating exons and grey blocks and lines indicating UTR and introns respectively. The numerical labels at the location of the rearrangements indicate those that were verified experimentally by PCR. Each number corresponds to a primer pair.

(c) PCR gel image validating the structural variant breakpoints shown in Figure 3a–b. Primer set 1 was designed to amplify a small fragment spanning the region of the deletion on chromosome 1. It was not expected to amplify the reference genome as the predicted fragment would be too large. Primer pairs 2 and 4 were designed to amplify chromosome 3 fragments from the reference genome only, whereas primer pairs 3, 5 and 6 were designed to amplify the region spanning the rearrangement in the mutant samples (γ -M8 and the LVD F_2 segregant). Numbers above the gel lanes indicate primer pair used; W = BTx-623-Z; M = γ -M8; F_2 = LVD plant from F_2 population; L = DNA size ladder.

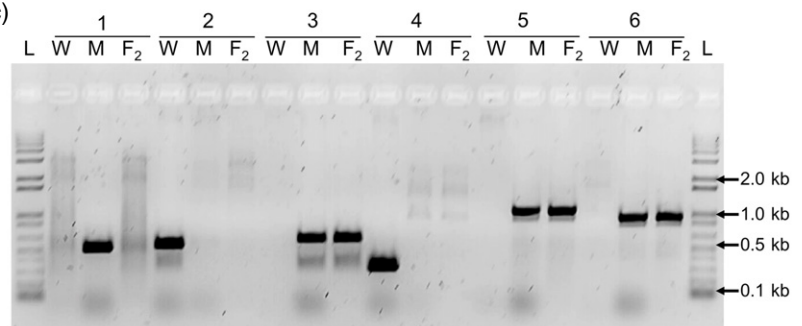
(a) Read alignment on reference genome



(b) Read alignment on reconstructed mutant genome



(c)



Allelism confirms that LVD in both γ -M8 and EMS-M30 is caused by mutations in the CYP90D2 gene

To confirm that γ -M8 and EMS-M30 mutant phenotypes were caused by mutations in the same gene, homozygous mutants were crossed and the F_1 progeny assessed (Figure 4a). F_1 plants were first screened by PCR using a polymorphic marker that distinguishes the two parental lines (Figure 4b). Seventy-five PCR-positive F_1 hybrid plants were then examined for the LVD phenotype. The mean vein densities in leaves of homozygous γ -M8, homozygous EMS-M30, and F_1 progenies of a γ -M8 $\times$ EMS-M30 cross and of an EMS-M30 $\times$ γ -M8 cross were not significantly different from each other (Table 2). As these F_1 plants represent crosses between two different parental lines, appropriate control crosses were also performed with WT lines BTx623-R and BTx623-Z. The progeny of control crosses between the WT lines exhibited WVD (Table 2). Heterosis in plant height was more pronounced in WT lines than in mutant lines, presumably due to the

decreased variance in plant height observed in the mutant lines (Table 2).

Phylogenetic analysis demonstrates that sorghum CYP90D2 is orthologous to rice CYP90D2

To determine phylogenetic relationships between the identified sorghum CYP90D2 gene and homologous sequences from other monocot and dicot species, a phylogenetic tree was generated. The tree demonstrates that the gene belongs to the cytochrome P450 superfamily, and that a gene duplication event that preceded diversification of the grasses resulted in the presence of two copies of this gene in each of the grass species used in this analysis (Figure 5). In rice (*Oryza sativa*), these duplicated genes (Os01 g10040 and Os05 g11130) have been named CYP90D2 and CYP90D3, respectively (Hong *et al.*, 2003), and have been shown to encode enzymes that catalyze intermediate steps in BR biosynthesis (Kim *et al.*, 1998, 2005; Hong *et al.*, 2003).

