STOMAGEN-like Regulates Stomatal Generation Causing Increased Water Use Efficiency during C_4 Evolution

Yong-Yao Zhao[1,2], Mingju Amy Lyu, FenFen Miao, Genyun Chen, Xin-Guang Zhu[1,2]

1. Center of Excellence for Molecular Plant Sciences, Chinese Academy of Sciences
2. State Key Laboratory for Plant Molecular Genetics

Abstract

Compared with C_3 plants, C_4 plants maintain lower stomatal conductance (g_s), attributed to decreased maximal stomatal conductance (g_{smax}), without losing photosynthetic CO_2 uptake rate. Which stomatal traits caused the decreased g_{smax} and how did the g_{smax} change along C_4 evolution as well as the molecular mechanism underlying this change remain unknown.

Stomatal patternings were examined in *Flaveria* genus, which contains species at different evolutionary stages from C_3 to C_4 photosynthesis. We further used comparative transcriptomics, transgenic experiments and semi-*in-vitro* analysis to identify the gene underlying decreased g_{smax} in C_4 species.

We found that the evolution from C_3 to C_4 species was accompanied by a stepwise decrease in g_{smax} and stomatal density (*SD*) and the dramatic decrease occurred between type II and C_4-like species. *SD* controled the decrease of g_{smax} during C_4 evolution; *FSTOMAGEN* gene positively regulated stomatal development and its changed expression underlay the decreased *SD* during C_4 evolution; STOMAGEN-like genes expression also decreased along photosynthetic evolution in other C_4 lineages.

Our study suggests that the lower *SD* is another major feature of C_4 evolution, besides C_4 enzymes and Kranz anatomy. *FSTOMAGEN* can be used to engineer *SD*, a C_4 feature required in the current effort of C_4 crop engineering.

Key words:

Water use efficiency, Stomatal density, C_4 photosynthesis, STOMAGEN, Stomatal conductance

The total word count for the main body of the text is 6346. The word count of Introduction, Materials and Methods, Results, Discussion and Acknowledgement is 811, 2293, 2100, 1099 and 43 respectively. This paper includes 8 colour figures. The supporting material includes 18 figures and 5 tables.

Introduction

Compared with C_3 plants, C_4 plants have higher light use efficiency (LUE) as a result of a CO_2 concentrating mechanism (*1*). At the same time, C_4 plants have higher water use efficiency (WUE), as a result of both lower g_s and higher photosynthetic CO_2 uptake rate (*A*) due to the CO_2 concentrating mechanism (*2, 3*). C_4 photosynthesis therefore optimizes water and carbon relations by reducing transpiration without compromising carbon assimilation (*4*). Improvements in WUE enables C_4 plants to better adapt to saline lands, hot and drought subtropical environments and open habitats compared with their close C_3 relatives, which raises the possibility that C_4 photosynthesis might have been selected as a water-conserving mechanism (*4*) and that water limitation might be the primary driver for the initial ecological advantage of C_4 over C_3 (*5-7*). Given that C_4 plants have higher photosynthetic energy conversion efficiency and higher potential for biomass productivity, their lower g_s is critical for the survival and fitness of C_4 plants in the field, especially in a world where water shortage is predicted to occur more often.

C_3 plants can gain improved WUE through reducing transpiration under long-term changed environmental conditions by decreasing g_{smax} (*8, 9*), which implies that the decreasing g_{smax} may be an option that plants use to reduce water loss. Indeed, phylogenetically informed comparison shows that C_4 plants have a lower g_{smax} compared with their close C_3 relatives (*10*), indicating that the lower g_{smax} contribute to the higher water-conserving ability of C_4 land plants (*3, 10*). Therefore, low g_{smax}, same as the C_4 metabolite shuttle and Kranz anatomy, may be a required feature for C_4 land plants. However, so far, in comparison with the extensive studies on the C_4 metabolite shuttle and Kranz anatomy, relatively little effort has been invested into studying the changes in g_{smax} and the mechanism underlying these changes during C_4

evolution (*11*).

In principle, g_{smax} positively correlates to operating g_s (*8, 12*) and the g_{smax} is determined by stomatal density (*SD*) and stomatal size (*SS*). In Eucalyptus and Arabidopsis, *SD* positively controls g_{smax} and operating g_s (*8, 13-15*). Higher *SD* appears to have been selected as a strategy to gain higher g_{smax} to cope with low atmospheric CO_2 levels over the last 400 million years (*16*). In contrast, when grown under drought, potato plants severely reduce *SS* to lower operating g_s and increase WUE (*17*). Under ABA treatment, a sufficiently small *SS* can lower g_{smax} to achieve higher WUE in *Tradescantia virginiana (18)*. Therefore, in different species or under different environmental conditions, either *SD* or *SS* is used as a primary determinant for g_{smax} in plants. In the case of C_4 photosynthesis, it also remains unknown whether SD or SS is the dominant characteristic responsible for the lower g_{smax}. For example, lower g_{smax} of different species in one C_4 clade were attributed to either smaller *SS* or lower *SD* (*10*), which may be a result of a compounded phylogenetic effect, i.e. species used in the study were evolutionarily distant from each other and hence genetic factors unrelated to photosynthetic type may influenced stomatal patterning(*3*).

Stomatal development is controlled by a well-studied genetic program (*19, 20*), which mainly involves a series of ligands and the transmembrane receptors, leading to downstream changes of mitogen-activated protein kinase (MAPK) and transcription factors (*19, 20*). TOO MANY MOUTHS (TMM) and ERECTA-family (ERfs: ER, ERL1, ERL2) are receptor-like kinases that negatively regulate stomatal development (*21, 22*), and the TMM interacts with ERfs (*21, 23*). The ligands that bind to ERfs and TMM receptors are mainly EPIDERMAL PATTERNING FACTOR/EPIDERMAL PATTERNING FACTOR-LIKE family (EPF/EPFL-family) peptides (*24*). The *EPF*/*EPFL*-family genes encode cysteine-rich peptides, which can fold into a three-dimensional structure through disulfide bond (*25*). Overexpression of *EPF1* decreases *SD*, whose loss of function mutation increases *SD* and results in stomatal cluster (*26*). *EPF2*, which acts earlier in stomatal development relative to *EPF1*, has a similar genetic effect as *EPF1* (*27*). STOMAGEN can compete with EPF1 and EPF2

for ERfs (*28*), and then prevents the phosphorylation of downstream receptors, which results in the changed expression of basic Helix-loop-Helix (bHLH) paralogs, SPCH, MUTE and FAMA, ultimately positively regulates stomatal development (*29-32*).

To study whether *SD* or *SS* underlies the decreased g_{smax} in plants with C_4 photosynthesis, here we use the *Flaveria* genus, a representative of dicotyledons which contains closely related species at different evolutionary stages from C_3 to C_4 photosynthesis. Using a genus containing species with intermediate photosynthetic types enables examination of changes in g_{smax} and stomatal patterning during C_4 evolution. This study aims to answer the following questions: 1) which stomatal patterning trait, i.e. *SD* and *SS*, underlies the decreased g_{smax} in C_4 evolution and how did the g_{smax} and stomatal patterning traits change during C_4 evolution? 2) What's the molecular mechanism underlying the change of the stomatal patterning trait? 3) Is this molecular mechanism conserved across different C_4 evolutionary events?

Result

Maximum stomatal conductance and stomatal pattern during the evolution of C_4 photosynthesis

As a model of C_4 evolution, there are distinct evolutionary phases in *Flaveria* genus (Fig 1a). The maximum value of operational g_s (marked with circles on figure) for C_3 *Flaveria* species under saturated light intensity was higher than that for C_4 species, indicating that the C_3 *Flaveria* species tends to have a higher g_s (Supplementary Fig 1a). The pore length of stomata was approximately equal to 0.41×stomatal length, and the guard cell width was approximately equal to 0.33×stomatal length (Fig. 1b). Through measuring stomatal density and stomatal length, we calculated the g_{smax} for *Flaveria* species. The results show that, whether growing under outdoor natural light or constant light in the greenhouse, the g_{smax} decreased gradually from C_3 through intermediate species to C_4 species (Fig 1d, e). A significant correlation was found between the g_{smax} and g_s under saturated light for the *Flaveria* species with different photosynthetic types (Fig 1f). Under saturated light, the g_s of *son* (Type I) was extremely high (marked with circle on figure), matching

with its highest g_{smax} within the *Flaveria* genus (Supplementary Fig 1a, Fig1d). Similar to that under the saturated light intensity, the WUE (A/g_s) of *rob* (C_3) measured under unsaturated light intensity was higher than that of *bid* (C_4), and the g_s of *rob* was 71% higher than that of *bid* (Supplementary Fig. 1c). The observed reductions in both g_s and g_{smax} of C_4 and C_4-like *Flaveria* species compared to C_3 and Type I species suggest that the decrease in g_{smax} contributes to the decrease in g_s during C_4 evolution.

In comparison to C_4 and C_4-like photosynthesis, the stomatal densities (*SD*) of C_3 and Type I photosynthesis on the abaxial leaf surface of *Flaveria* were nearly 150% higher (Fig. 2a). With the progression from C_3 to C_4 species, there was a distinctly gradual decrease in *SD* (Fig. 2a, b). The difference in *SD* between C_3 and Type I species wasn't as large as that between C_3 and C_4 species; and the *SD* of Type II species fell in between Type I and C_4-like species (Fig. 2b). The *SD* of C_4 and C_4-like species were similar, although the *SD* of C_4-like species was slightly less than that of C_4 (Fig. 2b). For the adaxial leaf surface, the *SD* of C_3 species was again higher than that of the C_4 species, although the difference between *SD* in the intermediate species was less significant (Supplementary fig. 4a, b). There was a strong positive correlation of g_{smax} and *SD* either on abaxial leaves (R^2=0.8955, p<0.0001) (Fig. 2C) or on adaxial leaves (R^2=0.9165, p<0.0001) (Supplementary fig. 4C) and a significant negative correlation of g_{smax} and *l* on abaxial leaves (R^2=0.2596, p<0.005) (Fig. 2d), indicating the *SD* determined the change of g_{smax}, not stomatal size (SS). *SD* and *l* showed a significant negative correlation on abaxial leaves (R^2=0.5093, p<0.0001) (Fig. 2e) or on adaxial leaves (R^2=0.2055, p<0.05) (Supplementary fig. 4e), the same as reported previously (*8, 13, 16*).

Flaveria species, i.e. *rob*, *ram* and *bid,* were chosen to represent C_3, intermediate and C_4 species, respectively, to analyse the ratio between stomatal number and the number of epidermal cells, i.e. the stomatal index (*SI*). The *SI* of *rob* was nearly 50% higher than *bid* and *ram* in plants grown in either greenhouse or outdoors, suggesting that the proportion of stomata occupying the epidermis, contributed to the highest *SD* in *rob* (Fig. 2g). In addition to the analysis of *Flaveria* plants grown from cuttings, we

also planted *Flaveria* germinated from seeds. Results show that the *SD* of C_3 species was higher than C_4 species, consistent with the results from cuttings (Supplementary Fig. 5).

Comparison of the sequence and transcription abundance of genes related to stomatal pattern between C_3 and C_4 *Flaveria* species

The signaling pathway for stomatal development has been studied extensively (*19*). We identified amino acid sequence encoded by homologs of stomatal development related genes in *Flaveria*. Amino acid sequence alignment analysis showed that epidermal patterning factor/epidermal patterning factor-Like gene family (*EPF*/*EPFL*-family) homologs was conserved at the C-terminal part of amino acid sequence (Fig. 4A, Supplementary fig. 6). Members of *EPF*/*EPFL*-family genes encode secretory proteins containing signal peptides at the N-terminus and functional domains at the C-terminus (*26, 27, 33*). Analysis shows that the hydrophobicity of EPFs/EPFL homologs in *Flaveria* was similar to EPF/EPFL in Arabidopsis (Supplementary fig 7), implying the homologs are also the secretory proteins and should have the same physiological function as EPF/EPFL. We hypothesized that either the amino acid sequence or the expression of stomatal development related genes changed during C_4 evolution. Analysis results reveals no differences within the functional region of amino acid sequence for STOMAGEN, EPF1 and EPF2 homologs between C_3 and C_4 species (Fig. 3f, Supplementary Fig. 6a, b), indicating the change of amino acid sequence is unlikely to be a cause of the changed *SD* in C_4 evolution.

As up-regulation of transcription has been a major approach to increase the expression of many genes participating in the building of C_4 phenotype during C_3 to C_4 evolution in *Flaveria* **(*34, 35*)**, we compared the transcript abundance of stomatal development related genes between C_3 and C_4 *Flaveria* species. The expression patterns of known C_4 related genes in our data analysis were same as the previous works, which indicates that our data collection and analysis procedure were reliable (Supplementary Fig. 12) (*34, 35*). We found that the homologs of most stomatal

development related genes showed higher transcript abundance in developing tissues as compared to developed tissues (Fig. 3b), which is consistent with their function of regulating stomatal differentiation in developing leaf (*26, 30, 32, 36*). *STOMAGEN*, *EPF1*, *SDD1* and *TMM* showed obviously lower expression levels in C_4 species than in C_3 species (Fig. 3c). Among these genes, *STOMAGEN* was the one that changed the most (Fig. 3c) and showed the most significant difference (Fig. 3d). Considering that the *EPF1*, *SDD1* and *TMM* are negative regulators for stomatal development (Supplementary table 4) (*24, 26-28*), the decrease in their expression should increase *SD*. But actually SD in C_4 species was lower than that in C_3, indicating these genes are not be the factors responsible for the decrease in *SD* of C_4 species. *STOMAGEN* is a positive regulator of *SD* (Supplementary table 4) (*24, 30, 31*), whose lower expression in C_4 species was consistent with the decreased *SD* in C_4 species (Fig. 2a, b, Supplementary Fig. 4a, b). RT-qPCR quantification of the expression levels of *STOMAGEN*, *EPF1*, *EPF2* homologs between C_3 and C_4 species grown under our conditions also confirmed the results from RNA-SEQ (Fig. 3e, Supplementary Fig. 15). In this study, we renamed *STOMAGEN* in *Flaveria* as *FSTOMAGEN*.

