# Expression of *SULTR2;2* in the Arabidopsis bundle sheath is mediated by a highly conserved positive regulator

Sandra Kirschner[1*], Helen Woodfield[2*], Katharina Prusko[1], Maria Koczor[1], Udo Gowik[1#], Julian M. Hibberd[2§], Peter Westhoff[1§]

[1]Institute for Plant Molecular and Developmental Biology, Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, D-40225 Düsseldorf, Germany.

[2]Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, United Kingdom.

[#]Present Address: Institute for Biology and Environmental Science (IBU), Plant Evolutionary Genetics, Carl von Ossietzky University Oldenburg, Germany

* Joint first authors

[§] For correspondence

Email addresses:

SK - sandra.kirschner@hhu.de

HW - hkwoodfield@gmail.com

KP - katharina.prusko@hhu.de

MK - maria.koczor@hhu.de

UG - udo.gowik@uni-oldenburg.de

JMH - jmh65@cam.ac.uk

PW - west@hhu.de

**Keywords:** *SULTR2.2*, Bundle sheath, gene expression, $C_3$ photosynthesis, *Arabidopsis thaliana*, *Flaveria bidentis*

**Abstract**

The bundle sheath provides a conduit linking veins and mesophyll cells. In $C_3$ *Arabidopsis thaliana* it also plays important roles in oxidative stress and sulphur metabolism. However, the mechanisms responsible for the patterns of gene expression that underpin these metabolic specialisations are poorly understood. Here we used the *A. thaliana SULTR2;2* gene as a model to better understand mechanisms that restrict expression to the bundle sheath. Deletion analysis indicated that the *SULTR2;2* promoter contains a short region necessary for expression in the bundle sheath. This sequence acts as a positive regulator and is tolerant to multiple consecutive deletions indicating considerable redundancy in the *cis*-elements involved. It is highly conserved in *SULTR2;2* genes of the Brassicaceae and is functional in the distantly related $C_4$ species *Flaveria bidentis* that belongs to the Asteraceae. We conclude that expression of *SULTR2;2* in the bundle sheath is underpinned by a highly redundant sequence that likely represents an ancient and conserved mechanism found in families as diverse as the Asteraceae and Brassicaceae.

44 **Introduction**

45 The evolution of multicellularity is associated with individual cell-types being able to undertake
46 specialised roles within a tissue. In leaves, bundle sheath (BS) cells form a wreath-like structure
47 around the vasculature that appears analogous to the endodermis of roots (Esau, 1965). The role of
48 BS cells is best characterized in $C_4$ species that partition photosynthesis between the BS and
49 mesophyll cells. In most $C_4$ plants, after $HCO_3^-$ is initially fixed into $C_4$ acids by phospho*enol*pyruvate
50 carboxylase in mesophyll cells, these $C_4$ acids then diffuse to the BS where $CO_2$ is released and
51 refixed by Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase (RuBisCO). Decarboxylation of $C_4$
52 acids in the BS generates a high concentration of $CO_2$ around RuBisCO that suppresses the
53 oxygenase activity of the enzyme and in so doing reduces photorespiration (Hatch, 1987). Thus, in
54 $C_4$ species, the BS is specialized to allow efficient fixation of $CO_2$ in the Calvin-Benson-Bassham
55 cycle. In some $C_4$ plants, the BS is also modified in terms of light capture. For example, in maize and
56 sorghum Photosystem II does not fully assemble in the BS (Kubicki *et al.*, 1994) but components of
57 cyclic electron transport are more abundant in the BS compared with mesophyll cells (Takabayashi
58 *et al.*, 2005). In addition to these changes associated with photosynthesis, the $C_4$ BS is also modified
59 to preferentially undertake starch synthesis and degradation, as well as the initial steps of sulphur
60 assimilation (Friso *et al.*, 2004).

61 In $C_3$ plants, the role of the BS is less clearly defined. It is thought to help maintain hydraulic
62 integrity of the xylem (Sade *et al.*, 2014), regulate flux of metabolites in and out of the leaf (Leegood,
63 2008) and act as a starch store (Miyake and Maeda, 1976). The $C_3$ BS is less important for
64 photosynthesis than that of $C_4$ species. However, although only around 15% of all chloroplasts of
65 the $C_3$ leaf are found in bundle sheath cells (Kinsman and Pyke, 1998), reducing photosynthesis in
66 these cells compromises growth and seed production (Janacek *et al.*, 2009). Thus, although less
67 photosynthetic than the $C_4$ BS, this physiological analysis indicates that the BS of $C_3$ plants is also
68 specialized. This notion is consistent with analysis of gene expression in this cell type. For example
69 quantification of transcripts available for translation indicate that the *A. thaliana* BS is likely important
70 in sulphur metabolism, glucosinolate biosynthesis, trehalose metabolism and detoxification of active
71 oxygen species (Aubry *et al.*, 2014). In summary, in both $C_3$ and $C_4$ plants mechanisms must operate
72 to restrict the expression of some genes to BS cells.