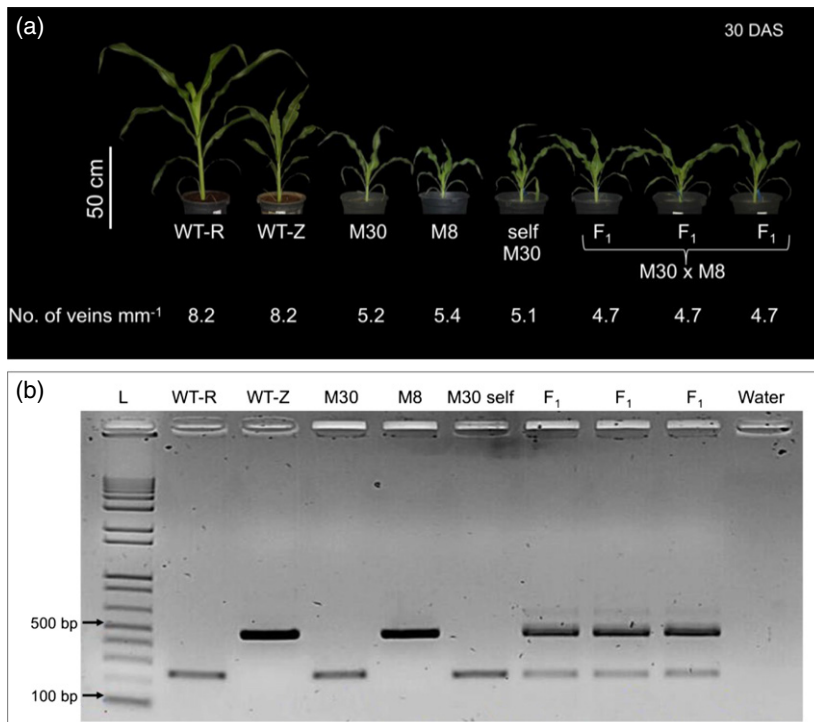


Figure 4. γ -M8 and EMS-M30 mutants are allelic. (a) Plant phenotype of WT, mutants and the F_1 population derived from a cross between γ -M8 and EMS-M30 mutants. (b) PCR verification of F_1 plants using polymorphic primers for WT-R and WT-Z. One of the F_1 plants derived from the crossed panicle was found to have self-pollinated, as verified by PCR (labeled 'self M30').

Table 2 Vein density and height of the mutants and WT, and F_1 progeny of crosses between mutants and and between WT lines

Genotype	Number of veins per mm (means $\pm$ SD)	Plant height 60 days after sowing (cm, means $\pm$ SD)
EMS-M30	5.4 $\pm$ 0.4	87.2 $\pm$ 12.0
γ -M8	5.2 $\pm$ 0.2	64.8 $\pm$ 7.0
F_1 (EMS-M30 $\times$ γ -M8)	4.7 $\pm$ 0.3	79.4 $\pm$ 4.0
F_1 (γ -M8 $\times$ EMS-M30)	4.7 $\pm$ 0.4	86.4 $\pm$ 8.0
BTx623-R	8.2 $\pm$ 0.7	266.7 $\pm$ 19.0
BTx623-Z	8.2 $\pm$ 0.6	116.6 $\pm$ 20.0
F_1 (BTx623-Z $\times$ BTx623-R)	7.7 $\pm$ 0.6	308.2 $\pm$ 30.0

Measurements were taken 30 days after sowing. The correlation coefficient between vein density and height (Pearson's correlation, r) was 0.75.

DISCUSSION

The specification of venation patterns in which two M cells separate each vascular bundle is a defining characteristic of all two-cell-type C_4 grasses. However, the genetic components that facilitate this patterning are largely unknown. Here, the results of two independent mutagenesis experiments converged to reveal a role for the *CYP90D2* gene in the specification of vein spacing in the C_4 plant sorghum. Both EMS and γ -rays were used as mutagens to maximize the types of mutations obtained. The fact that two comprehensive screens identified the same gene suggests that there may be a genetic constraint with regard to the way in