There are considerable differences in amino acid sequence of signal peptide of *STOMAGEN* homologs, especially between species with large evolutionary distance (Fig. 4a). The functional domain of *STOMAGEN* is however conserved, not only between different *Flaveria* species (Fig. 3f), but also between species that have a large evolutionary distance (Fig. 4a). In Arabidopsis, increased expression of *STOMAGEN* increases *SD*; experiments with exogenously applied STOMAGEN peptides shows that the signal peptide is not required for STOMAGEN to induce increased SD (*30, 31, 33, 36*). A phylogenic tree constructed for *STOMAGEN* from different plants can be divided into two clades, i.e. dicotyledon (blue clades) and monocotyledon (red clades) (Fig. 4b). Previous studies showed that overexpressing rice *STOMAGEN* homologs increases *SD* in Arabidopsis and decreased expression of rice STOMAGEN homologs reduces *SD* in rice (*37-39*), implying conserved function of *STOMAGEN* across monocots and dicots. When plants are grown under conditions that promote or inhibit stomatal development, the expression of *STOMAGEN* is known to increase or

decrease, respectively (*14, 36, 40*). A similar modulation of *STOMAGEN* expression levels was observed for *rob* when grown at different light conditions (Supplementary Fig 10).

To confirm that *FSTOMAGEN* can induce stomatal generation, we overexpressed the *FSTOMAGEN* in *Arabidopsis*. Specifically, we recombined the signal peptide of *STOMAGEN* from *Arabidopsis* to the functional region of *FSTOMAGEN* and named the hybrid gene as AFSTO (Fig, 4c). Overexpressing either the AFSTO or the intact *FSTOMAGEN* (FSTO) from *F.rob* into *Arabidopsis* resulted in similar increased *SD* (Fig. 4e, f, g, j, Supplementary fig 16). Identification of transcription showed successful transgenes and the expression level of the AFSTO or FSTO increased with the increase in SD (Fig. 4e, f, h, i). Furthermore, overexpression of *FSTOMAGEN* also induced clustered stomata (Fig. 4d), as also reported earlier for the over-expression of *STOMAGEN* (*30*). To further confirm that *FSTOMAGEN* can increase *SD* in *Flaveria*, we synthesized a peptide representing the FSTOMAGEN functional region, consisting of 45 amino acids. In order to confirm our synthesized peptides that have been folded were biologically active, the FSTOMAGEN peptides were applicated in vitro on the leaves of Arabidopsis and resulted in increased *SD* and cluster stomata (Fig 5b, c). Application of the peptides on the leaves of *bid* also increased *SD* and *SI* and caused stomatal cluster (Fig 5a, d, e). These results demonstrate that, in addition to its *in vivo* function, *in vitro* application of FSTOMAGEN also directly positively regulate *SD* in *Flaveria*. This is consistent with the notion that *STOMAGEN* is expressed in mesophyll cells and secreted out of cells to influence the stomatal development in the epidermal layer in *Arabidopsis* (*30, 31*).

The transcript abundance of *FSTOMAGEN* during C_4 evolution in *Flaveria*

We examined the copy number of *FSTOMAGEN* in *Flaveria* using blastn with the coding sequence of *FSTOMAGEN* as query sequence against the assembled contigs of each *Flaveria* species from 1KP. Results show that there was only one hit sequence with e-value less than 10^{-5} in each species, suggesting that there is a single copy of *FSTOMAGEN* in *Flaveria* (Supplementary Fig. 13). Therefore we can ignore

the impact of other *FSTOMAGEN* paralogs on stomatal development, and there will be no quantitative errors in the gene expression levels caused by other paralogs that were not considered. Examination of the RNA-seq data of *Flaveria* showed that the expression levels of *FSTOMAGEN* in C_3 species were markedly higher than those in C_4 species along different developmental stages of leaves (Fig. 6a). During leaf growth and development, the expression of *FSTOMAGEN* in both *bid* and *rob* increased first, reaching the highest level at leaf development stage 3 (leaf length 1.5 cm) or 4 (leaf length 2.0 cm), and then declined (Fig. 6a). The similar patterns of *FSTOMAGEN* expression in these two *Flaveria* species indicates that it would be valid to compare the *FSTOMAGEN* expression levels in leaves at the same development stage between *Flaveria* species. The expression level of *FSTOMAGEN* in *bid* (C_4) was lower than that in *rob* (C_3), and that of *ram* (intermediate) lied between them (Fig. 6b). This encouraged us to test more *Flaveria* species with different photosynthetic types. Similar results were obtained with either RNA-seq or RT-qPCR (Fig. 6c, Supplementary Fig. 14), where we found the gradual decrease of *FSTOMAGEN* expression from C_3 to intermediate species to C_4 species. Together these results suggest that the gradual change in the *FSTOMAGEN* expression levels between different species underlay the stepwise change in *SD* along C_4 evolution.

The expression of *FSTOMAGEN* homologs for species with different photosynthetic type in multiple C_4 lineages.

To gain insight into whether *STOMAGEN* homolog might be a generic factor leading to the difference in *SD* between C_3 and C_4 species in higher plants, we tested its expression level in different C_4 lineages other than the *Flaveria genus*. The *SD* of abaxial leaf in C_3 rice was 240% higher than that in C_4 maize, consistent with a 574% higher g_s in rice compared to maize (Fig. 7a, b). The stomatal length (l) in maize was 133% longer (Fig. 7a, b). We analysed the transcript abundance of *STOMAGEN* homologs, since there are two and three paralogs of *STOMAGEN* in rice and maize respectively (*37*), only one in rice and two in maize are expressed. The *STOMAGEN* homologs in rice had much higher transcript abundance than those in maize (Fig. 7d),

while the transcript abundance of other stomatal development related genes did not differ to the same extent between rice and maize (Fig. 4c). This result encouraged us to test more C_4 lineages (Fig. 7e). C_4 photosynthesis related genes were indeed higher in C_4 species, suggesting the data and analysis pipeline can be used to compare the differences in expression of genes between C_3 and C_4 photosynthetic types (Fig 7f). It was showed that the expression levels of *STOMAGEN* homologs for C_4 were lower than those for C_3, while intermediate had expression between them (Fig 7g).

Method

Plant materials, sample condition and data sources

All *Flaveria* species used in this study were kindly provided by Prof. Rowan F. Sage (University of Toronto). Plants used for stomatal phenotyping were re-grown from cuttings. Placed in an environment humidified with humidifiers to generate roots. After plants generated roots, they were transferred outdoors (Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai, China), and transplanted into 4-L pots filled with topsoil in April. In July, the number of stomata on abaxial and adaxial surfaces of leaves were counted. Another batch of *Flaveria*, which were also grown from cuttings, were transferred to the phytotron (Center of Excellence for Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai, China). Growth temperature of the phytotron was maintained at 25-28□, photoperiod was maintained at 16h light / 8h dark, relative humidity (RH) was at 60-65% and photosynthetic photon flux density (PPFD) was maintained at about 180 μmol m^{-2} s^{-1}. Two months later, the numbers of stomata on the abaxial and adaxial surfaces of leaves were counted. RNA was extracted from 1.5~3.0 cm long young leaves on the axial shoot which were still expanding. Three hours before sampling, we provided sufficient water to ensure that plants were not water limited, since drought and water status can affect expression levels of *STOMAGEN* and *SD* (*14*, *41*). Gene expression was also examined under normal growth conditions by RT-qPCR for the *rob*, *ram* and *bid*, which are representative of C_3, intermediate and C_4 species in the *Flaveria* genus. These three species were grown in the same phytotron with the same environmental

conditions, except that we did not water plants before sampling. To test whether plants grown from cuttings or from seeds show difference in the *SD*, we further grew *Flaveria* species from seeds in phytotron and examined stomatal properties. Same environmental conditions were maintained for phytotron as above. Two months after the germination of seeds, we counted stomatal number.

All plants were well watered to avoid drought stress during their growth. We supplied the same amount of commercial nutrient solution (Peters Professional, SCOTTS, N:P:K = 20:20:20+TE, at recommended dose for application) to the different species to avoid nutrient deficiency.

Rice (*Oryza Sativa Nipponbare*) and maize (*Zea maize* B73) were grown in the greenhouse (Center of Excellence for Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai, China) under a PPFD of 550 umol/m^2/sec, a photoperiod of 12h light / 12h dark, and a daytime temperature of around 31℃ and a night-time temperature of around 22℃. The Relative humidity was maintained at around 50-55%. The plants were regularly watered to avoid drought stress.

In addition to the expression data collected by in this study, we also used public data for comparison of transcript abundance. Considering that stomatal development related genes are mainly expressed in young tissues, in each of these used public data source, samples for young leaves were included. For the *Flaveria*, the samples data from four source containing the leaves of plants grown under normal condition were analysed. The first data source is the One Thousand Plants (1KP) Consortium, and we subsequently analyzed the data according to the methods described in previous work (*42*, *43*); The second and third data sources were Mallman et al (2014) (*35*) and Udo Gowik et al (*34*); The fourth data source was Kumari Billakurthi et al (2018) (*44*). The last three data sources are all from Heinrich-Heine-University. For the second and third data source, the second and fourth visible leaves were sampled; the second leaf was not completely mature, and stomatal developmental related genes were still actively expressed. In addition to *Flaveria*, transcripts of C_3 and C_4 species from other C_4 lineage sequenced from 1KP were assembled using Trinity (*45*) with default

parameters. Transcript abundances were analyzed by mapping the reads to assembled contigs for corresponding samples using the RSEM package with mapping option of bowtie 1 (*46*), and outliers that less than Q1 (lower quartile) - 1.5IQR (interquartile range) and greater than Q3 (upper quartile) + 1.5IQR (interquartile range) were removed. Expression levels for stomatal development related genes in developing leaves for rice and maize were obtained from Wang et al. (*47*).

Conservation and hydrophobicity analysis of proteins

Amino acid sequence of the homologs of *STOMAGEN*, *EPF1* and *EPF2* were aligned by clustalx, and graphs for the multiple alignments were drawn with ESPrit 3.0 (*48*) (http://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi). Hydrophobicity of proteins was determined by ProtScale (*49*) (https://web.expasy.org/protscale/).

Construction of the phylogenetic tree of *STOMAGEN*

In order to construct the gene evolutionary tree of *STOMAGEN*, we searched the orthologs of *STOMAGEN* in 28 representative species among Viridisplantae as demonstrated in (*50*),The protein sequences of representative species were downloaded from Phytozome (https://phytozome.jgi.doe.gov/). Orthologs of STOMAGEN were predicted applying OrthoFinder (V 0.7.1) (*51*)using default parameters. Proteins sequences of STOMAGEN orthologs were then aligned using MUSCLE (*52*). We than inferred the gene phylogeny using both maximum likelihood (ML) and Bayesian-inference (BI) methods. The ML tree was constructed applying RAxML (V2.2.3) (*53*), where bootstrap scores were obtained by running 100,000 independent trees. BI tree was built applying MrBayes package (v 3.2.1) (*54*), where posterior probabilities were calculated by running 2000,000 generations with the first 2000 generations being used as burn-in. In both cases, PROTGAMMAILG model (General Time Reversible amino acid substitution model with assumption that variations in sites follow gamma distribution) was used to infer the gene phylogeny. The protein sequence of AT4G14723 (*CLL2*) from A. thaliana was used as outgroup as it showed the highest identity with STOMAGEN in A. thaliana, with a protein

identity of 26.97% and E-value of 0.037 based on Blastp (*55*). Protein sequences for the *Flaveria* species were predicted using OrfPredictor (*56*) based on de novo assembled transcripts as described in (*43*).

Measurement of stomatal conductance (g_s)

The LI-6400 Portable Photosynthesis System (Li-Cor, Inc., Lincoln, NE, USA) was used to measure g_s on the youngest fully expanded leaves from approximately three-months-old *Flaveria* plants. The cuvette temperature was maintained at 25□, the PPFD was at ~500 μmol/m²s and the CO_2 concentration was at around 400 ppm. Chamber humidity was matched to ambient air humidity. Leaf-to-air vapor pressure difference (VPD) was maintained around 1.1 kPa. Before the g_s measurements, *Flaveria* plants were maintained under constant light with a PPFD of ~500 μmol/m^{2}s until g_s reached a steady state. External environments during the g_s measurements for rice and maize were controlled to be their growth conditions. In maize and rice, the16-day-old third leaves which were the youngest mature leaves, were used for gas exchange measurements (*47*).

Phenotyping stomatal traits and the calculation of g_{smax}

The youngest fully expanded leaves (usually the third or fourth leaves from the apex in *Flaveria*) at the main stem were used to phenotype stomata related traits. In *Flaveria* and maize, stomata related traits were observed using both nail polish method and epidermis directly peeled from leaves (*6, 13*). For the first method, nail polish was painted on both leaf surfaces to produce the impressions of epidermis. After drying of the nail polish, the nail polish was peeled off the leaf and placed on a microscope slide for examination. For the second method, leaf epidermis was peeled off manually and transferred onto microscope slide. For the count of SD in Arabidopsis, the rosette leaves was soaked in 9:1 (V/V) ethanol:acetic acid overnight, washed with water and cleared in chloral hydrate solution (glycerol:chloral hydrate:water, 1:8:1) overnight, and then stomata of Arabidopsis were stained with Safranin-O, finally leaves were placed on a microscope slide for imaging. Light

microscope (smarte; CNOPTEC), camera and image processed software (OPTPRO; CNOPTEC) were used to observe and image epidermis at 100x magnification in *Flaveria*. Center of a leaf was chosen to count stomatal density and measure stomatal length. Five stomata randomly chosen along a diagonal of the varnish peels or epidermis were measured for stomatal length.