73 To date, most studies of the mechanisms responsible for preferential gene expression in the BS
74 have occurred in $C_4$ species (Hibberd and Covshoff, 2010). In the $C_4$ dicotyledon *Flaveria trinervia*
75 the glycine decarboxylase P-subunit (*GLDPA*) gene contains two promoters, one proximal to the
76 coding region, and the other more distal. Activity of the distal promoter is high but not cell-type
77 specific. However, in the presence of the proximal promoter, transcripts derived from the distal
78 promoter are degraded in mesophyll cells by nonsense-mediated RNA decay of incompletely spliced
79 transcripts (Engelmann *et al.*, 2008; Wiludda *et al.*, 2012). Despite the phylogenetic distance
80 between the Brassicaceae and the Asteraceae the *GLDPA* promoter from *F. trinervia* is able to
81 generate BS-specific activity in $C_3$ *A. thaliana* (Engelmann *et al.*, 2008; Wiludda *et al.*, 2012). In
82 *Amaranthus hybridus* 5' and 3' untranslated regions of the *RbcS1* gene act to restrict accumulation
83 of the glucoronidase (GUS) reporter to the BS of $C_4$ *F. bidentis* and appear to function as enhancers
84 of translation (Patel *et al.*, 2006). Lastly, in $C_4$ *Gynandropsis gynandra* preferential expression of
85 *NAD-ME1&2* genes in the BS is associated with coding sequence rather than UTRs or promoter
86 elements (Brown *et al.*, 2011). The motifs underpinning this regulation are a pair of duons that play
87 a dual role in coding for amino acids as well as the spatial patterning of gene expression associated
88 with the $C_4$ leaf. Although these duons are present in $C_3$ *A. thaliana* and many other land plants they
89 do not act to generate cell-specific expression in the ancestral $C_3$ state (Brown *et al.*, 2011; Reyna-
90 Llorens *et al.*, 2016). In summary, current evidence indicates that gene expression in the BS of $C_4$
91 species is controlled by a variety of mechanisms, some of which involve regulatory codes that are
92 derived from the ancestral $C_3$ state.

93 However, our understanding of how gene expression is restricted to the BS in $C_3$ species is poor.
94 A small number of promoters including *SHORT-ROOT* (Dhondt *et al.*, 2010), *SCARECROW*
95 (Wysocka-Diller *et al.*, 2000) and *SULTR2;2* (Takahashi *et al.*, 2000) have been reported to drive
96 BS specific expression in *A. thaliana,* but the molecular nature of *cis*-regulatory elements controlling
97 their expression is unclear. An increased understanding of these processes would not only advance
98 our understanding of mechanisms underpinning cell specific gene expression in multicellular leaves,
99 but also provide insight into whether $C_4$ gene expression is built on pre-existing mechanisms found
100 in $C_3$ species.

To better understand mechanisms associated with gene expression in BS cells of $C_3$ *A. thaliana* we analysed the *SULTR2;2* gene that encodes a low-affinity sulphur transporter (Takahashi *et al.*, 2000). Elements in the promoter sequence that regulate the spatial patterning, and the strength of gene expression were identified. Specifically, preferential expression in the BS is mediated by a repetitive region that is highly conserved within orthologous genes from species of the Brassicaceae. This region acts to enhance expression in the BS rather than repressing expression in mesophyll cells. Furthermore, the *SULTR2;2* promoter from *A. thaliana* generates BS specificity in the $C_4$ species *F. bidentis* that belongs to the Asteraceae. The most parsimonious explanation for this finding is that a common transcription factor is shared by these phylogenetically dispersed species, and that it functions in both the $C_3$ and $C_4$ BS.

112 **Results**

113 Nucleotides -2815 to +123 relative to the predicted *SULTR2;2* translational start site have
114 previously been reported to generate expression in the BS of *A. thaliana* (Takahashi *et al.*, 2000).
115 We confirmed this finding (Figure 1A-C). To test if sequence after the predicted start codon is
116 required for expression in the BS a construct that terminated at nucleotide -1 relative to the predicted
117 translational start was generated (Figure 1D). Staining showed that each construct led to strong
118 accumulation of GUS in the BS of mature rosette leaves (Figure 1B,C,E,F). Staining was evident in
119 vascular tissue as well as the BS but there was no evidence that GUS accumulated in mesophyll or
120 epidermal cells from either construct (Figure 1B,C,E,F). Promoter activity was quantified using the
121 MUG fluorimetric assay. GUS activity driven by the construct that contained 603 fewer base pairs at
122 the 5' end but an additional 123 nucleotides of coding sequence was about tenfold higher (Figure
123 1G). This difference could either be due to a negative regulator located between -3418 and -2185
124 nucleotides, or a positive regulator in sequence downstream of the predicted translational start site.
125 A translational fusion between the yellow fluorescent protein (YFP) and the nuclear localized Histone
126 2B protein under control of the *SULTR2;2* promoter (Figure 1H) labelled nuclei of BS cells and
127 vascular tissue (Figure 1I) and indicated that the presence of GUS in the BS was due to gene
128 expression and not diffusion of the dye outwards from to vascular tissue. Consistent with previous
129 reports (Chytilova *et al.*, 1999) it was noticeable that vascular nuclei were elongated and rod-like
130 compared with the more spherical ones in the BS (Figure 1I). We conclude that elements between
131 nucleotides -2815 and the translational start site of *SULTR2;2* are sufficient to drive gene expression
132 in vascular and BS cells of *A. thaliana*.

133

134 **A short region that is necessary and sufficient to activate gene expression in the BS**

135 To better understand the sequences responsible for generating expression in BS cells, a 5'
136 deletion series was generated (Figure 2A). Each of the deleted regions are hereafter referred to as
137 regions 1 to 5. Removal of region 1 had no significant effect on either activity or spatial accumulation
138 of GUS (Figure 2A&C) indicating that no essential *cis*-regulatory elements are located within this
139 section. Deletion of region 2 resulted in total loss of GUS activity and staining (Figure 2A&D), and
140 removal of regions 3 and 4 had no further effect (Figure 2A, E&F). These data therefore indicate that

141 nucleotides in region 1 do not impact on promoter activity, but that region 2 is necessary for BS
142 expression in mature leaves. A separate deletion series involving slightly smaller regions supported
143 this conclusion, with expression in the BS being lost once nucleotides between -1813 and -1335
144 were deleted (Supplementary Figure 1).