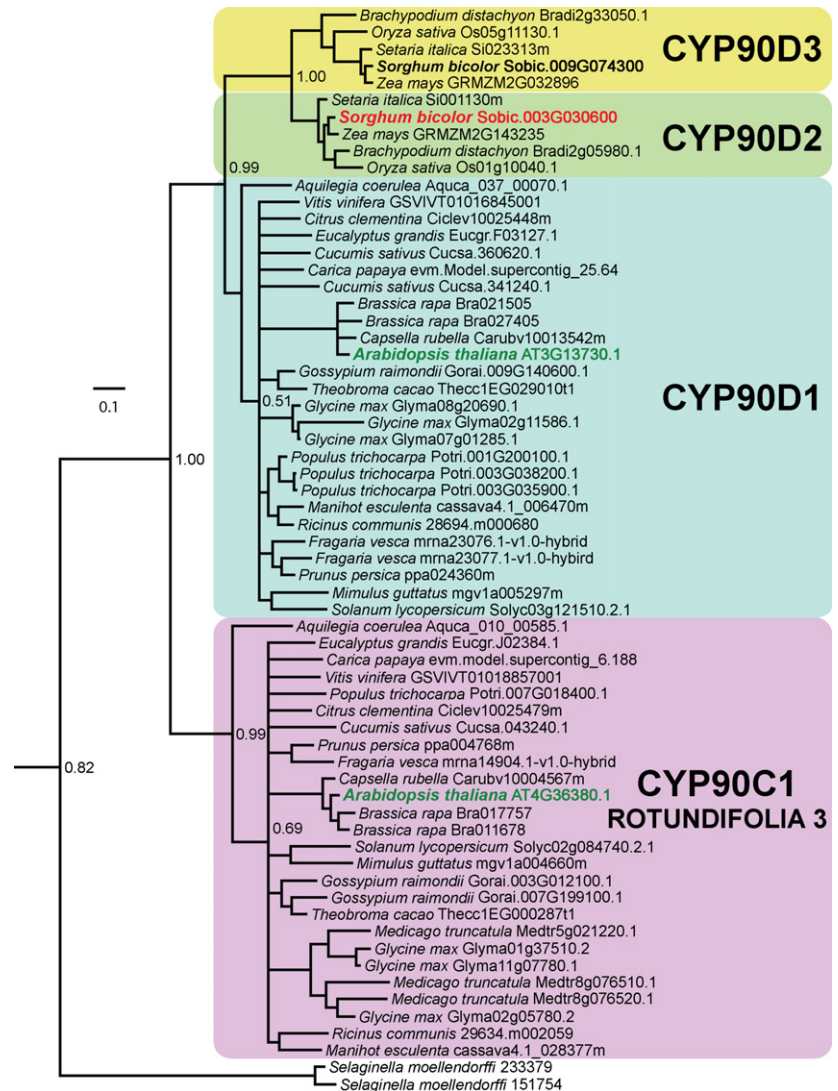
which vein density may be perturbed in C_4 plants whilst maintaining plant viability. This, combined with a lack of *bona fide* examples of reversion from C_4 to C_3 species (Ingram *et al.*, 2011), may reflect an implicit evolutionary difficulty in the reversion of C_4 to C_3 photosynthesis.

In rice, *CYP90D2* has been shown to mediate the C_3 oxidation of 6-deoxoteasterone/teasterone, an intermediate in the BR biosynthesis pathway (Sakamoto *et al.*, 2012). Notably, both γ -M8 and EMS-M30 sorghum mutants were dwarfed (the γ -irradiated mutant had a more severe phenotype, Figure 1a), resembling known BR-deficient mutants of Arabidopsis (Szekeres *et al.*, 1996), tomato (*Solanum lycopersicum*) (Bishop *et al.*, 1996) and rice (Hong *et al.*, 2003). Other BR-related mutant phenotypes such as reduced internode elongation and delayed flowering (Li *et al.*, 2010), were also observed in γ -M8 and EMS-M30. The LVD mutants identified in this study exhibit many features characteristic of BR-deficient mutants.

The most closely related gene to *CYP90D2* in Arabidopsis is *CYP90D1* (AT3G13730), which is expressed predominantly in the leaf vasculature (Kim *et al.*, 2005). The role of BR in shoot vascular differentiation (by inducing xylem cell differentiation and repressing phloem differentiation) has been demonstrated previously in Arabidopsis (Caño-Delgado *et al.*, 2004). Moreover, Arabidopsis mutants with BR deficiencies have a reduced number of vascular bundles (Ibañez *et al.*, 2009), while transgenic lines with enhanced BR signaling exhibited an increased number of vascular bundles in the flowering stem (Ibañez *et al.*, 2009). In addition to this, the BR-deficient mutant *cpd* exhibits a signifi-

Figure 5. Bootstrapped majority-rule, maximum-likelihood phylogenetic tree of the CYP90C and CYP90D sub-family.

The scale bar indicates the number of amino acid changes per orthologous site. The values shown at selected nodes indicate the proportion bootstrap support from 100 replicates. Arabidopsis genes are indicated in green. The mutated sorghum gene identified in this study is indicated in red. Clades are shaded according to prior nomenclature of the named Arabidopsis or rice genes (Kim *et al.* 1998, 2005; Hong *et al.* 2003).



cant reduction in the complexity of leaf veins (Zhiponova *et al.*, 2013). In rice, a *CYP90D2* null mutant has an altered number of bulliform cells, resulting in a rolled and twisted leaf morphology (Li *et al.*, 2013). The nature of anatomical changes observed in the sorghum mutants (altered vein density mainly due to changes in the size and number of M cells) indicates a conserved role for BR in vascular differentiation and patterning.