G_{smax} was calculated using the follow equation (*16*):

$$g_{smax} = \frac{d}{v} \cdot \mathrm{D} \cdot a_{max} / (e + \frac{\pi\sqrt{a_{max}/\pi}}{2})$$

The constants d is the diffusivity of H_2O in air; the constant v is the molar volume of air; D is the stomatal density; a_{max} is the maximum stomatal pore area which is calculated as $a_{max} = \pi p^2/4$ in *Flaveria* species, and the p is the stomatal pore length; e is the stomatal pore depth, which is known as equal to guard cell width.

RNA isolation, semiquantitative RT-PCR and quantitative real-time PCR

The young leaf samples taken from the main stem in *Flaveria* or the developing rosette leaves in Arabidopsis were immediately frozen in liquid nitrogen and used directly for RNA extraction. Total RNA was extract from the young leaves using the PureLink RNA Mini kit (Thermo Fisher Scientifc). The first-strand cDNA was synthesized by the TransScript One-step gDNA Removal and cDNA Synthesis SuperMix (TRANSGEN). Semiquantitative RT-PCR was performed with Phanta® Max Super-Fidelity DNA Polymerase (vazyme). The actin8 was chosen as the loading control. Real-time PCR was performed on the CFX connect™ (Bio-Rad) with the UNICON™ qPCR SYBR ®Green Master Mix (YEASEN) according to the manufacturer's instructions. The calculation of expression according to the $2^{-\Delta\Delta CT}$.

The expression stability of internal reference genes for the quantitative real-time PCR was assessed by the software: geNORM and NormFinder (*57, 58*). Specifically, we first calculated the coefficient of variations (CV) of all genes in different *Flaveria* species. After that, all genes were sorted according to the calculated CV values. 150 genes with the lowest CV were selected as potential candidate reference genes. Among these, those with mean expression level >300 were chosen as candidate reference genes. We also included those commonly used house-keeping genes, which

include *UBC9* (AT4G27960), *UBQ11* (AT4G05050), *ACT7* (AT5G09810), *GAPDH* (AT1G13440), *EF1a* (AT5G60390), *CACMS* (AT5G46630) (Supplementary fig. 6a, b, d), as the internal reference genes in this study. The specificity of the primers (Supplementary table 2) for all these candidate reference genes were identified by agarose electrophoresis (Supplementary fig. 6c). With Genorm (Supplementary fig. 6e) and Normfinder (Supplementary fig. 6f), we concluded that *ACT7*, *EF1a* and *HEME2* showed stable expression levels across *rob*, *ram* and *bid*. *EF1a* and *ACT7* are classic reference genes, and they showed minor variations under different conditions. The geometrical means of the expression level of *ACT7* and *EF1a* were finally used as normalization factors. All primers used for RT-qPCR are listed in Supplementary table 2. For each gene, three technical replicates and at least three biological replicates were performed.

Vector construction and transformation

A list of primer sequences were used for plasmid construction (Supplementary table 2). The *STOMAGEN* and *FSTOMAGEN* DNA sequences, named ATSTO-cDNA and FloSTO-cDNA respectively, were PCR-amplified from cDNA in Arabidopsis and *rob*. To obtain the AFSTO fragment, the signal peptide region of *STOMAGEN* was cloned from ATSTO-cDNA by PCR-amplification, and the functional region of *FSTOMAGEN* was PCR-amplified from FloSTO-cDNA. The signal peptide and functional region were combined by fusion PCR. To obtain an intact *FSTOMAGEN* (FSTO) gene, the whole gene in *rob*, including the signal peptide, propeptide and the *STOMAGEN* functional region, was amplified from the cDNA of *rob*. To over-express AFSTO and FSTO, the AFSTO and FSTO were cloned separately to a pcambia-3300 binary vector which has a CaMV 35S promoter. This construct was introduced into *A. tumefaciens* strain GV3101, transformed into Arabidopsis using a floral dipping procedure. Transgenic lines were selected on soil that had been treated with a diluted solution of phosphinothricin (diluted 3000 times) over night. About a month after the seeds were sowed on the soil, the stomata were observed on the biggest mature rosette leaf. The phenotype of Arabidopsis T1 and T2 generations was observed at the 400x

magnification using the light microscope. For each epidermal peel, five fields were counted for density.

Refolding of synthetic *FSTOMAGEN*

Chemically synthesized peptide (Synthesized peptide sequence: IGSVKPTCTY NECRGCRSRCRAEQVPVEGNDPINSAYHYRCICHR, Sangon Biotech, Shanghai, China) was dissolved in 20 mM Tris-HCl, pH 8.8, and 50 mM NaCl and then dialysed (Sangon Biotech,0.5 KD-1 KD) for 1 day at 4□ against 0.5 mM glutathione (reduced and oxidized forms, Macklin) and 200 mML-arginine (meilunbio) at pH 8.0, which were further dialysed three times against 20 mM Tris-HCl, pH 8.8, and 50 mM NaCl for 1.5 days to remove glutathione.

Stomata induction assay

When the first true leaf or cotyledons appeared in *bid* and Arabidopsis that had germinated in 1/2 MS sterilized solid medium, F*STOMAGEN* peptide was applied on the plants. After the *bid* and Arabidopsis was further incubated in 1/2 MS liquid medium at 22℃ for 5 days, stomatal number on the abaxial surfaces for the first true leaf and cotyledons was counted under a differential interference contrast microscope (DIC)**.**

Plotting and statistical analysis

The correlation, linear regressions and t-test (a=0.05) was performed by GraphPad Prism 8. Differences between means of multi-groups were assessed with one-way analysis of variance (ANOVA). Duncan's multiple range test was performed using the agricolae package of the R (https://www.r-project.org/) to detect differences between two groups. The A and g_s at the saturated light intensity were digitized from figure 1 and figure 3 in (*2, 59*) using the OriginPro 2018.

Accession numbers

The *Flaveria* read data have been submitted to the National Center for Biotechnology Information Short Read Archive under accession numbers: SRP036880, SRP036881, SRP036883, SRP036884, SRP036885, SRP037526, SRP037527, SRP037528, SRP037529, SRP006166, SRP169568. The *Flaveria* RNA-Seq datasets from 1KP Consortium were available at National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) under accession number: GSE54339. The rice and maize RNA-seq datasets under accession number: GSE54274.

Discussion

C_4 plants have higher water use efficiency and light use efficiency, as result of the combined effect of decreased operating g_s and a CO_2 concentrating mechanism (Fig 1b, c) (*3*, *4*). Previous work suggests that the g_{smax} in C_4 plants are lower than their C_3 relatives (*10*). Comparative studies using plants from different C_4 clades show that lower g_{smax} is due to either a smaller stomatal size (*SS*) for some C_4 clades, or lower stomatal density (*SD*) for other clades (*10*). We collected and analysed *SD* data from literature for a large number of C_3 and C_4 species (*10, 60-66*). The average *SD* on both sides of leaf surface of C_3 species was significantly higher than that of C_4 species (Supplementary Fig. 1). Recent studies have shown that when the physiological and ecological differences between C_3 and C_4 plants are compared, the phylogenic effect must be considered (*3, 67, 68*), *i.e.* comparing either stomatal density or size without considering the phylogeny inevitably has the compounded effect of other genetic factors other than photosynthetic types. Here we used the *Flaveria* genus which contains species with close evolutionary distance and intermediate photosynthetic types to examine changes in stomatal traits during C_4 evolution. We showed that, in addition to the grass, the g_{smax} of C_4 plants in dicots was lower than that of C_3 plants, and there was a stepwise decrease during C_4 evolution indeed (Fig 1). Compared to the *SD* of C_3 species, the *SD* of C_4 species decreased (Fig. 2a, b). This decrease in *SD* is due to less stomata generation rather than augmented guard cell and epidermal cells (Fig. 2g). In two other closely related Cleomaecae species, *Gynandropsis gynandra*

(C_4) also has a lower *SD* than *Tareneya hassleriana* (C_3), consistent with the difference in their g_s (*69*).

The lower *SD* in phylogenetically closely related C_4 species compared to C_3 species raises the possibility that, during evolution from C_3 to C_4 species, the *SD* may gradually decrease. This is indeed shown in (Fig. 2a, b) where with the gradual transition of photosynthetic type from C_3 to type I, to type II, to C_4-like and to C_4, there is a gradually decrease in *SD*. Usually stomata conductance is significantly higher on the abaxial surface, as shown in maize, broad bean, and wild tomato species (*70-72*). Here we found that there is a highly significant difference in *SD* on the abaxial surface between the C_3 species and the C_4 species; by comparison, on the adaxial surface, though the difference between *SD* in the C_3 and C_4 species was smaller, C_4 species still had lower *SD* (Supplemental Fig 4,b). In contrast, the *SS* as represented by stomatal length was longer in intermediate species and C_4 species compared to C_3 species in *Flaveria*, consistent with previous reports that *SD* negatively correlates with *SS* (*8, 16*), suggesting that the *SS* is not the crucial contributor to the decreased g_{smax} and operational g_s in C_4 evolution. These evidences show that lower *SD* is the determinant of the commonly observed lower g_{smax} and correspondingly higher water use efficiency in C_4 compared to C_3 plants (*2, 73*). Therefore, lower SD could be another major feature of C_4 evolution, besides C_4 enzymes and Kranz anatomy.

Many earlier studies suggest that water limitation might have been a primary selection pressure for the evolution of C_4 photosynthesis (*5-7*). In *F. angustifolia*, a type I intermediate *Flaveria* species, when grown under outdoor natural light conditions, *SD*, g_{smax} and g_s were still similar to those in two C_3 *Flaveria* species, i.e. *F. cronquistii* and *F. pringlei* (*2*) (Fig. 1d, 3b), as also the case in *Helliotropin* (*74*). In contrast, the CO_2 concentration point of *F. angustifolia* was dramatically decreased (*75*) compared to their C_3 ancestors, as a result of an operating CO_2 concentrating mechanism of C_2 (*76*). This suggests that during the evolution from C_3 to intermediate species, where the movement of GDC from mesophyll cell to bundle sheath cell occurred, the primary evolutionary driving force should be to elevate CO_2 levels

around Rubisco, rather than H_2O saving (*77*). After this step, reducing water loss might became crucial. The emergence of C_2 mechanisms can increase leaf photosynthetic CO_2 uptake rate and biomass production, which can in principle create a higher water demand and increased the danger of embolism in the intermediate species, especially in the arid and semi-arid areas (*4*). This will create an even greater challenge once the complete C_4 photosynthetic metabolism emerged where photosynthetic CO_2 uptake rate and biomass production are further increased. Therefore, decreased *SD* and further increased water use efficiency observed in this study and earlier studies (*2, 59*) should be a required change during the transition from the intermediate species and later C_4 species.

Since *SD* is controlled by a molecular signaling network composed of many interacting components, many options can be used to alter *SD*. Out of these different options to regulate *SD*, here we present results which indicate that a down-regulation of the expression of *FSTOMAGEN*, a *STOMAGEN*-like gene, has been selected as a molecular switch to reduce *SD* during C_4 evolution in *Flaveria*, which is also the case in other C_4 lineages. When synthetic FSTOMAGEN peptide was applied on the leaves or *FSTOMAGEN* was over-expressed in Arabidopsis, the increases in SD were observed. Our evidences showed that the *FSTOMAGEN* positively regulates stomatal development. The result suggests that the declining expression of *STOMAGEN* homologs underlies the observed decrease in *SD* during the evolution of C_4 photosynthesis. In theory, Any of genes controlling stomatal patterning can be potentially selected to alter *SD* by regulation of expression during the emergence of C_4 photosynthesis. And in fact, these genes have innate mechanisms to alter their expression under different environments, such as low CO_2 (*78*). However, *STOMAGEN* homologs were shown to be the regulator selected to reduce *SD*. The two specialities of *STOMAGEN* probably have accounted for the phenomenon. First, *STOMAGEN* can regulates the entire molecular process of stomatal development, rather than *EPF1*, *EPF2*, *TMM* and the members of *ERECTA*, which individually only regulates particular phases of the stomatal development (*20, 21, 23, 24, 26, 27*). Second, which is potentially more important, *STOMAGEN* is expressed in the

mesophyll cell, while all other regulators controlling stomatal development are expressed in the cell of stomatal lineage (*24*, *26*, *27*, *30*, *31*). Having a genetic regulator controlling stomatal density in the mesophyll cell, where many of the C_4 related genes are expressed, can enable an easier coordination between stomatal development and C_4 photosynthesis. Future work is still needed to elucidate molecular level coordinating the expression of *STOMAGEN* and C_4 photosynthetic related genes.