145 The lack GUS accumulation in the BS and loss of promoter activity once region 2 is removed
146 could be because this region contains *cis*-elements that generate expression specifically in BS cells
147 or because it drives ubiquitous expression but regions 3 to 5 contain elements that restrict activity to
148 the BS. To investigate this possibility, 5' rapid amplification of cDNA ends was first used to define
149 the transcriptional start site of *SULTR2;2* (Supplemental Figure 2). No single strong transcription
150 start site was detected but rather multiple transcripts were initiated from position -125 onwards
151 (Supplementary Figure 2). Nucleotides spanning -349 to -1 were therefore considered likely to be
152 sufficient for transcriptional initiation and are hereafter referred to as the core promoter. Fusion of
153 region 2 to this core promoter led to MUG conversion that was comparable to that from the full-length
154 promoter (Figure 3A and Figure 2A) and was also sufficient to direct accumulation of GUS to veins
155 and the BS (Figure 3B). To exclude the possibility that the core promoter includes *cis*-elements
156 necessary for BS-specific expression, region 2 was also fused to the minimal CaMV35S promoter
157 which does not drive significant expression. Although this led to fourfold lower GUS activity than the
158 full-length promoter (Figure 3A) GUS accumulation was still restricted to veins and the BS (Figure
159 3C). We conclude that although nucleotides -349 to -1 are not able to generate expression in the BS
160 (Figure 2F), they likely represent the core promoter of *SULTR2;2*. In contrast, nucleotides -2053 to
161 -1312 are sufficient to restrict expression to the BS and vascular tissue of *A. thaliana*.

162

163 ***AtSULTR2;2* contains multiple redundant sequences mediating BS expression**

164 Having established that nucleotides -2053 to -1312 relative to the predicted translational start
165 site are both necessary and sufficient for BS expression, an unbiased approach to further dissect
166 this region was adopted. Ten consecutive deletions were made to this region and each fused to the
167 core promoter of *SULTR2;2* (Figure 4A). Strikingly, none of these deletions resulted in total loss of
168 GUS activity (Figure 4A), nor was GUS staining lost from the BS (Figure 4B-K) indicating that despite
169 an absence of repeated *cis*-elements in this region significant functional redundancy in regulatory

170 elements must mediate BS-specific expression. However, it was notable that GUS activity declined
171 to varying degrees compared with the intact region (Figure 4A), suggesting that either these
172 redundant *cis*-elements act additively, or that this region contains quantitative elements regulating
173 gene expression.

174 To better understand the extent to which *cis*-elements in this section of the promoter act
175 redundantly, larger deletions were made from position -2053 (Figure 5A). This generated five
176 regions, hereafter referred to as sub-regions 2.1 to 2.5. Deletion of sub-region 2.1 resulted in a strong
177 reduction of GUS activity (Figure 5A) but BS-specific accumulation of GUS was maintained (Figure
178 5C). This finding implies that sub-region 2.1 contains a quantitative enhancer element. Deleting sub-
179 region 2.2 had no clear additional impact on either BS-specificity or activity (Figure 5A&D). However,
180 the subsequent deletion of sub-regions 2.3 and 2.4 caused loss of GUS activity and also loss of
181 GUS staining in BS and vascular tissue (Figure 5A,E&F). These data indicate that *cis*-regulatory
182 elements mediating BS expression are situated in sub-region 2.3, or that quantitative elements in
183 this region mask qualitative elements in distal sub-regions. To address these options, 3' deletions of
184 region 2 were also created (Figure 5G). As sub-regions 2.1 and 2.2 had little impact on BS
185 expression, the last three sub-regions were fused to the minimal CaMV35S promoter. Expression
186 from each of these three constructs was low (Figure 5G) however BS expression could be observed
187 from the construct containing sub-regions 2.3, 2.4 and 2.5 (Figure 5H). Removal of sub-region 2.5
188 resulted in loss of GUS in rosette leaves although cotyledons still showed patchy staining restricted
189 to the BS and vascular tissue (Figure 5I). Once sub-region 2.4 was removed GUS activity was further
190 reduced and GUS staining was no longer detectable even in seedlings (Figure 5G&J). We conclude
191 that sub-regions 2.3 and 2.4, which contain a total of 350 base pairs, contain *cis*-regulatory elements
192 necessary for BS expression of *AtSULTR2;2*.

193 The *cis*-regulatory elements necessary for BS-specific expression of *AtSULTR2;2* appear to be
194 located in a 350 nucleotide region of the promoter. As finer-scale deletions had failed to identify the
195 exact *cis*-elements responsible for this phenotype (Figure 4) a phylogenetic approach was
196 undertaken. Orthologues of *AtSULTR2;2* were identified from seven species of the Brassicaceae.
197 Alignments of sequences 5 kb upstream of each orthologue indicated that with the exception of *A.*
198 *lyrata* that contains a 446 nucleotide insertion, region 2 is highly conserved (Figure 6). However, no

199 short sequences or motifs within this sequence that may restrict expression to the BS could be
200 identified (Figure 6). Although the results of this alignment therefore do not identify a specific *cis*-
201 element that could be bound by a transcription factor responsible for generating BS expression, they
202 do support the functional analysis and implicate the whole of region 2 as a critical component of the
203 *SULTR2;2* promoter for BS expression.
204
205 **The *AtSULTR2;2* promoter is capable of driving BS specific expression in $C_4$ *Flaveria bidentis***
206 As the *GLDPA* promoter from the $C_4$ species *F. trinervia* is able to confer BS specific expression
207 in $C_3$ *A. thaliana* (Engelmann *et al.*, 2008) we next tested whether the *A. thaliana SULTR2;2* promoter
208 would lead to BS expression in $C_4$ *F. bidentis*. GUS activity in transgenic *F. bidentis* plants was about
209 fourfold higher than that in *A. thaliana* (Figure 7A&B). However, histochemical analysis of mature
210 leaves revealed a very similar expression pattern to that in *A. thaliana* with strong GUS accumulation
211 in BS and vascular tissue but not in mesophyll cells (Figure 7C*)*. This indicates that transcription
212 factors from *F. bidentis* recognise BS *cis*-regulatory elements from the Brassicaceae. The most
213 parsimonious explanation for this finding is that these sequences represent part of an ancient and
214 conserved mechanism that restricts gene expression to BS cells of dicotyledenous leaves.
215