Although BR has been implicated in vascular differentiation processes, auxin signaling is more commonly associated with the specification of vein spacing patterns (Scarpella *et al.*, 2006). Intriguingly, however, studies in Arabidopsis revealed that an interaction between auxin transport and BR signaling may be necessary to establish periodic vascular patterns in the shoot (Ibañez *et al.*, 2009). This suggestion is consistent with the fact that loss-of-function *CYP90D2* mutants in sorghum show a domain shift in the leaf, such that blade tissue differentiates more

like sheath tissue (i.e. the leaves were thicker with lower vein density). In the C_4 plant maize, this same phenotype was observed in gain-of-function *KNOTTED1* mutants (Freeling and Hake, 1985), and in plants treated with auxin transport inhibitors (Tsiantis *et al.*, 1999). Furthermore, the *KNOTTED1* transcription factor has been shown to directly suppress BR signaling in rice (Tsuda *et al.*, 2014). We therefore hypothesize that co-ordination of BR and auxin signaling networks is required for accurate domain specification in the sorghum leaf.

EXPERIMENTAL PROCEDURES

Plant materials

Seeds of *Sorghum bicolor* accession BTx623 were gifts from William L. Rooney (Department of Soil and Crop Sciences, Texas A&M University, College Station, TX, USA) and Xin Zhanguo (United States Department of Agriculture, Agricultural Research

Service (ARS), Washington D.C., USA). The two stocks were named BTx623-Rooney (BTx623-R) and BTx623-Zhanguo (BTx623-Z), respectively.

Mutagenesis experiment

BTx623-Z seeds were treated with 400 Gy of γ rays, and BTx623-R seeds were treated with 0.28% EMS, and were grown to obtain M₁ plants. The panicles of M₁ plants were bagged for self-pollination and harvested individually to generate M₂ lines. Twenty-four seeds were grown for each M₂ line. Seedlings from the M₂ generation were screened for LVD as described below. A single mutant line exhibiting LVD from each mutagenesis pool (γ -M8 and EMS-M30) was selected for further characterization. These individual lines were crossed as described by Rizal *et al.* (2014) to obtain both self- and cross-pollinated seeds. γ -M8 and EMS-M30 mutants were also selfed to maintain homozygous mutant stocks, and were also crossed with WT to generate F₁ heterozygous lines. The F₁ lines were selfed to produce segregating F₂ populations.

Phenotypic analysis of leaf venation patterns

To ensure uniformity of the vein density measurements, all measurements were made on the mid-section of the 12th leaf once it was at the youngest fully expanded stage. Other veins, including transverse veins, were not assessed. Vein density was approximated as the number of minor veins that were orientated orthogonal to a 2 mm transect line in the mid-section of the 12th fully expanded leaf. For initial mutant screening, at least four 2 mm windows per leaf were counted by visualizing the upper leaf surface using a handheld Dino-Lite[®] (<http://www.dino-lite.com/>) USB microscope. Plants exhibiting a vein density of fewer than seven veins per mm were selected as candidate LVD mutants for further analysis. Transverse sections were cut by hand, and images were obtained using an Olympus DP71 camera attached to an Olympus BX51 microscope (<http://www.olympus-lifescience.com/>). For detailed phenotypic analysis of selected mutants, between 9 and 20 12th leaves for each genotype were assessed. Multiple features were recorded, including the number of M cells between each vein pair, and BS cells around each vein were counted. Linear inter-veinal distance, vein width, vein height, M cell width and vascular bundle area were also measured for each mutant line (Table S1). In all cases, unpaired two-sample Student's *t*-tests were used to assess the probability that mutant and WT lines were different. Fluorescence images of calcofluor-stained cellulose and chlorophyll autofluorescence were taken using an Olympus BX61 microscope with DAPI and RFP filters, respectively.