Reference

1. O. V. Sedelnikova, T. E. Hughes, J. A. Langdale, Understanding the Genetic Basis of C_4 Kranz Anatomy with a View to Engineering C_3 Crops. *Annu Rev Genet* **52**, 249-270 (2018).
2. P. J. Vogan, R. F. Sage, Water-use efficiency and nitrogen-use efficiency of C_3-C_4 intermediate species of Flaveria Juss. (Asteraceae). *Plant, cell & environment* **34**, 1415-1430 (2011).
3. S. H. H. Taylor, S. P. Rees, M Ripley, B. S. Woodward, F. I. Osborne, C. P., Ecophysiological traits in C_3 and C_4 grasses: a phylogenetically controlled screening experiment. *New Phytol* **185**, 780-791 (2010).
4. C. P. Osborne, L. Sack, Evolution of C_4 plants: a new hypothesis for an interaction of CO_2 and water relations mediated by plant hydraulics. *Philos Trans R Soc Lond B Biol Sci* **367**, 583-600 (2012).
5. H. Zhou, B. R. Helliker, M. Huber, A. Dicks, E. Akcay, C_4 photosynthesis and climate through the lens of optimality. *Proc Natl Acad Sci U S A* **115**, 12057-12062 (2018).
6. G. Reeves *et al.*, Natural Variation within a Species for Traits Underpinning C_4 Photosynthesis. *Plant physiology* **177**, 504-512 (2018).
7. B. P. Williams, I. G. Johnston, S. Covshoff, J. M. Hibberd, Phenotypic landscape inference reveals multiple evolutionary paths to C_4 photosynthesis. *Elife* **2**, e00961 (2013).
8. P. J. Franks, P. L. Drake, D. J. Beerling, Plasticity in maximum stomatal conductance constrained by negative correlation between stomatal size and density: an analysis using Eucalyptus globulus. *Plant, cell & environment* **32**, 1737-1748 (2009).
9. T. Doheny-Adams, L. Hunt, P. J. Franks, D. J. Beerling, J. E. Gray, Genetic manipulation of stomatal density influences stomatal size, plant growth and tolerance to restricted water supply across a growth carbon dioxide gradient. *Philos Trans R Soc Lond B Biol Sci* **367**, 547-555 (2012).
10. S. H. Taylor *et al.*, Photosynthetic pathway and ecological adaptation explain stomatal trait diversity amongst grasses. *New Phytol* **193**, 387-396 (2012).
11. D. A. Way, What lies between : the evolution of stomatal traits on the road to C4 photosynthesis. *New phytologist* **193**, 291-293 (2012).
12. G. J. Dow, D. C. Bergmann, J. A. Berry, An integrated model of stomatal development and leaf physiology. *New Phytol* **201**, 1218-1226 (2014).
13. P. J. Franks, W. Doheny-Adams T, Z. J. Britton-Harper, J. E. Gray, Increasing water-use efficiency directly through genetic manipulation of stomatal density. *New Phytol* **207**, 188-195 (2015).

607 14. E. T. Hamanishi, B. R. Thomas, M. M. Campbell, Drought induces alterations in the stomatal
608 development program in Populus. *J Exp Bot* **63**, 4959-4971 (2012).
609 15. Y. Tanaka, S. S. Sugano, T. Shimada, I. Hara-Nishimura, Enhancement of leaf photosynthetic
610 capacity through increased stomatal density in Arabidopsis. *New Phytologist* **198**, 757-764
611 (2013).
612 16. P. J. F. a. D. J. Beerling, Maximum leaf conductance driven by CO_2 effects on stomatal size and
613 density over geologic time. *PNAS* **106**, 10343–10347 (2009).
614 17. S. h. Zhang, X. f. Xu, Y. m. Sun, J. l. Zhang, C. z. Li, Influence of drought hardening on the
615 resistance physiology of potato seedlings under drought stress. *Journal of Integrative*
616 *Agriculture* **17**, 336-347 (2018).
617 18. Peter J. Franks, G. D. Farquhar, The Effect of Exogenous Abscisic Acid on Stomatal
618 Development, Stomatal Mechanics, and Leaf Gas Exchange in Tradescantia virginiana. *Plant*
619 *physiology* **125**, 935–942 (2001).
620 19. L. J. Pillitteri, K. U. Torii, Mechanisms of stomatal development. *Annu Rev Plant Biol* **63**,
621 591-614 (2012).
622 20. N. Zoulias, E. L. Harrison, S. A. Casson, J. E. Gray, Molecular control of stomatal development.
623 *Biochem J* **475**, 441-454 (2018).
624 21. G. Lin *et al.*, A receptor-like protein acts as a specificity switch for the regulation of stomatal
625 development. *Genes Dev* **31**, 927-938 (2017).
626 22. E. D. Shpak, J. M. McAbee, L. J. Pillitteri, K. U. Torii, Stomatal patterning and differentiation by
627 synergistic interactions of receptor kinases. *Science* **309**, 290-293 (2005).
628 23. J. S. Lee *et al.*, Direct interaction of ligand-receptor pairs specifying stomatal patterning.
629 *Genes Dev* **26**, 126-136 (2012).
630 24. T. Shimada, S. S. Sugano, I. Hara-Nishimura, Positive and negative peptide signals control
631 stomatal density. *Cell Mol Life Sci* **68**, 2081-2088 (2011).
632 25. A. L. Rychel, K. M. Peterson, K. U. Torii, Plant twitter: ligands under 140 amino acids enforcing
633 stomatal patterning. *J Plant Res* **123**, 275-280 (2010).
634 26. K. Hara, R. Kajita, K. U. Torii, D. C. Bergmann, T. Kakimoto, The secretory peptide gene EPF1
635 enforces the stomatal one-cell-spacing rule. *Genes Dev* **21**, 1720-1725 (2007).
636 27. L. Hunt, J. E. Gray, The signaling peptide EPF2 controls asymmetric cell divisions during
637 stomatal development. *Curr Biol* **19**, 864-869 (2009).
638 28. J. S. Lee *et al.*, Competitive binding of antagonistic peptides fine-tunes stomatal patterning.
639 *Nature* **522**, 439-443 (2015).
640 29. L. J. Pillitteri, K. U. Torii, Breaking the silence: three bHLH proteins direct cell-fate decisions
641 during stomatal development. *Bioessays* **29**, 861-870 (2007).
642 30. S. S. Sugano *et al.*, Stomagen positively regulates stomatal density in Arabidopsis. *Nature* **463**,
643 241-244 (2010).
644 31. T. Kondo *et al.*, Stomatal density is controlled by a mesophyll-derived signaling molecule.
645 *Plant Cell Physiol* **51**, 1-8 (2010).
646 32. L. Hunt, K. J. Bailey, J. E. Gray, The signalling peptide EPFL9 is a positive regulator of stomatal
647 development. *New Phytol* **186**, 609-614 (2010).
648 33. S. Ohki, M. Takeuchi, M. Mori, The NMR structure of stomagen reveals the basis of stomatal
649 density regulation by plant peptide hormones. *Nat Commun* **2**, 512 (2011).

650 34. U. Gowik, A. Bräutigam, K. L. Weber, A. P. M. Weber, P. Westhoff, Evolution of C_4
651 Photosynthesis in the Genus Flaveria: How Many and Which Genes Does It Take to Make C_4?
652 *The Plant Cell* **23**, 2087-2105 (2011).
653 35. J. Mallmann *et al.*, The role of photorespiration during the evolution of C_4 photosynthesis in
654 the genus Flaveria. *Elife* **3**, e02478 (2014).
655 36. M. Hronkova *et al.*, Light-induced STOMAGEN-mediated stomatal development in
656 Arabidopsis leaves. *J Exp Bot* **66**, 4621-4630 (2015).
657 37. J. Lu *et al.*, Homologous genes of epidermal patterning factor regulate stomatal development
658 in rice. *J Plant Physiol* **234-235**, 18-27 (2019).
659 38. L. Xuxin, The Effects for Stomata Developmental on Arabidopsis by Overexpression of Rice
660 EPFL Family Genes. *Molecular Plant Breeding* **15**, 2042- 2047 (2017).
661 39. X. Yin *et al.*, CRISPR-Cas9 and CRISPR-Cpf1 mediated targeting of a stomatal developmental
662 gene EPFL9 in rice. *Plant Cell Rep* **36**, 745-757 (2017).
663 40. J. Y. Zhang, S. B. He, L. Li, H. Q. Yang, Auxin inhibits stomatal development through
664 MONOPTEROS repression of a mobile peptide gene STOMAGEN in mesophyll. *Proc Natl Acad*
665 *Sci U S A* **111**, E3015-3023 (2014).
666 41. Z. Xu, G. Zhou, Responses of leaf stomatal density to water status and its relationship with
667 photosynthesis in a grass. *J Exp Bot* **59**, 3317-3325 (2008).
668 42. M. J. Lyu *et al.*, RNA-Seq based phylogeny recapitulates previous phylogeny of the genus
669 Flaveria (Asteraceae) with some modifications. *BMC Evol Biol* **15**, 116 (2015).
670 43. A. L. Ming-Ju *et al.*, The Coordination and Jumps along C_4 Photosynthesis Evolution in the
671 Genus Flaveria. *bioRxiv*, 460287 (2018).
672 44. K. Billakurthi *et al.*, Transcriptome dynamics in developing leaves from C_3 and C_4 Flaveria
673 species reveal determinants of Kranz anatomy. *bioRxiv*, (2018).
674 45. M. G. Grabherr *et al.*, Full-length transcriptome assembly from RNA-Seq data without a
675 reference genome. *Nat Biotechnol* **29**, 644-652 (2011).
676 46. B. Li, C. N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or
677 without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
678 47. L. Wang *et al.*, Comparative analyses of C_4 and C_3 photosynthesis in developing leaves of
679 maize and rice. *Nat Biotechnol* **32**, 1158-1165 (2014).
680 48. X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript
681 server. *Nucleic Acids Res* **42**, W320-324 (2014).
682 49. D. R. F. Kyte J, A simple method for displaying the hydropathic character of a protein. *Journal*
683 *of Molecular Biology* **157**, 105-132 (1982).
684 50. S. Perveen *et al.*, Overexpression of maize transcription factor mEmBP-1 increases
685 photosynthesis, biomass, and yield in rice. *J Exp Bot* **71**, 4944-4957 (2020).
686 51. D. M. Emms, S. Kelly, OrthoFinder: solving fundamental biases in whole genome comparisons
687 dramatically improves orthogroup inference accuracy. *Genome Biol* **16**, 157 (2015).
688 52. R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput.
689 *Nucleic Acids Res* **32**, 1792-1797 (2004).
690 53. A. Stamatakis, RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with
691 thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-2690 (2006).
692 54. F. Ronquist *et al.*, MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice
693 across a large model space. *Syst Biol* **61**, 539-542 (2012).

694 55. S. A. Shiryev, J. S. Papadopoulos, A. A. Schaffer, R. Agarwala, Improved BLAST searches using
695 longer words for protein seeding. *Bioinformatics* **23**, 2949-2951 (2007).
696 56. X. J. Min, G. Butler, R. Storms, A. Tsang, OrfPredictor: predicting protein-coding regions in
697 EST-derived sequences. *Nucleic Acids Res* **33**, W677-680 (2005).
698 57. J. L. J. a. T. F. Ø. Claus Lindbjerg Andersen, Normalization of real-time quantitative reverse
699 transcription-PCR data a model-based variance estimation approach to identify genes suited
700 for normalization, applied to bladder and colon cancer data sets. *Cancer Res* **64**, 5245-5250
701 (2004).
702 58. D. P. K. Vandesompele J, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F., Accurate
703 normalization of real-time quantitative RT-PCR data by geometric averaging of multiple
704 internal control genes. *Genome Biology* **3**, research0034.0031–0034.0011 (2002).
705 59. D. A. Way, G. G. Katul, S. Manzoni, G. Vico, Increasing water use efficiency along the C_3 to C_4
706 evolutionary pathway: a stomatal optimization perspective. *J Exp Bot* **65**, 3683-3693 (2014).
707 60. V. S. R. D. a. M. Santakumari, Stomatal characteristics of some dicotyledonous plants in
708 relation to the C_4 and C_3 pathways of photosynthesis. *Plant and cell physiology* **18**, 935-938
709 (1977).
710 61. M. E. K. a. S. A. M. Marc D. Abrams, Relating Wet and Dry Year Ecophysiology to Leaf
711 Structure in Contrasting Temperate Tree Species. *Ecology* **75**, 123-133 (1994).
712 62. C. T. a. D. R. Murray, Effects of Elevated Atmospheric CO_2 Concentration on Leaf Anatomy and
713 Morphology in Panicum Species Representing Different Photosynthetic Modes. *International*
714 *Journal of Plant Sciences* **160**, 1063-1073 (1999).
715 63. A. S. Soares *et al.*, Adaxial/abaxial specification in the regulation of photosynthesis and
716 stomatal opening with respect to light orientation and growth with CO_2 enrichment in the C_4
717 species Paspalum dilatatum. *New Phytologist* **177**, 186-198 (2007).
718 64. I. Gindel, Stomatal Number and Size as Related to Soil Moisture in Tree Xerophytes in Israel
719 *Ecological Society of America* **50**, 263-267 (1968).
720 65. V. S. Pathare, N. Koteyeva, A. B. Cousins, Increased adaxial stomatal density is associated with
721 greater mesophyll surface area exposed to intercellular air spaces and mesophyll
722 conductance in diverse C4 grasses. *The New phytologist* **225**, 169-182 (2019).
723 66. J. A. Cardoso, M. Pineda, L. Jimenez Jde, M. F. Vergara, I. M. Rao, Contrasting strategies to
724 cope with drought conditions by two tropical forage C_4 grasses. *AoB Plants* **7**, plv107 (2015).
725 67. E. J. Edwards, C. J. Still, M. J. Donoghue, The relevance of phylogeny to studies of global
726 change. *Trends Ecol Evol* **22**, 243-249 (2007).
727 68. E. J. Edwards, C. J. Still, Climate, phylogeny and the ecological distribution of C4 grasses. *Ecol*
728 *Lett* **11**, 266-276 (2008).
729 69. S. Aubry *et al.*, A Specific Transcriptome Signature for Guard Cells from the C_4 Plant
730 Gynandropsis gynandra. *Plant physiology* **170**, 1345-1357 (2016).
731 70. S. P. Driscoll, A. Prins, E. Olmos, K. J. Kunert, C. H. Foyer, Specification of adaxial and abaxial
732 stomata, epidermal structure and photosynthesis to CO_2 enrichment in maize leaves. *J Exp*
733 *Bot* **57**, 381-390 (2006).
734 71. W.-H. W. Xi-Qing Wang, and Sarah M. Assmann, Differential Responses of Abaxial and Adaxial
735 Guard Cells of Broad Bean to Abscisic Acid and Calcium. *Plant physiology* **118**, 1421–1429
736 (1998).