## Discussion

### *SULTR2;2* contains a short region that activates expression in the bundle sheath

*SULTR2;2* encodes a low affinity transporter that facilitates movement of sulphate from the vascular bundle to palisade cells in the leaf (Takahashi *et al.*, 2000). Consistent with this function, analysis of both GUS and the histone2B::YFP fusion indicated that the *SULTR2;2* promoter directs expression to veins as well as to BS cells. It was notable that none of the various deletions we made to this promoter led to expression being restricted to either veins or bundle sheath cells. The *GLDPA* promoter from *F. trinervia* also drives expression in both veins and BS cells of *A. thaliana* (Engelmann *et al.*, 2008). These findings therefore imply that the *SULTR2;2* and *GLDP* genes may be controlled by gene regulatory networks that are shared by these cell-types. It is possible that these cell-types share some gene regulatory networks because they are derived from the same lineage (Dengler and Nelson, 1999; Soros and Dengler, 2001).

Within the *SULTR2;2* promoter, one specific region that impacted on gene expression in the BS was identified. This sequence, consisting of 350 nucleotides, is both necessary and sufficient for restricting expression to this cell-type. Small consecutive deletions within this region failed to abolish this spatial patterning implying that multiple contiguous elements act redundantly to generate strong and stable expression in the BS. The only detectable impact of deleting any part of this region was for strength of expression to be reduced. We therefore propose that either multiple independent BS modules contained within this region act additively, or that distinct quantitative elements are co-located, and at least partially overlapping with, *cis*-elements that determine this cell-specificity. Redundancy of this sort has previously been reported for the promoter of *Phenylalanine Ammonia Lyase*2 that drives xylem-specific expression in tobacco (Leyva *et al.*, 1992; Hatton *et al.*, 1995), *DORNRÖSCHEN-LIKE* of *A. thaliana* that contains three functionally redundant enhancers (Comelli *et al.*, 2016), and *EVEN-SKIPPED (EVE)* from *Drosophila melanogaster* where a minimal enhancer is sufficient to direct expression of *EVE* to the second stripe, but surrounding binding sites increase the robustness of this patterning during genetic and environmental perturbations (Ludwig *et al.*, 2011). Thus, although the exact role of redundancy in the regulation of *SULTR2;2* in the BS is unclear, it may also increase robustness in the control of gene expression during environmental perturbations, and/or increase patterning precision (Barolo, 2012; Payne and Wagner, 2015).

245 Compared with other examples of elements that restrict gene expression to BS cells of $C_3$
246 species, this single block of sequence from *SULTR2;2* that acts as a positive regulator of
247 transcription appears to operate via relatively simple mechanisms. For example, the *F. trinervia*
248 *GLDPA* generates BS-specific expression (Engelmann *et al.*, 2008) because of a complex interplay
249 between transcriptional and post-transcriptional processes. These are mediated by distal and
250 proximal sequences relative to the translational start site, leading to repression of *GLDP* expression
251 in mesophyll cells (Wiludda *et al.*, 2012). We therefore propose that the positive regulator located
252 upstream of *SULTR2;2* could be used as a synthetic module to manipulate or engineer processes
253 in BS of *A. thaliana*.
254

255 **The bundle sheath element of *SULTR2;2* is conserved in the Brassicaceae and functional in**
256 **the Asteraceae**

257 Alignment of *SULTR2;2* promoters from multiple species of the Brassicaceae did not reveal an
258 individual shared *cis*-element but rather highlighted sequence that was conserved across the whole
259 of region 2. This sequence conservation argues for relatively strong purifying selection compared
260 with the rest of the *SULTR2;2* promoter and also implies that this region may function as part of a
261 widely conserved positive regulator of gene expression in the BS of the Brassicaceae. Consistent
262 with this proposal and indicating that these regulatory elements may be even more ancient, when
263 the *SULTR2;2* promoter was placed into the phylogenetically distant $C_4$ species *F. bidentis* it was
264 also recognized by *trans*-factors that restricted gene expression to the $C_4$ BS. This is analogous to
265 the behavior of the *GLDPA* promoter from $C_4$ *F. trinervia* which is able to restrict expression to the
266 BS of *A. thaliana* (Engelmann *et al.*, 2008; Wiludda *et al.*, 2012). Currently, the crown ages of the
267 rosids and asterids are estimated to be 108-117 and 107-117 million years ago (Sanderson *et al.*,
268 2004; Wikström *et al.*, 2001) indicating that these clades diverged in the Early Cretaceous. Whilst
269 for both *SULTR2;2* and *GLDP* it is possible that different mechanisms lead to BS specific expression
270 in species of the Brassicaceae and Asteraceae it seems more likely that expression of each gene is
271 determined by ancient and highly conserved *cis*-regulatory codes that that have been maintained
272 since these clades diverged from their last common ancestor. Although the plant vasculature is
273 thought to have started to evolve from 450 to 430 million years ago (Furuta *et al.*, 2014; Ye, 2002)

to our knowledge there are no clear estimates of when the BS originated. It would be intriguing if regulatory networks operating in both the veins and BS cells are uncovered that can be associated with the evolution of the vasculature in early diverging lineages of land plants.