Measurements of photosynthetic rate

Gas exchange measurements were performed on the 12th fully expanded leaf inside the glasshouse between 9:00 and 13:00 h. Plants were watered before the measurements. Three replicates were measured for each mutant and WT plant using a Li-COR 6400XT infra-red gas analyzer (Li-COR Inc., <http://www.licor.com/>). CO₂ assimilation curves were obtained for a range of CO₂ concentrations between 10 and 1600 $\mu\text{mol CO}_2 \text{ mol}^{-1}$ air, at a constant light intensity of 2000 $\mu\text{mol photons m}^{-2} \text{ sec}^{-1}$. Light-dependent CO₂ assimilation was measured at decreasing light intensity at a constant CO₂ concentration of 400 $\mu\text{mol CO}_2 \text{ mol}^{-1}$ air.

Selection of individuals for bulk segregant analysis

The distribution of vein densities in individuals of the F₂ segregating population was fitted to a mixture model comprising two

normal distributions using a generalized additive model for location, scale, and shape (GAMLSS) (version 4.1-1) (Stasinopoulos and Rigby, 2007). Using this approach, all individuals were assigned a probability of having either LVD or WVD, and individuals from both populations were selected for bulk segregant analysis only if the probability of their belonging to either the LVD or WVD pool was ≥ 0.999 . For γ -ray segregants, two pools each containing 30 individuals were constructed from each of the LVD and WVD populations. Similarly, for EMS segregants, two pools each comprising 28 individuals were made from each of the LVD and WVD populations.

DNA extraction and sequencing

Total genomic DNA was extracted from young leaves using a cetyl trimethylammonium bromide protocol (Murray and Thompson, 1980). The DNA quantity and quality were evaluated using a Nanodrop ND-8000 spectrophotometer (Thermo Scientific, <http://www.thermoscientific.com>) and agarose gel electrophoresis. Equal quantities of DNA from each individual were pooled for each group (LVD and WVD), and sequenced using Illumina (<https://www.illumina.com>) HiSeq platforms at the Beijing Genomics Institute, Shenzhen, China. DNA from the WT parental lines BTx623-Z and BTx623-R was also extracted and sequenced.

Generation of a reference genome for sorghum WT lines BTx623-Z and BTx623-R

The parental lines used in this experiment were expected to contain a number of variants not found in the published genome reference for sorghum. To prevent these variants from influencing downstream analyses, the respective parental WT lines were also sequenced, and a reference genome sequence for each specific parental line was generated by correcting for the WT-specific alleles (Table S9). Reads were first filtered/trimmed for base quality using the FASTXtoolkit (version 0.0.13; http://hannonlab.cshl.edu/fastx_toolkit/index.html). The parameters used were base quality ≥ 20 , minimum length after 3' trimming ≥ 30 , and fraction of read length with high-quality bases ≥ 0.85 . Longer reads (approximately 100 bp) were aligned using BWA-MEM (Li, 2013), whereas shorter reads (< 50 bp) were aligned using Bowtie 2 (Langmead and Salzberg, 2012). Only paired-end reads that mapped consistently with the library construction protocol were used for further analysis. To correct differences between the published sorghum genome reference and the experimental lines, Pilon (version 1.5) (Walker *et al.*, 2014) was run iteratively using the 'fix all' option. Pilon was run iteratively, as changes at each iteration corrected ambiguous and erroneous bases and enabled subsequent increases in the number of correctly mapped reads and hence further corrections.

Variant discovery

Paired-end reads from γ -M8 and EMS-M30 genome sequences were first filtered/trimmed for base quality using Trimmomatic-0.32 (Bolger *et al.*, 2014; parameters: leading = 10, trailing = 10, sliding window = 5:15, minlen = 25), and then mapped using BWA-MEM (Li, 2013) on the respective WT genomes with a minimum mapping quality of 20 (Table S10). Subsequently, they were subjected to indel realignment and base quality recalibration using the 'Genome Analysis Toolkit' (GATK, version 2.7.4) (DePristo *et al.*, 2011). To recalibrate base quality scores the set of variants identified by both GATK and samtools (Li *et al.*, 2009), and having high quality (genotype quality ≥ 50 ; variant quality ≥ 500 and ≥ 200 for GATK and samtools, respectively; depth

between 0.5 and 2 times the mean depth) were chosen as the reference set. Variant calls were made using GATK's HaplotypeCaller. To reduce the number of false positives arising from WT alleles that were not corrected by Pilon, variants were also called for respective WTs. Stringent filters were applied to the calls to discard those that (i) also occurred in the WT sample, (ii) had an allele frequency < 0.3, (iii) had an unlikely coverage depth or extreme read depth (outside the range of 0.2–5 times the depth), and (iv) failed the internal programmatic (hard) filters recommended by GATK.