72. C. D. Muir, M. À. Conesa, J. Galmés, Independent evolution of ab- and adaxial stomatal density enables adaptation. *biorxiv*, (2015).
73. R. J. Webster *et al.*, High C_3 photosynthetic capacity and high intrinsic water use efficiency underlies the high productivity of the bioenergy grass Arundo donax. *Sci Rep* **6**, 20694 (2016).
74. P. J. Vogan, M. W. Frohlich, R. F. Sage, The functional significance of C_3-C_4 intermediate traits in Heliotropium L. (Boraginaceae): gas exchange perspectives. *Plant, cell & environment* **30**, 1337-1345 (2007).
75. J. W. Maurice S. B. Ku, Ziyu Dai, Rick A. Scott, Chun Chu, and Gerald E. Edwards, Photosynthetic and Photorespiratory Characteristics of Flaveria species. *Plant physiology* **96**, 518-528 (1991).
76. O. Keerberg, T. Parnik, H. Ivanova, B. Bassuner, H. Bauwe, C_2 photosynthesis generates about 3-fold elevated leaf CO_2 levels in the C_3-C_4 intermediate species Flaveria pubescens. *J Exp Bot* **65**, 3649-3656 (2014).
77. R. F. Sage, T. L. Sage, F. Kocacinar, Photorespiration and the evolution of C_4 photosynthesis. *Annu Rev Plant Biol* **63**, 19-47 (2012).
78. C. B. Engineer *et al.*, Carbonic anhydrases, EPF2 and a novel protease mediate CO_2 control of stomatal development. *Nature* **513**, 246-250 (2014).

Acknowledgements

The work was funded by Strategic Priority Research Program of the Chinese Academy of Sciences (grant number: XDB27020105) and the general program of National Science Foundation of China (31870214). We thank Prof. Rowan F. Sage (University of Toronto) for providing the Flaveria cuttings.

Author contributions

YYZ and XGZ designed the study. YYZ observed and measured the stomatal characteristic. YYZ carried out the molecular experiment and the treatment of synthetic peptides in vivo. FFM conducted the transgene experiment. GYC guided the treatment of peptides in vivo. Amy MJL analyzed the one kp RNA-seq data. YYZ and XGZ interpreted results and wrote the article.

Competing interests

The authors declare no competing financial interests.

Figure Legend

Fig 1: Maximum stomatal conductance (g_{smax}) of *Flaveria* species

a, Schematic of the progresssion of C_4 evolution in *Flaveria* genus. **b**, A stomatal graph taken with a light microscope showing stomatal parameters for *Flaveria* species which are needed to be measured to calculate the g_{smax}. The graph of stomata is *rob*. *l* (stomatal length), PL (pore length) and GCW (guard cell width). **c**, Correlation of the guard cell width and stomatal length (left), correlation of the stomatal length and pore length (right). **d**,The change of g_{smax} on the abaxial surface in *Flaveria* species with different photosynthetic types from C_3 to C_4 grown outdoors (left) and in greenhouse (right). **e**, The difference in g_{smax} of the abaxial surface between species with different photosynthetic types in the *Flaveria* genus grown outdoors (left) and in greenhouse (right). **f**, Correlation of g_{smax} and g_s for *Flaveria* species with different photosynthetic types. Error bar indicates s.d. Different letters represent statistically significant differences (one-way ANOVA, Duncan's multiple range test (α = 0.05)). Pearson correlation coefficient and the significance were noted in each panel.

Fig 2: Stomatal density (*SD*) and its patterning.

a, The difference in *SD* of abaxial surface between species with different photosynthetic types in *Flaveria* genus grown outdoors (left) and in greenhouse (right). **b**, The change of *SD* of abaxial surface from C_3 to C_4 in *Flaveria* species grown outdoors (left) and in greenhouse (right). (n=4, biologically independent replicates). **c,** Correlation of g_{smax} and *SD* for different *Flaveria* species. **d**, Correlation of g_{smax} and *l* for different *Flaveria* species. **e**, Correlation of *l* and *SD* for different *Flaveria* species. **f,** The photographs of stomatal patterning on abaxial surface in *rob*, *ram* and *bid*. The photographs from the left to right are for *rob*, *ram* and *bid* grown outdoors (top) and in greenhouse (bottom). **g**, Comparison of stomatal index (*SI*) between *rob*, *ram*, and *bid*. Different colours represent plants grown outdoors (blue) or in greenhouse (green). *l*, (n=20, biologically independent replicates). *SI*, (n=3, biologically independent replicates). Error bar indicates s.d. Different letters represent

statistically significant differences (one-way ANOVA, Duncan's multiple range test ($\alpha = 0.05$)). Pearson correlation coefficient and the significance are noted in each panel.

Fig 3: Transcript abundance of stomatal developmental related genes between C_3 and C_4 *Flaveria* species.

a, A model of the signaling pathway for the stomatal development in Arabidopsis. A protodermal cell (PDC, white) becomes a meristemoid mother cell (MMC, orange), which divides asymmetrically into two daughter cells, with one being a meristemoid (M, red) and the other being a stomatal-lineage ground cell (SLGC). The M cell develops into a guard mother cell (GMC, yellow); the SLGC develops into a pavement cell (PC). The GMC conducts a symmetrical division to produce two equal-sized guard cells (GC, green). EPF1, EPF2 and STOMAGEN competitively bind to ER and TMM, which can deliver a signal to the YDA MAPK cascade. SPCH, MUTE and FAMA ultimately are inactivated through phosphorylation by the YDA MAPK cascade. **b**, Comparison of the expression of stomatal developmental related genes between juvenile and mature leaves. **c**, The foldchanges of stomatal development related genes for the C_3 and C_4 *Flaveria* species. Fold change means the ratio between the gene expression in a C_3 *Flaveria* specie and gene expression in a C_4 *Flaveria* specie under the same leaf development stage and growth condition as the C_3 *Flaveria* species. The expression data were from (*44*), (*35*), (*34*) and One Thousand Plants (1KP). **d,** The volcano plot for the expression of the stomatal development related genes between C_3 and C_4 *Flaveria* species. The expression data were from (*35*) and (*44*). **e**, Relative expression levels determined by RT-qPCR and expression level determined by RNA-seq of FSTO (FSTOMAGEN) in *rob* and *bid* (n=4, biologically independent replicates). **f**, Protein conservation analysis of the homologs of FSTOMAGEN between different species in the *Flaveria* genus. The amino acid regions marked by green line and red line represent the signal peptide and the functional domain of STOMAGEN, respectively. Red letters represent conserved amino acids in most species; white letters with red background represent amino acids conserved across all species.

Fig 4: The Function of *FSTOMAGEN* in controlling the *SD*

a, Protein conservation analysis of the homologs of FSTOMAGEN in different species. The amino acid regions marked by green line and red line represent the signal peptide and the functional domain of FSTOMAGEN, respectively. Red letters represent conserved amino acids in most species, white letters with red backgrounds represent amino acid conserved across all species. **b**, Phylogenetic tree for FSTOMAGEN. blue represents dicotyledon and red represents monocotyledon. **c**, The structure of recombinant genes used to transform Arabidopsis. AFSTO represents the recombination of the signal peptide of STOMAGEN from Arabidopsis and the functional domain of FSTOMAGEN from *F.robusta*. FSTO represents the intact STOMAGEN from *F.robusta*. **d**, Photographs showing clustered stomata on a leaf for the transgenic Arabidopsis. Scar bar represents 10 um. **e**, **h**, The *SD* (**e**) and RNA-level (**h**) of the different transgenic Arabidopsis (FSTO) lines. **f**, **i**, The *SD* (**f**) and RNA-level (**i**) of the different transgenic Arabidopsis (AR) lines. **g**, Comparison of the transgenic FSTO lines and AR lines. **j**, Photographs of stomata in a leaf from Arabidopsis with FSTO and AFSTO over-expressed. Scar bar represents 20 um. Error bar indicates s.d. The n.s. represent statistically no significant differences ($P>0.05$, t-test, two-tailed distribution, unpaired).

Fig 5: The FSTOMAGEN peptide was applied to plants in vitro.

a, Photographs of leaf surface taken with differential interference contrast microscope showing the increased stomata density in *F. bidentis* (*bid*) leaves with FSTOMAGEN applied. Scale bar represents 20 μm. **b,** Photographs of leaf surface taken with differential interference contrast microscope showing the increased stomata density in Arabidopsis leaves with FSTOMAGEN applied. Scale bar represents 20 μm. **c**, The stomatal density for Arabidopsis leaves with FSTOMAGEN applied at different concentrations (n=8, biologically independent replicates). **d**, The stomatal density for *bid* leaves with FSTOMAGEN applied at different concentrations (n=8, biologically independent replicates). **e**, The stomatal index (SI) for *bid* leaves with FSTOMAGEN

applied at different concentrations (n=8, biologically independent replicates). Error bar indicates s.d. The asterisks represent statistically significant differences (P<0.05, t-test, two-tailed distribution, unpaired).

Fig 6: Expression level of *FSTOMAGEN* at different stages of C_4 evolution.

a, Expression levels of *FSTOMAGEN* at the different developmental stages of leaves in *rob* and *bid*. P denotes leaf primordia. Approximate leaf length for the different leaf stages was 0.3 cm (stage 1), 0.5 cm (stage 2), 1.5 cm (stage 3), 2 cm (stage 4), 3 cm (stage 5), 4 cm (stage 6), 6 cm (stage7) , 8 cm (stage 8) and 10 cm (stage 9). **b**, Comparison of relative expression levels of *FSTOMAGEN* between *rob*, *ram* and *bid*. (n=3, biologically independent replicates). **c,** Comparison of the expression levels of *FSTOMAGEN* between different species with different photosynthetic types in *Flaveria* genus. For RT-qPCR, C_3 species incude *pri* (n=5) and *rob* (n=13), intermediates species includes *ang* (n=3), *ram* (n=5) and *vag* (n=3), C_4 species include *tri* (n=4), *aus* (n=3), *bid* (n=4) and *koc* (n=3). For RNA-seq, C_3 species include *pri* (n=4) and *rob* (n=4), intermediates species includes *ano* (n=4), *pub* (n=4), *chl* (n=4), *bro* (n=4), C_4 speices include *tri* (n=4) and *bid* (n=4). The relative expression indicates the measurements with RT-qPCR, and the others were measured with RNA-seq. The geometric means of *EF1a* and *ACT7* were used as the reference in the calculation of expression levels for other genes. Error bars indicate s.d. The asterisks represent statistically significant differences (P<0.05, t-test, two-tailed distribution, unpaired). Different letters represent statistically significant differences (one-way ANOVA, Duncan's multiple range test (α = 0.05)).

Fig 7, Comparison of stomatal pattern and expression levels of stomatal development related genes in the C_4 lineages other than *Flaveria*

a, Stomatal characteristics for rice and maize. *SD*, (n=8, biologically independent replicates); *l*, (n=25, biologically independent replicates); *SI*, (n=3, biologically independent replicates); g_s, (n=8, biologically independent replicates). Error bar

indicates s.d. The asterisks represent statistically significant differences (P<0.05, t-test, two-tailed distribution, unpaired). **b**, Photographs of stomatal pattern in rice and maize. Scale bars represent 50 μm. **c**, Comparison of transcript abundance of stomatal development related genes between rice and maize. The developmental stages of leaves were divided into 11 and 15 parts in rice and maize, respectively. R represents the leaves of rice, M represents the leaves of maize. **d**, Comparison of the expression levels of *STOMAGEN* homologs between rice and maize. There are one and two paralogs of *STOMAGEN* that is expressing in rice and maize, respectively. *OsSTO* represents the homologs of *STOMAGEN* in rice. *ZmSTO* represents the homologs of *STOMAGEN* in maize. **e**, The families and numbers of species with different photosynthetic type that we used to analyse. **f**, Comparison of the expression of C_4 related genes between species with different photosynthetic types. **g**, Comparison of the expression of *STOMAGEN* homologs between species with different photosynthetic types. The asterisks represent statistically significant differences (P<0.05, t-test, two-tailed distribution, unpaired).

Fig 8：Schematic of the role of *FSTOMAGEN* in the changes of WUE during C_4 evolution. During the C_4 evolution, *FSTOMAGEN* (*FSTO*) changes g_{smax} through regulating stomatal density, which in turn affectes gas exchange and improves the water use efficiency of C_4 plants. C_3-C_4 denotes the intermediate plants between C_3 and C_4 plants during C_4 evolution.

Legends for Supplemental Figures

Supplementary fig 1. Comparison of *SD* between C_3 and C_4 species. **a**, Comparison of *SD* on the abaxial surface of leaves between 70 C_3 and 47 C_4 species, with data collected from literature (see details in the discussion section). b, Comparison of *SD* on adaxial surface of leaves between 26 C_3 and 27 C_4 species with data collected from literature (see details in the discussion section). Error bar indicates s.d. The asterisks represent statistically significant differences (P<0.05, t-test, two-tailed distribution, unpaired).

Supplementary fig 2. a, Stomatal conductance (g_s) responses to net photosynthesis rate (A) under saturated light. The circles on the graph indicate larger g_s values in different photosynthetic types (Date from (*2, 59*)). **b**, The correlation of g_{smax} (abaxial + adxial) and g_s. **c**, G_s responses to A at the growth light in two *Flaveria* species (*rob* and *bid*) (n=6).