277 **Materials and methods**

278 **Cloning of promoter-reporter gene constructs and 5' Rapid Amplification of cDNA ends**

279 All DNA fragments created by PCR were confirmed by DNA sequencing. Generation of full-length
280 promoter construct via PCR was constructed with *Arabidopsis thaliana* Columbia-0 (Col-0) genomic
281 DNA. Subsequent constructs were generated using this as a template. Restriction sites were added
282 to the respective fragments by PCR and fragments were inserted into pBI121 or a partially modified
283 pBI121. Region 2 with internal deletions were synthesised by GenScript and swapped with the full-
284 length region 2 of 2::5.2::*GUS* to generate the internal deletion constructs. Total RNA from leaves of
285 wild type *A. thaliana* Col-0 plants was extracted, DNase I treated and purified with the RNeasy®
286 Plant Mini Kit (Qiagen, Hilden, Germany). cDNA was generated from 1 µg RNA and then 5' RACE-
287 PCR performed using the Advantage® 2 DNA Polymerase Mix (Clontech Laboratories, Mountain
288 View, USA) or Phusion® HF DNA Polymerase (Thermo Fisher Scientific, Waltham, USA). Two
289 nested 3' oligonucleotides, AtSultr2;2-11 and AtSultr2;2-13, both binding in the cDNA of
290 *AtSULTR2;2*, were used. PCR products were cloned and confirmed via colony PCR. Correct clones
291 were subjected for plasmid preparation and sequencing.

292

293 **Transformation and plant growth**

294 *A. thaliana* ecotype Col-0 was transformed using floral dipping (Clough and Bent, 1998;
295 Logemann *et al.*, 2006) using *Agrobacterium tumefaciens* strain AGL1. *F. bidentis* was transformed
296 as described previously (Chitty *et al.*, 1999). Successful transformations of *A. thaliana* or *F. bidentis*
297 were tested by PCR. Before transplanting to soil, positive transformants of *A. thaliana* were selected
298 on kanamycin. Seeds were sterilized by washing twice for 5 min with 20 % (v/v) DanKlorix (Colgate-
299 Palmolive, New York City, USA) and 0.02 % (v.v) Triton-X 100 (Fluka Analyticals, Buchs,
300 Switzerland) and four times with sterile water. Stratification of seeds was performed at 4 °C for 48
301 hours prior to spreading them on ½ MS pH 5.7 containing 10 g $l^{-1}$ (w/v) sucrose (Sigma-Aldrich, St.
302 Louis, USA), 0.5 g $l^{-1}$ MES (Biomol, Hamburg, Germany), 2.13 g/l (w/v) Murashige-Skoog basal salts
303 (Duchefa Biochemie, Haarlem, Netherlands), 0.75 % agar (SERVA Electrophoresis, Heidelberg,
304 Germany), 50 µg $ml^{-1}$ kanamycin (Sigma-Aldrich, St. Louis, USA) and 100 mg $l^{-1}$ Cefotaxim

305 (Fresenius Kabi Deutschland, Bad Homburg, Germany). Plants were transferred to 14 h light/10 h
306 dark and temperatures of 23 °C day and 20 °C night and a light intensity of 90 µmol $m^{-2}$ $s^{-1}$.
307
308 **Visual and quantitative analysis of reporter genes**
309 To take account of effects caused by transgene insertion into different genomic locations at
310 between nine and forty-one independent $T_0$ plants were analysed for each construct. Histochemical
311 analysis was performed as described previously (Engelmann *et al.*, 2008). For *A. thaliana* either
312 three to four-week old rosette leaves or ten to fourteen-day old seedlings and for *F. bidentis* the sixth
313 leaves of 40-50 cm tall plants were used. Transverse sections were prepared manually using a razor
314 blade. Stained leaves were imaged using light microscopy. Quantification of GUS activity was
315 performed via fluorimetric assay (Jefferson *et al.*, 1987) using two to four leaves of three- to four-
316 week old $T_0$ *A. thaliana* plants or the fifth leave of 40-50 cm tall $T_0$ *F. bidentis* plants, respectively.
317 The Mann-Whitney U test was used to determine statistical differences between datasets. Imaging
318 of H2B::YFP was performed on a Zeiss LSM 780 confocal laser-scanning microscope, and YFP
319 fluorescence excited at 514 nm and emission detected between 517 and 569 nm.
320
321
322 **Acknowledgments:** Work performed in PWs laboratory (SK, KP, MK, UG) was supported by the
323 Deutsche Forschungsgemeinschaft through the Cluster of Excellence in Plant Sciences
324 Duesseldorf-Cologne (EXC 1028). HW supported by a BBSRC PhD studentship. We thank the
325 Center for Advanced imaging (HHU Düsseldorf) for provision and technical assistance of the Zeiss
326 LSM 780 laser-scanning microscope.
327

328 **References**

329 **Aubry, S., Smith-Unna, R.D., Boursnell, C.M., Kopriva, S. and Hibberd, J.M.** (2014) Transcript
330 residency on ribosomes reveals a key role for the Arabidopsis thaliana bundle sheath in sulfur
331 and glucosinolate metabolism. *Plant J.*, **78**, 659–673.

332 **Barolo, S.** (2012) Shadow enhancers: Frequently asked questions about distributed cis-regulatory
333 information and enhancer redundancy. *BioEssays*, **34**, 135–141.

334 **Brown, N.J., Newell, C.A., Stanley, S., Chen, J.E., Perrin, A.J., Kajala, K. and Hibberd, J.M.**
335 (2011) Independent and Parallel Recruitment of Preexisting Mechanisms Underlying $C_4$
336 Photosynthesis. *Science* **331**, 1436–1439.

337 **Chitty, J.A., Furbank, R.T., Marshall, J.S., Chen, Z. and WC, T.** (1999) Genetic transformation of
338 the $C_4$ plant, *Flaveria bidentis*. *Plant J.*, **6**, 949–956.