Phylogenetic inference

An iterative hidden Markov model search was performed starting with the query gene Sobc.003G030600 (Kelly *et al.*, 2011). The identified sequences were aligned using MergeAlign to reduce the impact of evolutionary model selection on multiple sequence alignment inference (Collingridge and Kelly, 2012). This alignment was subjected to bootstrapped maximum-likelihood phylogenetic inference using FastTree (Price *et al.*, 2009) utilizing the JTT model of amino acid substitution, CAT rates and 100 replicates.

PCR verification of the candidate mutations

PCR primers were designed using primer3 (version 0.4.0) (Untergasser *et al.*, 2012) with default parameters. PCR amplification was performed using appropriate primer pairs and genomic DNA extracted from mutants, BC₁F₂ progenies and corresponding WTs as appropriate. In all cases, PCR was performed using iTaq DNA polymerase (Intron Biotechnology, <http://eng.intronbio.com>). The PCR products were visualized using SYBR[®] Safe DNA gel stain (Invitrogen, <https://www.thermofisher.com>) after electrophoresis in 1.5% agarose gels. To validate the sequence of the PCR products, amplicons were excised from the gel and purified using a QIAquick gel extraction kit (Qiagen, <https://www.qiagen.com/>). The purified PCR products were sequenced by Macrogen (<http://www.macrogen.com/>). The sequences of the PCR products were then assembled using the sequence assembly tool CAP3 (Huang and Madan, 1999) with default program parameters.

Sequence data access

The raw read data for this project have been submitted to the National Center for Biotechnology Information Sequence Read Archive under accession number BioProject ID PRJNA279538. The Sequence Read Archive experiment numbers are SRX971817, SRX973468, SRX974489, SRX974516, SRX974517 and SRX974522.

ACKNOWLEDGEMENTS

We acknowledge the technical assistance of Michael Alcasid for crossing, Joan Turingan and Irma Canicosa for gas exchange measurements, Juvy Reyes, Menard dela Rosa and Gemma Lorenzana for sampling of leaves and DNA extraction, and Abraham Darius Llave and Diana Michelle Magbanua for bioinformatics analysis of γ -M8 lines. This work was funded by a Bill and Melinda Gates Foundation award to J.S., W.P.Q., J.A.L. and S.Ke.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Fluorescence micrograph image of transverse sections of mutant and non-mutant parental WT leaves.

Figure S2. Measurements of photosynthetic performance of mutant and WT plants.

Figure S3. Frequency distribution of vein densities.

Figure S4. Genome-wide plot of the alternate allele frequency of SNPs observed in the EMS-M30 LVD and WVD pools.

Figure S5. Genome-wide plot of the alternate allele frequency of SNPs and indels observed in the γ -M8 LVD pool.

Table S1. Leaf anatomy traits of the sorghum mutants and the corresponding wild-types.

Table S2. EMS-M30 LVD and WVD segregants selected for bulked segregant analysis.

Table S3. γ -M8 LVD and WVD segregants selected for bulked segregant analysis.

Table S4. Whole-genome sequencing of pooled LVD and WVD samples.

Table S5. Annotation of tightly linked SNP variants in the LVD pool of EMS-M30.

Table S6. Validation of SNPs in the EMS-M30 mutant by PCR amplicon sequencing.

Table S7. Variants linked to the unknown causal mutation in the LVD pool of γ -M8.

Table S8. Annotation of structural variants in the LVD pool of γ -M8.

Table S9. Number of variants in parental genomes versus the starting reference genome.

Table S10. Summary of sequencing data after filtering for base quality, and alignment onto respective WT reference genomes.