Supplementary fig 3. g_{smax} on the adxial surface of leaves in *Flaveria*. **a**, The difference in g_{smax} in species with different photosynthetic types grown outdoors (left) or in greenhouse (right)**. b,** The changes of g_{smax} on the adaxial leaf surface from C_3 to C_4 species in the *Flaveria* genus grown outdoors (left) and in greenhouse (right). (n=3, biologically independent replicates). Correlation of g_{smax} and g_s on the adaxial surface of leaves. Different letters represent statistically significant differences (one-way ANOVA, Duncan's multiple range test (α = 0.05)).

Supplementary fig 4. *SD* on the adaxial surface of leaves in *Flaveria*. **a**, The difference in *SD* in species with different photosynthetic types grown outdoors (left) or in greenhouse (right)**. b**, The changes of *SD* on the adaxial leaf surface from C_3 to C_4 species in the *Flaveria* genus grown outdoors (left) and in greenhouse (right). (n=3, biologically independent replicates). **c**, Correlation of g_{smax} and *SD* on the adaxial surface of leaves. **d**, Correlation of g_{smax} and l on the adaxial surface of leaves. **e**, Correlation of l and *SD* on the adaxial surface of leaves. Different letters represent statistically significant differences (one-way ANOVA, Duncan's multiple range test (α = 0.05)).

Supplementary fig 5. *SD* in *rob* (C_3), *ram* (intermediate) and *bid* (C_4) grown from seeds. The *SD* on abaxial (ab) (n=5, biologically independent replicates) and adaxial (ad) (n=4, biologically independent replicates) surfaces were observed in *rob*, *ram* and *bid* grown in greenhouse. Different colours represent different photosynthetic types. Different letters represent statistically significant differences (one-way ANOVA, Duncan's multiple range test (α = 0.05)).

Supplementary fig 6. Conservative analysis of amino acid sequences for EPF1 and EPF2. **a**. Comparison of amino acid sequences of homologs of EPF1 between

different *Flaveria* species and Arabidopsis. **b**. Comparison of amino acid sequences of homologs of EPF2 between *Flaveria* species and Arabidopsis. White letters with red background represent amino acids conserved across all species. Red letters represent amino acids with similar biochemical property.

Supplementary fig 7. Hydrophobicity analysis of proteins. **a**, the protein of stomagen (left) and the protein of FSTOMAGEN (right). **b**, the protein of EPF1 (left) and the protein of FEPF1 (right). **c**, the protein of EPF2 (left) and the protein of FEPF2 (right).

Supplementary fig 8. Assessment of the stability of expression levels of reference genes. **a**,**b**, Comparison of the stability of expression levels of candidate reference genes in the *Flaveria* species by RNA-seq. **c**, Gel electrophoresis showing the specificity of the primers for candidate reference genes. From left to right, *ACT7*, *EF1a*, *J3*, *PNdO*, *CPN20*, *UBC9*, *GAPDH*, *F2N1.5*, *PSRP4*, *CACMS*, *HEME2*, and *UBQ11*. **d**, The coefficient of variations (CV) for the candidate reference genes. **e**, The expression stability for candidate reference genes calculated by Genorm. **f**, The expression stability for candidate reference genes calculated by Normfinder.

Supplementary fig 9. The expression levels of stomatal development related genes in *Flaveria* by RNA-seq. Expression levels of stomatal development related genes in species with different photosynthetic types in the *Flaveria* genus.

Supplementary fig 10. Relative expression levels of *FSTOMAGEN* and *SD* for *rob* (C_3) under different light conditions. (A) Relative expression levels for *FSTOMAGEN* under two different light conditions. (B) *SD* under two different light conditions. Natural light: the maximal photosynthetic photon flux density in the greenhouse was about 1300~1400 mmol m^{-2} s^{-1}. Light in the phytotron: 20~100 mmol m^{-2} s^{-1}. Reference gene used to calculate the expression level was *ACT7*. The asterisks represent statistically significant differences (P<0.05, t-test, two-tailed distribution, unpaired).

Supplementary fig 11. Relative expression levels of stomata development related genes determined with different reference genes and different primers. **a, b, c,** Quantification with different primers and the same *EF1a* reference gene. Relative

expressions shown in **a and b** were determined using the same primer pairs while that in **c** used other primer pairs (See details of the primers used in Table S2). **d,** Relative expression levels determined with the gene *J3* used as the reference gene.

Supplementary fig 12. Transcript abundance of genes in the C_4 cycle and C_4 related transporters. J represents juvenile leaves and M represents mature leaves. Different colours represent species with different photosynthetic types in the *Flaveria* genus. Red colour represents C_3 species, grey colour represents C_3-C_4 species, green colour represents C_4-like species, and blue colour represents C_4 species.

Supplementary fig 13. Alignment of the functional domain of *STOMAGEN* with transcripts from different *Flaveria* species by Blast. From the top to bottom, *F. robusta*, *F. pringlei*, *F. angustifolia*, *F. ramosissima*, *F. vaginata*, *F. australasica*, *F. bidentis*.

Supplementary fig 14. The expression level of *FSTOMAGEN* in different *Flaveria* species based on RT-qPCR (**a**) or RNA-seq (**b**). **c**, The correlation of *SD* and *FSTOMAGEN* expression in *Flaveria* species.

Supplementary fig 15. Comparison of the expression level of *FEPF1* and *FEPF2* between *rob* and *bid Flaveria* species.

Supplementary fig 16. The analysis for the *SD* of T2 generation of transgenic Arabidopsis (AFSTO and FSTO). **a**, Comparison of *SD* between WT and transgenic Arabidopsis AR, FSTO respectively. **b**, RNA level identification of transgenic Arabidospsis. Two independent lines was analyzed for each trangene event.

Supplementary fig 17. The increase of *SD* in cotyledon of *F. bidentis* (*bid*) with in vitro application of FSTOMAGEN. Photographs of the leaf surface in the untreated (control) and treated bid plants. scale bar, 20 mm.

Supplementary fig 18. The gene tree of STOMAGEN based on bayesian-inference method.

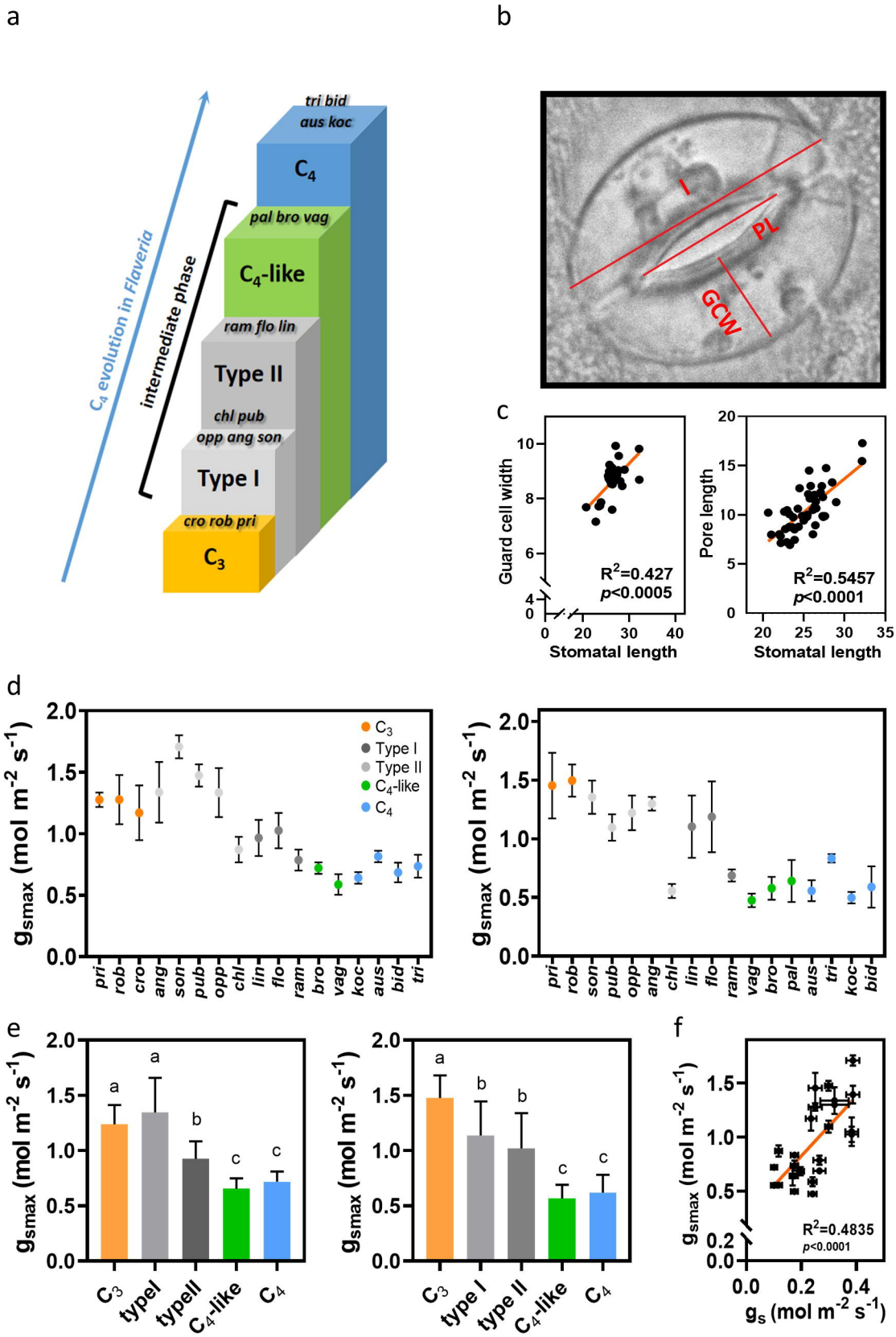
Supplementary table 1. Statistics for the mapping of the RNA-seq data of the juvenile and mature leaves in different species of the *Flaveria* genus.

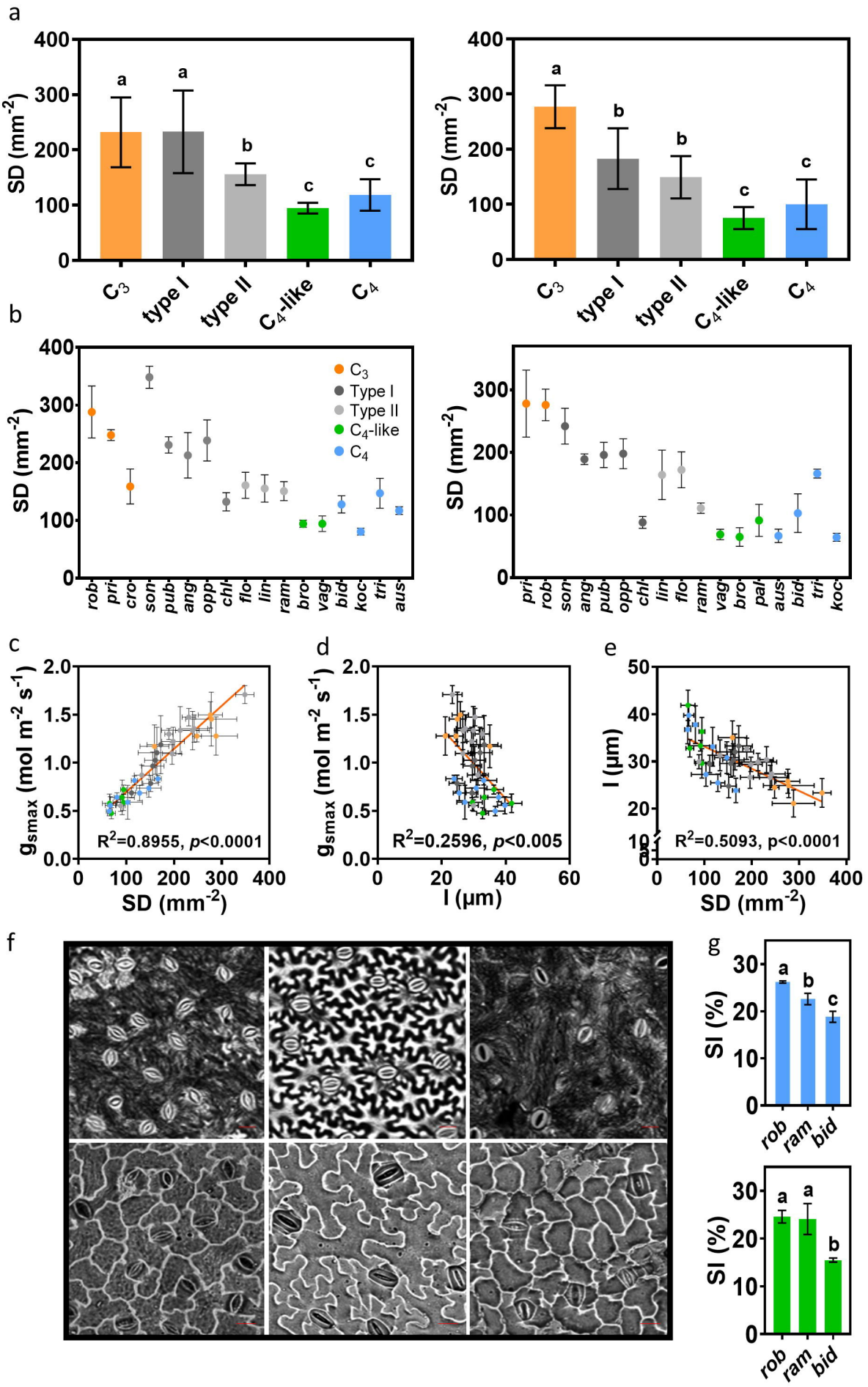
Supplementary table 2. Primers used in this study