339 **Chytilova, E.V.A., Macas, J. and Galbraith, D.W.** (1999) Green Fluorescent Protein Targeted to
340 the Nucleus, a Transgenic Phenotype Useful for Studies in Plant Biology. *Ann. Bot.*, **83**, 645–
341 654.

342 **Clough, S.J. and Bent, A.F.** (1998) Floral dip: a simplified method for Agrobacterium-mediated
343 transformation of Arabidopsis thaliana. *Plant J.*, **16**, 735–743.

344 **Comelli, P., Glowa, D., Chandler, J.W. and Werr, W.** (2016) Founder-cell-specific transcription of
345 the DORNRÖSCHEN-LIKE promoter and integration of the auxin response. *J. Exp. Bot.*, **67**,
346 143–155.

347 **Dengler, N. and Nelson, T.** (1999) Leaf structure and development In $C_4$ plants. In R. F. Sage and
348 R. K. Monson, eds. *$C_4$ Plant Biology*. San Diego: Academic Press.

349 **Dhondt, S., Coppens, F., Winter, F. De, Swarup, K., Merks, R.M., Inze, D., Bennett, M.J. and**
350 **Beemster, G.T.** (2010) SHORT-ROOT and SCARECROW regulate leaf growth in
351 Arabidopsis by stimulating S-phase progression of the cell cycle. *Plant Physiol*, **154**, 1183–
352 1195.

353 **Engelmann, S., Wiludda, C., Burscheidt, J., et al.** (2008) The gene for the P-subunit of glycine
354 decarboxylase from the $C_4$ species *Flaveria trinervia*: analysis of transcriptional control in
355 transgenic *Flaveria bidentis* ($C_4$) and *Arabidopsis* ($C_3$). *Plant Physiol.*, **146**, 1773–1785.

356 **Esau, K.** (1965) *Plant Anatomy* 2nd Ed., New York: John Wiley and Sons.

357 **Friso, G., Giacomelli, L., Ytterberg, A.J., Peltier, J.B., Rudella, A., Sun, Q. and Wijk, K.J.**
358 (2004) In-depth analysis of the thylakoid membrane proteome of Arabidopsis thaliana
359 chloroplasts: new proteins, new functions, and a plastid proteome database. *Plant Cell*, **16**,
360 478–499.
361 **Furuta, K.M., Hellmann, E. and Helariutta, Y.** (2014) Molecular Control of Cell Specification and
362 Cell Differentiation During Procambial Development. *Annu. Rev. Plant Biol.*, **65**, 607–638.
363 **Hatch, M.D.** (1987) $C_4$ photosynthesis: a unique blend of modified biochemistry, anatomy and
364 ultrastructure. *Biochim. Biophys. Acta*, **895**, 81–106.
365 **Hatton, D., Sablowski, R., Yung, M.-H., Smith, C., Schuch, W. and Bevan, M.** (1995) Two
366 classes of *cis*-sequences contribute to tissue-specific expression of a *PAL2* promoter in
367 transgenic tobacco. *Plant J.*, **7**, 859–876.
368 **Hibberd, J.M. and Covshoff, S.** (2010) The Regulation of Gene Expression Required for $C_4$
369 Photosynthesis. *Annu. Rev. Plant Biol.*, **61**, 181–207.
370 **Higo, K., Ugawa, Y., Iwamoto, M. and Korenaga, T.** (1999) Plant cis-acting regulatory DNA
371 elements (PLACE) database: 1999. *Nucleic Acids Res.*, **27**, 297–300.
372 **Janacek, S.H., Trenkamp, S., Palmer, B., et al.** (2009) Photosynthesis in cells around veins of
373 the $C_3$ plant *Arabidopsis thaliana* is important for both the shikimate pathway and leaf
374 senescence as well as contributing to plant fitness. *Plant J.*, **59**, 329–343.
375 **Jefferson, R.A., Kavanagh, T.A. and Bevan, M.W.** (1987) GUS fusions: ß-Glucuronidase as a
376 sensitive and versatile gene fusion marker in higher plants. *EMBO J.*, **6**, 3901–3907.
377 **Kinsman, E.A. and Pyke, K.A.** (1998) Bundle sheath cells and cell-specific plastid development in
378 Arabidopsis leaves. *Development*, **125**, 1815–1822.
379 **Kubicki, A., Steinmuller, K. and Westhoff, P.** (1994) Differential transcription of plastome-
380 encoded genes in the mesophyll and bundle-sheath chloroplasts of the monocotyledonous
381 NADP-malic enzyme-type $C_4$ plants maize and sorghum. *Plant Mol. Biol.*, **25**, 669–679.
382 **Leegood, R.C.** (2008) Roles of the bundle sheath cells in leaves of $C_3$ plants. *J. Exp. Bot.*, **59**,
383 1663–1673.
384 **Leyva, A., Liang, X., Pintor-Toro, J.A., Dixon, R.A. and Lamb, C.J.** (1992) *cis*-element
385 combinations determine phenylalanine ammonia-lyase gene tissue-specific expression

386 patterns. *Plant Cell*, **4**, 263–271.

387 **Logemann, E., Birkenbihl, R.P., Ülker, B. and Somssich, I.E.** (2006) An improved method for
388 preparing Agrobacterium cells that simplifies the Arabidopsis transformation protocol. *Plant*
389 *Methods*, **2**, 16.

390 **Ludwig, M.Z., Manu, Kittler, R., White, K.P. and Kreitman, M.** (2011) Consequences of
391 Eukaryotic Enhancer Architecture for Gene Expression Dynamics, Development, and Fitness.
392 *PLOS Genet.*, **7**, e1002364.