REFERENCES

- Abe, A., Kosugi, S., Yoshida, K.K. *et al.* (2012) Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat. Biotechnol.* **30**, 174–178.
- Bishop, G.J., Harrison, K. and Jones, J.D. (1996) The tomato *Dwarf* gene isolated by heterologous transposon tagging encodes the first member of a new cytochrome P450 family. *Plant Cell*, **8**, 959–969.
- Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
- Brown, W.V. (1975) Variations in anatomy, associations, and origins of Kranz tissue. *Am. J. Bot.* **62**, 395–402.
- Caño-Delgado, A., Yin, Y., Yu, C., Vafeados, D., Mora-García, S., Cheng, J.-C., Nam, K.H., Li, J. and Chory, J. (2004) BRL1 and BRL3 are novel brassinosteroid receptors that function in vascular differentiation in Arabidopsis. *Development*, **131**, 5341–5351.
- Chollet, R. and Ogren, W.L. (1975) Regulation of photorespiration in C₃ and C₄ species. *Bot. Rev.* **41**, 137–179.
- Collingridge, P.W. and Kelly, S. (2012) MergeAlign: improving multiple sequence alignment performance by dynamic reconstruction of consensus multiple sequence alignments. *BMC Bioinformatics*, **13**, 117.
- Dengler, N.G., Dengler, R.E., Donnelly, P.M. and Hattersley, P.W. (1994) Quantitative leaf anatomy of C₃ and C₄ grasses (Poaceae): bundlesheath and mesophyll surface area relationships. *Ann. Bot.* **73**, 241–255.
- DePristo, M.A., Banks, E., Poplin, R. *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498.
- Fouracre, J.P., Ando, S. and Langdale, J.A. (2014) Cracking the Kranz enigma with systems biology. *J. Exp. Bot.* **65**, 3327–3339.
- Freeling, M. and Hake, S. (1985) Developmental genetics of mutants that specify knotted leaves in maize. *Genetics*, **111**, 617–634.
- Haberlandt, G. (1896) *Physiologische Pflanzenanatomie*. Leipzig, Germany: Wilhelm Engelmann.
- Hatch, M. and Osmond, C. (1976) Compartmentation and transport in C₄ photosynthesis. In *Transport in Plants* (Huber, U. and Stockings, U., eds). Berlin/Heidelberg: Springer-Verlag, pp. 145–184.
- Hattersley, P.W. and Watson, L. (1975) Anatomical parameters for predicting photosynthetic pathways of grass leaves: the 'maximum lateral cell