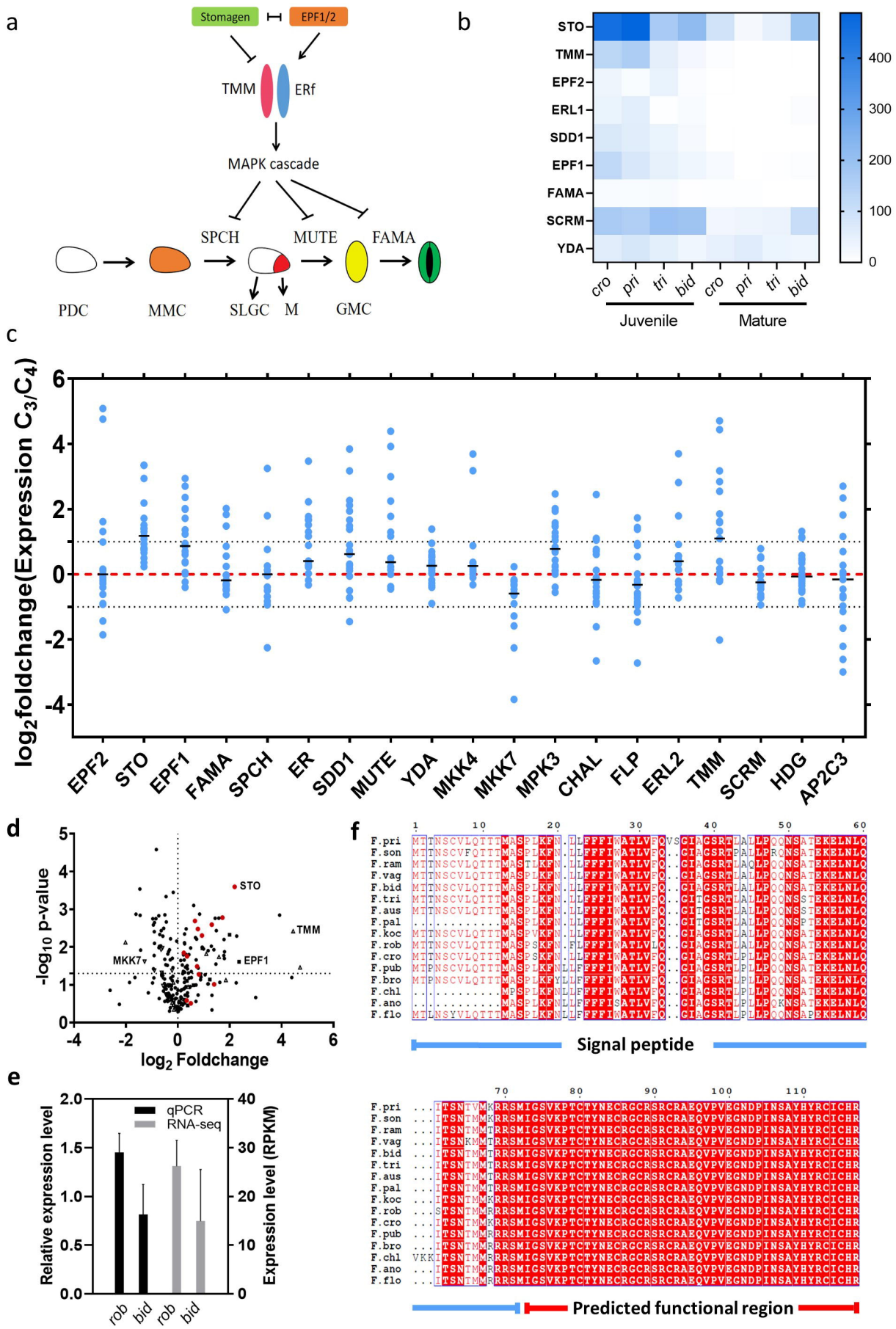
Supplementary table 3. The effect of genes related to stomatal development in Arabidopsis on stomatal phenotypes.

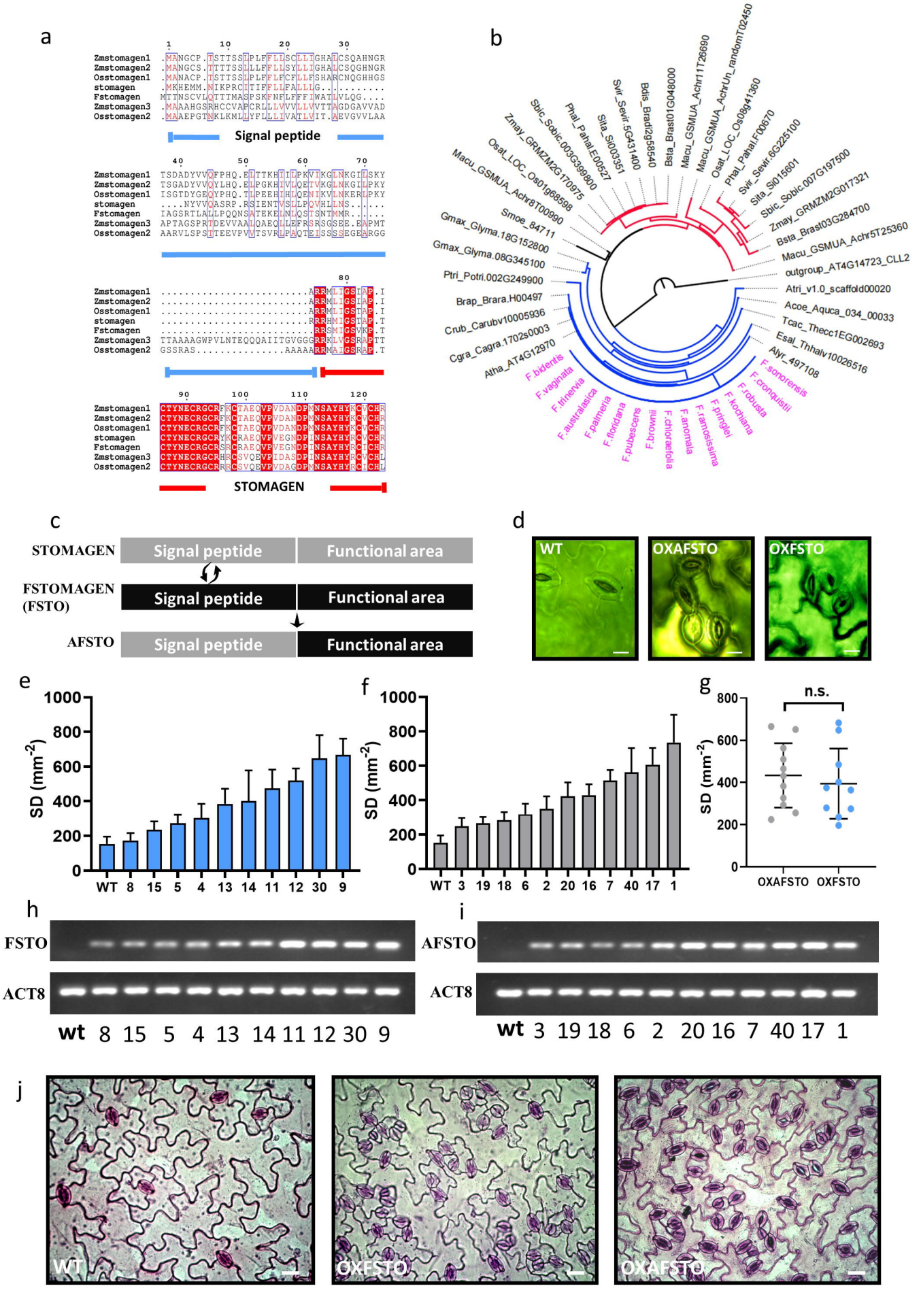
Supplementary table 4. The N50 of the mapping of the RNA-SEQ data taken from 1KP database for the juvenile and mature leaves

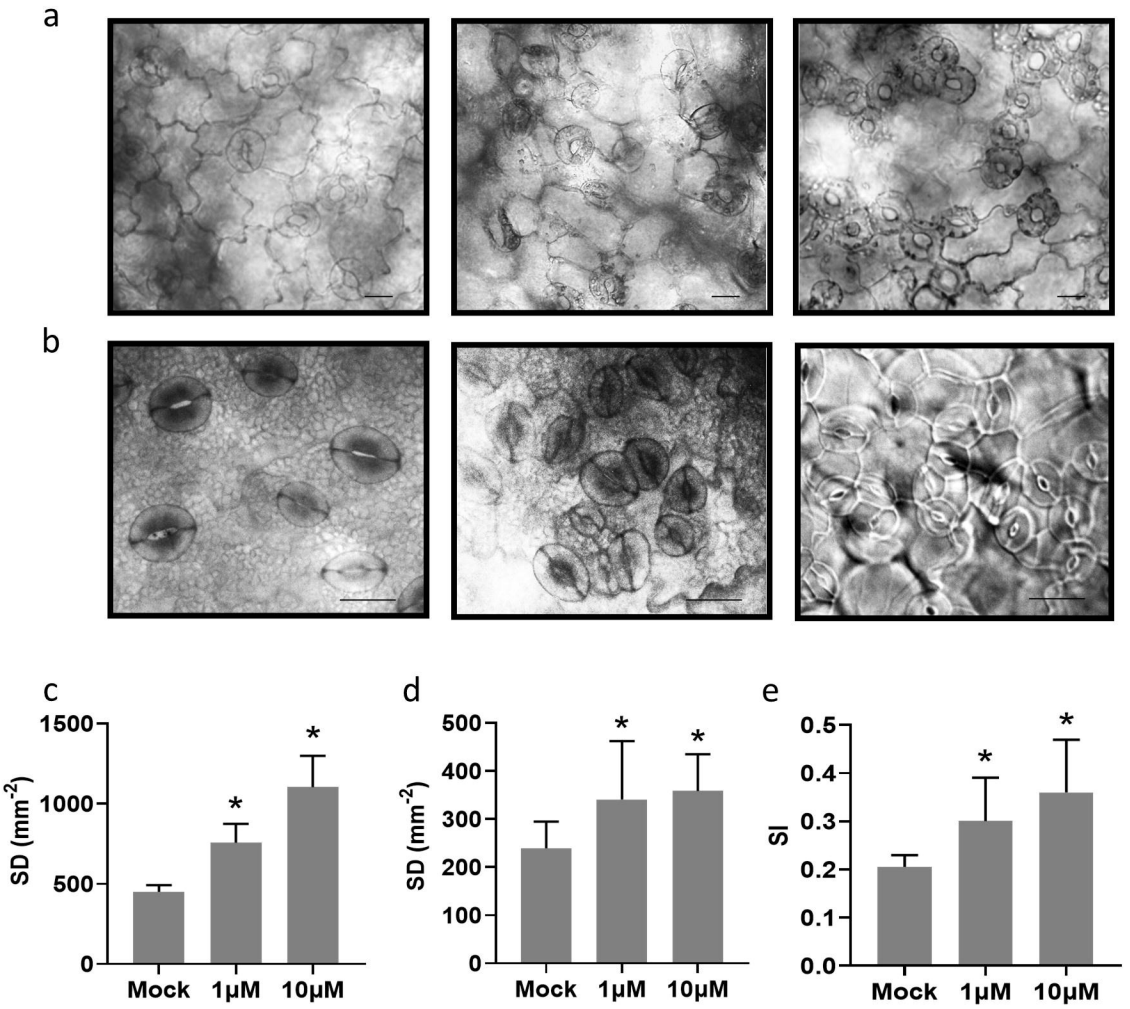
Supplementary table 5. The mapping rates of the mapping of the RNA-SEQ data taken from 1KP database for the juvenile and mature leaves.