393 **Miyake, H. and Maeda, E.** (1976) Development of bundle sheath chloroplasts in rice seedlings.
394 *Can. J. Bot.*, **54**, 556–565.

395 **Patel, M., Siegel, A.J. and Berry, J.O.** (2006) Untranslated regions of *FbRbcS1* mRNA mediate
396 bundle sheath cell-specific gene expression in leaves of a $C_4$ plant. *J. Biol. Chem.*, **281**,
397 25485–25491.

398 **Payne, J.L. and Wagner, A.** (2015) Mechanisms of mutational robustness in transcriptional
399 regulation. *Front. Genet.*, **6**, 322.

400 **Reyna-Llorens, I., Burgess, S.J., Williams, B.P., Stanley, S., Boursnell, C. and Hibberd, J.M.**
401 (2016) Ancient coding sequences underpin the spatial patterning of gene expression in $C_4$
402 leaves. *bioRxiv*, doi: https://doi.org/10.1101/085795.

403 **Sade, N., Shatil-Cohen, A., Attia, Z., et al.** (2014) The role of plasma membrane aquaporins in
404 regulating the bundle sheath-mesophyll continuum and leaf hydraulics. *Plant Physiol.,* **166,**
405 1609-20. doi:10.1104/pp.114.248633.

406 **Sanderson, M.J., Thorne, J.L., Wikström, N. and Bremer, K.** (2004) Molecular evidence on
407 plant divergence times. *Am. J. Bot.*, **91**, 1656–1665.

408 **Soros, C.L. and Dengler, N.G.** (2001) Ontogenetic derivation and cell differentiation in
409 photosynthetic tissues of $C_3$ and $C_4$ Cyperaceae. *Am J Bot*, **88**, 992–1005.

410 **Takabayashi, A., Kishine, M., Asada, K., Endo, T. and Sato, F.** (2005) Differential use of two
411 cyclic electron flows around photosystem I for driving $CO_2$ concentration mechanism in $C_4$
412 photosynthesis. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 16898–16903.

413 **Takahashi, H., Watanabe-Takahashi, A., Smith, F.W., Blake-Kalff, M., Hawkesford, M.J. and**
414 **Saito, K.** (2000) The roles of three functional sulphate transporters involved in uptake and

415 translocation of sulphate in Arabidopsis thaliana. *Plant J.*, **23**, 171–182.

416 **Wikström, N., Savolainen, V. and Chase, M.W.** (2001) Evolution of the angiosperms: calibrating
417 the family tree. *Proc. R. Soc. London. Ser. B Biol. Sci.*, **268**, 2211 LP-2220.

418 **Wiludda, C., Schulze, S., Gowik, U., Engelmann, S., Koczor, M., Streubel, M., Bauwe, H. and**
419 **Westhoff, P.** (2012) Regulation of the photorespiratory *GLDPA* gene in $C_4$ *Flaveria*: an
420 intricate interplay of transcriptional and posttranscriptional processes. *Plant Cell*, **24**, 137–151.

421 **Wysocka-Diller, J.W., Helariutta, Y., Fukaki, H., Malamy, J.E. and Benfey, P.N.** (2000)
422 Molecular analysis of SCARECROW function reveals a radial patterning mechanism common
423 to root and shoot. *Development*, **127**, 595–603.

424 **Ye, Z.-H.** (2002) VASCULAR TISSUE DIFFERENTIATION AND PATTERN FORMATION IN
425 PLANTS. *Annu. Rev. Plant Biol.*, **53**, 183–202.

426 **Figure legends**

427 **Figure 1: The *SULTR2;2* promoter drives expression in veins and bundle sheath cells of**
428 ***Arabidopsis thaliana*. (A-C)** Nucleotides -2185 to +123 of *SULTR2;2* are sufficient to generate
429 preferential accumulation of the GUS reporter in bundle sheath cells. **(D-F)** Nucleotides -2185 to
430 +123 of *SULTR2;2* are sufficient to generate preferential accumulation of the GUS reporter in bundle
431 sheath cells. **(G)** Quantitative analysis of expression from each construct based on the GUS activity
432 assay. **(H)**. Schematic of the *A. thaliana SULTR2;2* full-length promoter fused to *H2B::YFP* that
433 targets YFP to the nucleus. **(I)** H2B::YFP fusion marks nuclei from the larger bundle sheath exhibits
434 (white arrowhead) as well as the smaller more elongated nuclei of the vasculature (red arrowhead).
435 Data from GUS activity assays include the median (M) indicated by red lines and the number (n) of
436 independent lines. Histological GUS assays were allowed to proceed for 23 h (B), 16 h (C) and 5 h
437 (E) 6 h (F). Scale bars represent 500 (C) or 50 µm (D,E,F).

438

439 **Figure 2: A 715 nucleotide region in the *SULTR2;2* promoter that is necessary for bundle**
440 **sheath expression. (A)** Schematic of the 5' deletion series (left) with GUS activities (right). **(B-F)**
441 GUS accumulation from each deletion. Deletion of region 2 abolished accumulation of GUS. Data
442 from GUS activity assays include the median (M) indicated by red lines and the number (n) of
443 independent lines. Histological GUS assays were conducted on leaves for 23 h (B), 4 h (C) and 6
444 days. Histological GUS assays were allowed to proceed for 23 h (B), 4 h (C) and 6 days (E&F). Scale
445 bars represent 50 µm.

446

447 **Figure 3: The *SULTR2;2* promoter contains a region that is sufficient to activate expression**
448 **in bundle sheath cells. (A)** Schematic of constructs containing nucleotides -2053 to -1312
449 combined with either the core promoter of *SULTR2;2* or the minimal CaMV35S promoter (left) and
450 quantitative analysis of expression from each construct based on the GUS activity assay (right).
451 Orange arrowheads within the core promoter indicate transcription start sites obtained by 5'-RACE.
452 **(B-C)** Both constructs are sufficient to generate GUS accumulation in the bundle sheath. Data from
453 GUS activity assays include the median (M) indicated by red lines and the number (n) of independent
454 lines. Histological GUS assays were allowed to proceed for 5 h (B) and 22 h (C). Scale bars represent

455 50 µm.