- count' and the 'maximum cells distant count'. *Phytomorphology*, **25**, 325–333.
- Hong, Z., Ueguchi-Tanaka, M., Umemura, K. *et al.* (2003) A rice brassinosteroid-deficient mutant, *ebisu dwarf* (*d2*), is caused by a loss of function of a new member of cytochrome P450. *Plant Cell*, **15**, 2900–2910.
- Huang, X. and Madan, A. (1999) CAP3: a DNA sequence assembly program. *Genome Res.* **9**, 868–877.
- Ibañez, M., Fabregas, N., Chory, J. and Caño-Delgado, A.I. (2009) Brassinosteroid signaling and auxin transport are required to establish the periodic pattern of Arabidopsis shoot vascular bundles. *Proc. Natl Acad. Sci. USA*, **106**, 13630–13635.
- Ingram, A.L., Christin, P.A. and Osborne, C.P. (2011) Molecular phylogenies disprove a hypothesized C₄ reversion in *Eragrostis walteri* (Poaceae). *Ann. Bot.* **107**, 321–325.
- Kelly, S., Wickstead, B. and Gull, K. (2011) Archaeal phylogenomics provides evidence in support of a methanogenic origin of the Archaea and a thaumarchaeal origin for the eukaryotes. *Proc. Biol. Sci.* **278**, 1009–1018.
- Kim, G., Tsukaya, H. and Uchimiya, H. (1998) The *rotundifolia* gene of *Arabidopsis thaliana* encodes a new member of the cytochrome P450 family that is required for the regulated polar elongation of leaf cells. *Genes Dev.* **12**, 2381–2391.
- Kim, G.T., Fujioka, S., Kozuka, T., Tax, F.E., Takatsuto, S., Yoshida, S. and Tsukaya, H. (2005) CYP90C1 and CYP90D1 are involved in different steps in the brassinosteroid biosynthesis pathway in *Arabidopsis thaliana*. *Plant J.* **41**, 710–721.
- Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
- Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 1303.3997v2.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
- Li, J., Li, Y., Chen, S. and An, L. (2010) Involvement of brassinosteroid signals in the floral-induction network of Arabidopsis. *J. Exp. Bot.* **61**, 4221–4230.
- Li, H., Jiang, L., Youn, J.H. *et al.* (2013) A comprehensive genetic study reveals a crucial role of CYP90D2/D2 in regulating plant architecture in rice (*Oryza sativa*). *New Phytol.* **200**, 1076–1088.
- Ludlow, M.M. (1985) Photosynthesis and dry matter production in C₃ and C₄ pasture plants, with special emphasis on tropical C₃ legumes and C₄ grasses. *Aust. J. Plant Physiol.* **12**, 557–572.
- Monson, R.K. (1999) The origins of C₄ genes and evolutionary pattern in the C₄ metabolic phenotype. In *C₄ Plant Biology* (Sage, R.F. and Monson, R.K., eds). Millbrae, CA: California Academic Press, pp. 377–410.
- Murray, M. and Thompson, W. (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* **8**, 4321–4326.
- Pengelly, J.J.L., Kwasny, S., Bala, S., Evans, J.R., Voznesenskaya, E.V., Koteyeva, N.K., Edwards, G.E., Furbank, R.T. and von Caemmerer, S. (2011) Functional analysis of corn husk photosynthesis. *Plant Physiol.* **156**, 503–513.
- Price, M.N., Dehal, P.S. and Arkin, A.P. (2009) FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650.
- Rizal, G., Karki, S., Alcasid, M., Montecillo, F., Acebron, K., Larazo, N., Garcia, R., Slamet-Loedin, I.H. and Quick, W.P. (2014) Shortening the breeding cycle of sorghum, a model crop for research. *Crop Sci.* **54**, 520–529.
- Sage, R.F., Christin, P.-A. and Edwards, E.J. (2011) The C₄ plant lineages of planet Earth. *J. Exp. Bot.* **62**, 3155–3169.
- Sakamoto, T., Ohnishi, T., Fujioka, S., Watanabe, B. and Mizutani, M. (2012) Rice CYP90D2 and CYP90D3 catalyze C-23 hydroxylation of brassinosteroids *in vitro*. *Plant Physiol. Biochem.* **58**, 220–226.
- Scarpella, E., Marcos, D., Friml, J. and Berleth, T. (2006) Control of leaf vascular patterning by polar auxin transport. *Genes Dev.* **20**, 1015–1027.
- Slack, C.R., Hatch, M.D. and Goodchild, D.J. (1969) Distribution of enzymes in mesophyll and parenchyma-sheath chloroplasts of maize leaves in relation to the C₄-dicarboxylic acid pathway of photosynthesis. *Biochem. J.* **114**, 489–498.
- Slewinski, T.L., Anderson, A.A., Zhang, C. and Turgeon, R. (2012) Scarecrow plays a role in establishing Kranz anatomy in maize leaves. *Plant Cell Physiol.* **53**, 2030–2037.
- Slewinski, T.L., Anderson, A.A., Price, S., Withee, J.R., Gallagher, K. and Turgeon, R. (2014) Short-Root1 plays a role in the development of vascular tissue and Kranz anatomy in maize leaves. *Mol. Plant*, **7**, 1388–1392.
- Stasinopoulos, D.M. and Rigby, R.A. (2007) Generalised additive models for location scale and shape (GAMLSS) in R. *J. Stat. Soft.* **23**, 1–46.
- Szekeress, M., Németh, K., Koncz-Kálmán, Z. *et al.* (1996) Brassinosteroids rescue the deficiency of CYP90, a cytochrome P450, controlling cell elongation and de-etiolation in Arabidopsis. *Cell*, **85**, 171–182.
- Tsiantis, M., Brown, M.I.N., Skibinski, G. and Langdale, J.A. (1999) Disruption of auxin transport is associated with aberrant leaf development in maize. *Plant Physiol.* **121**, 1163–1168.
- Tsuda, K., Kurata, N., Ohyanagi, H. and Hake, S. (2014) Genome-wide study of KNOX regulatory network reveals brassinosteroid catabolic genes important for shoot meristem function in rice. *Plant Cell*, **26**, 3488–3500.
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M. and Rozen, S.G. (2012) Primer3 – new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115.
- Voznesenskaya, E.V., Franceschi, V.R., Kiirats, O., Freitag, H. and Edwards, G.E. (2001) Kranz anatomy is not essential for terrestrial C₄ plant photosynthesis. *Nature*, **414**, 543–546.
- Walker, B.J., Abeel, T., Shea, T. *et al.* (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE*, **9**, e112963.
- Wang, P., Kelly, S., Fouracre, J.P. and Langdale, J.A. (2013) Genome-wide transcript analysis of early maize leaf development reveals gene cohorts associated with the differentiation of C₄ Kranz anatomy. *Plant J.* **75**, 656–670.
- Zhiponova, M.K., Vanhoutte, I., Boudolf, V. *et al.* (2013) Brassinosteroid production and signaling differentially control cell division and expansion in the leaf. *New Phytol.* **197**, 490–502.