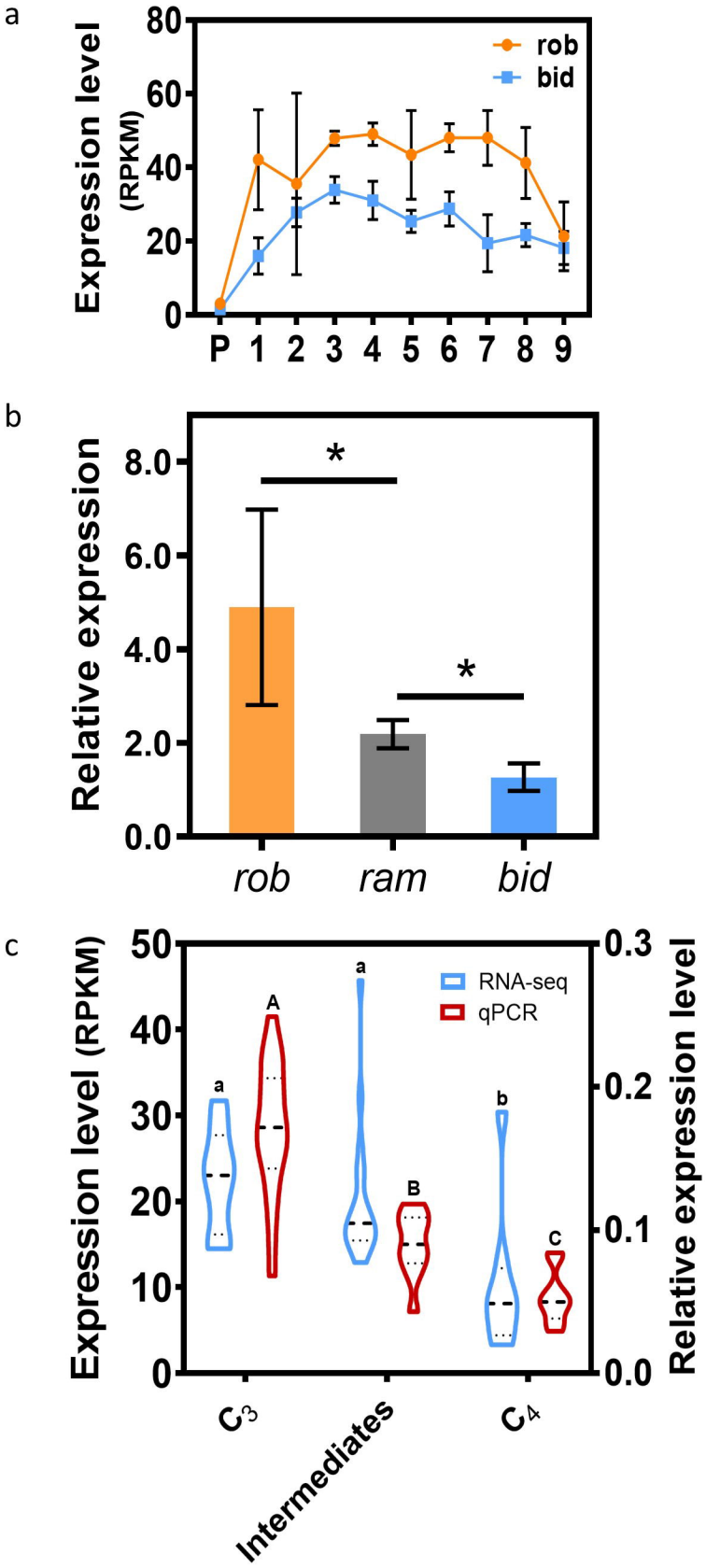

a
b
tri bid
aus koc
C4
pal bro vag
C4-like
ram flo lin
Type II
chl pub
opp ang son
Type I
cro rob pri
C3
C4 evolution in Flaveria
intermediate phase
I
PL
GCW
c
Guard cell width
Stomatal length
R2=0.427
p<0.0005
Pore length
Stomatal length
R2=0.5457
p<0.0001
d
gsmax (mol m-2 s-1)
C3
Type I
Type II
C4-like
C4
pri rob cro ang son pub opp chl lin flo ram bro vag koc aus bid tri
pri rob son pub opp ang chl lin flo ram vag bro pal aus tri koc bid
e
gsmax (mol m-2 s-1)
C3 typeI typeII C4-like C4
C3 type I type II C4-like C4
f
gsmax (mol m-2 s-1)
gs (mol m-2 s-1)
R2=0.4835
p<0.0001

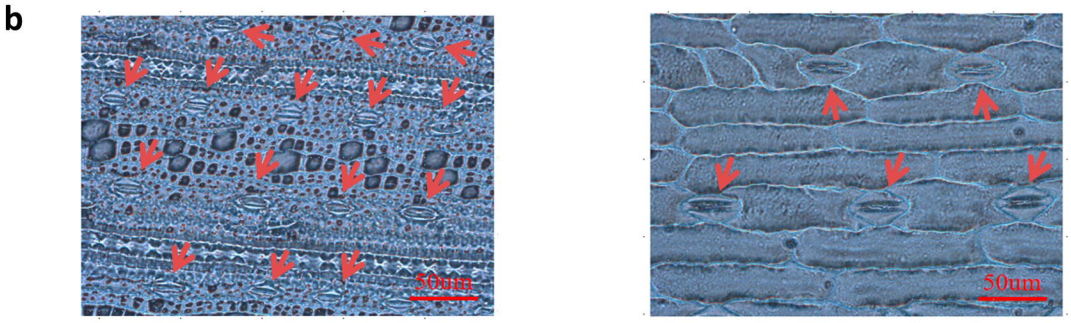

a
SD (mm-2)
C3
type I
type II
C4-like
C4
b
Type I
Type II
C4-like
rob pri cro son pub ang opp chl flo lin ram bro vag bid koc tri aus
pri rob son ang pub opp chl lin flo ram vag bro pal aus bid tri koc
c
gsmax (mol m-2 s-1)
R2=0.8955, p<0.0001
d
R2=0.2596, p<0.005
e
l (µm)
R2=0.5093, p<0.0001
f
g
SI (%)
rob ram bid

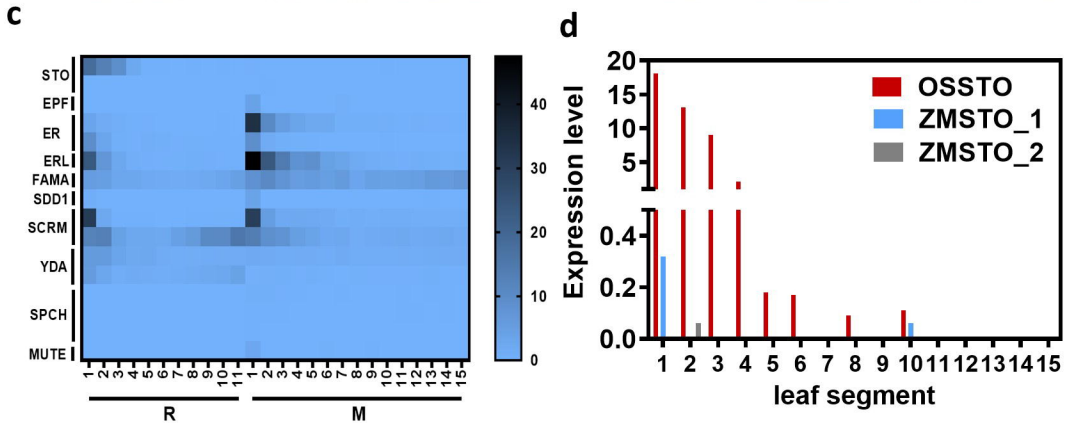

a
Stomagen
EPF1/2
TMM
ERf
MAPK cascade
SPCH
MUTE
FAMA
PDC
MMC
SLGC
M
GMC
b
STO
TMM
EPF2
ERL1
SDD1
EPF1
FAMA
SCRM
YDA
cro
pri
tri
bid
Juvenile
Mature
400
300
200
100
0
c
log2foldchange(Expression C3/C4)
EPF2
STO
EPF1
FAMA
SPCH
ER
SDD1
MUTE
YDA
MKK4
MKK7
MPK3
CHAL
FLP
ERL2
TMM
SCRM
HDG
AP2C3
d
-log10 p-value
log2 Foldchange
STO
TMM
MKK7
EPF1
e
qPCR
RNA-seq
Relative expression level
Expression level (RPKM)
rob
bid
rob
bid
f
Signal peptide
Predicted functional region


a
Zmstomagen1
Zmstomagen2
Osstomagen1
stomagen
Fstomagen
Zmstomagen3
Osstomagen2
Signal peptide
STOMAGEN
b
Bsta_Brast01G048000
Macu_GSMUA_Achr11T26690
Macu_GSMUA_AchrUn_randomT02450
Osat_LOC_Os08g41360
Phal_Pahal.F00670
Svir_Sevir.6G225100
Sita_SI015601
Sbic_Sobic.007G197500
Zmay_GRMZM2G017321
Bsta_Brast03G284700
Macu_GSMUA_Achr5T25360
outgroup_AT4G14723_CLL2
Atri_v1.0_scaffold00020
Acoe_Aquca_034_00033
Tcac_Thecc1EG002693
Esal_Thhalv10026516
Alyr_497108
F.sonorensis
F.cronquistii
F.robusta
F.kochiana
F.pringlei
F.ramosissima
F.anomala
F.chlorafolia
F.brownii
F.pubescens
F.floridana
F.palmeria
F.australasica
F.trinervia
F.vaginata
F.bidentis
Atha_AT4G12970
Cgra_Cagra.1702s0003
Crub_Carubv10005936
Brap_Brara.H00497
Ptri_Potri.002G249900
Gmax_Glyma.08G345100
Gmax_Glyma.18G152800
Smoe_84711
Macu_GSMUA_Achr8T00990
Osat_LOC_Os01g68598
Zmay_GRMZM2G170975
Sbic_Sobic.003G399800
Phal_Pahal.E00527
Sita_SI003351
Svir_Sevir.5G431400
Bdis_Bradi2g58540
c
STOMAGEN
Signal peptide
Functional area
FSTOMAGEN (FSTO)
Signal peptide
Functional area
AFSTO
Signal peptide
Functional area
d
WT
OXAFSTO
OXFSTO
e
SD (mm-2)
WT 8 15 5 4 13 14 11 12 30 9
f
SD (mm-2)
WT 3 19 18 6 2 20 16 7 40 17 1
g
n.s.
SD (mm-2)
OXAFSTO OXFSTO
h
FSTO
ACT8
wt 8 15 5 4 13 14 11 12 30 9
i
AFSTO
ACT8
wt 3 19 18 6 2 20 16 7 40 17 1
j
WT
OXFSTO
OXAFSTO


a
b
c
SD (mm-2)
1500
1000
500
0
Mock
1μM
10μM
d
SD (mm-2)
500
400
300
200
100
0
Mock
1μM
10μM
e
SI
0.5
0.4
0.3
0.2
0.1
0.0
Mock
1μM
10μM

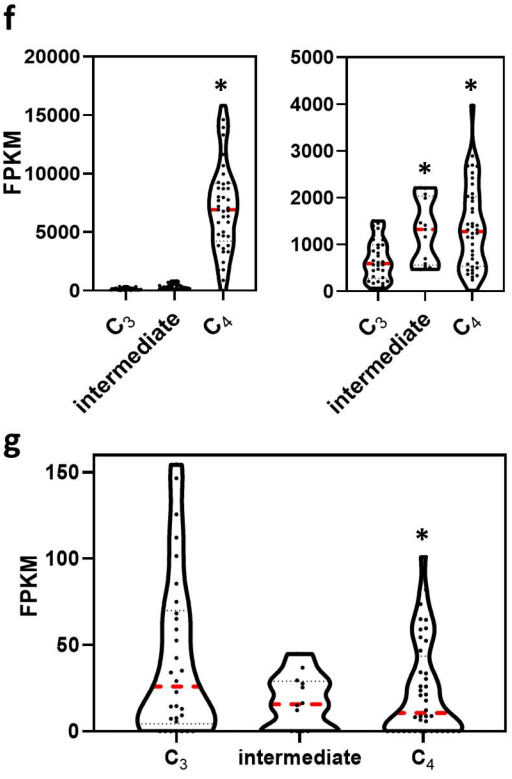

a
Expression level
(RPKM)
80
60
40
20
0
rob
bid
P 1 2 3 4 5 6 7 8 9
b
Relative expression
8.0
6.0
4.0
2.0
0.0
*
*
rob
ram
bid
c
Expression level (RPKM)
50
40
30
20
10
0
Relative expression level
0.3
0.2
0.1
0.0
RNA-seq
qPCR
a
A
a
B
b
C
C3
Intermediates
C4

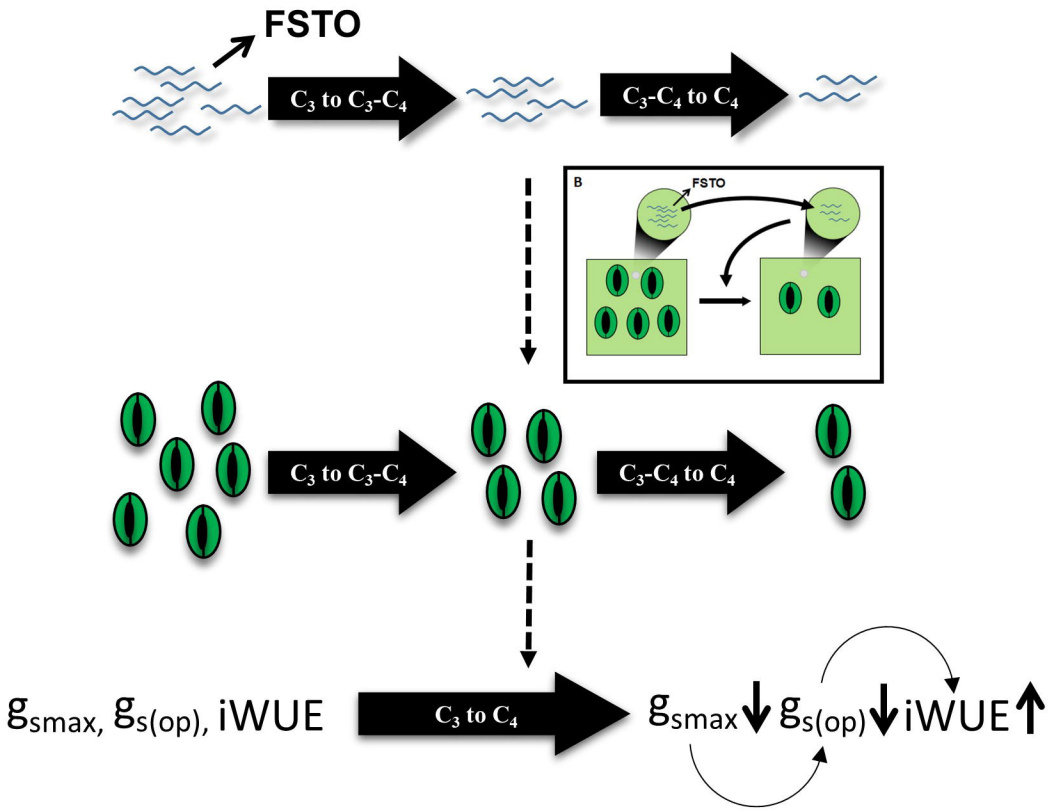




e

Family	Species number
Poaceae	4
Aizoaceae	7
Amaranthacea	14
Chenopodiaceae	6
Molluginaceae	7
Nyctaginaceae	8
Portulacaceae	11
Zygophyllaceae	1
Euphorbiaceae	1
Cleomaceae	3
Boraginaceae	14
Scrophulariaceae	1




FSTO
C3 to C3-C4
C3-C4 to C4
B
FSTO
C3 to C3-C4
C3-C4 to C4
gsmax, gs(op), iWUE
C3 to C4
gsmax↓ gs(op)↓ iWUE↑