456

457 **Figure 4: *SULTR2;2* contains a region with multiple redundant regions capable of directing**
458 **bundle sheath expression. (A)** Schematic of internal deletion series constructs (left) and
459 quantitative analysis of expression from each construct based on the GUS activity assay (right).
460 Except for the last deletion representing 66 base pairs, each construct lacks consecutive 75 base
461 pair sequences. These internal deletions modify the GUS activity **(A)**, but none abolish accumulation
462 of GUS in the bundle sheath **(B-K)**. Data from GUS activity assays include the median (M) indicated
463 by red lines and the number (n) of independent lines. Histological GUS assays were allowed to
464 proceed for 23 h (B), 16 h (C), 48 h (D), 23 h (E), 48 h (F), 8 h (G), 3 h (H), 7 h (I), 3 h (J) and 19 h
465 (K). Scale bars represent 50 µm.

466

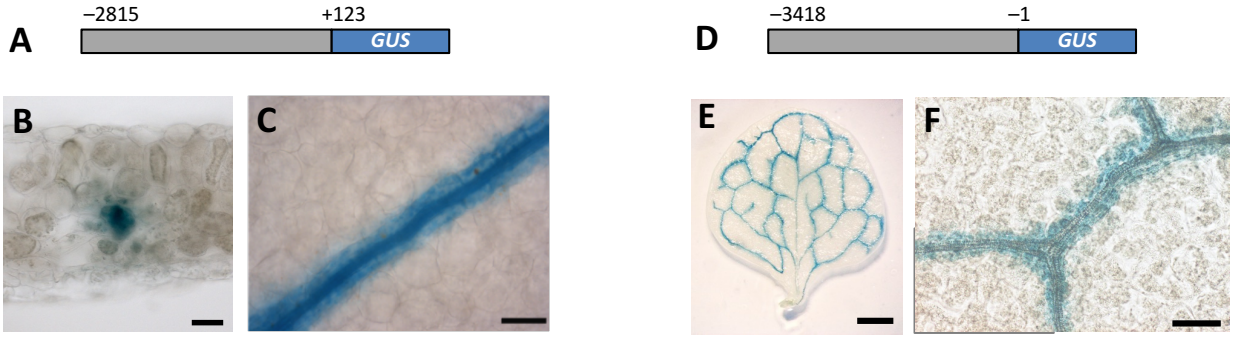
467 **Figure 5: A short region of the *SULTR2;2* promoter that is necessary for expression in the**
468 **bundle sheath. (A**) Schematic of deletions made to region 2 of the *SULTR2;2* promoter (left) and
469 quantitative analysis of expression from each construct based on the GUS activity assay (right).
470 Activity was no longer detectable when subregions 4 and 5 were removed. **(B-F)** Histochemical
471 staining of leaves indicated that subregion 3 is required for expression in the bundle sheath. **(G)**
472 Schematic of 3' deletion constructs placed upstream of the minimal CaMV35S promoter (left) and
473 quantitative analysis of expression from each construct based on the GUS activity assay (right). **(H-**
474 **I)** Nucleotides from -1845 to -1495 are sufficient to drive expression in the bundle sheath of leaves
475 (H) and cotyledons (I) respectively. **(J)** Deletion of nucleotides -1692 to 1312 abolishes accumulation
476 of GUS in the BS. Data from GUS activity assays include the median (M) indicated by red lines and
477 the number (n) of independent lines. Histological GUS assays were allowed to proceed for 4 h (B),
478 47 h (C), 4 h (D) 6 d (E, F), 5 d (H), 2 d (I) and 29 h (J). Scale bars represent 50 µm.

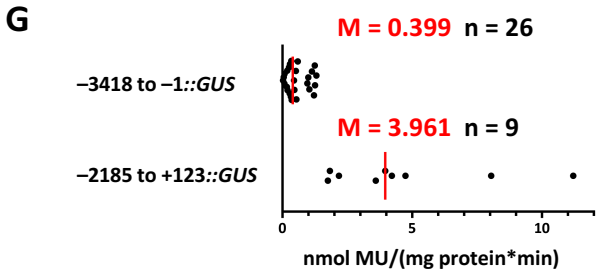
479

480 **Figure 6: The region necessary for BS expression of *SULTR2;2* is highly conserved in the**
481 **Brassicaceae**. The promoter of *AtSULTRT2;2* was aligned against sequences ~5 kb upstream of
482 genes from seven additional species of the *Brassicaceae*. With the exception of *Arabidopsis lyrata*,
483 which contains a 446 bp insertion, region 2 is highly conserved. Black boxes indicate strong similarity

484 of sequences, grey boxes sequences not matching the consensus sequence and black lines gaps.
485 The level of similarity is also indicated on top of the alignment. High peaks in green mark strong
486 similarity, low red peaks poor similarity.
487
488 **Figure 7: The *A. thaliana SULTR2;2* promoter generates strong bundle sheath expression in**
489 **the $C_3$ species *Flaveria bidentis*. (A)** Schematic of the sequence placed into *Flaveria bidentis*. **(B)**
490 Quantitative analysis of expression from each construct based on the GUS activity assay. To
491 facilitate comparison GUS data from the same construct in $C_3$ *A. thaliana* are included. **(C)**
492 Representative image of transverse section of *Flaveria bidentis* after histochemical staining for GUS.
493 Data from GUS assays include the median (M) indicated by red lines and the number (n) of
494 independent lines. Histological GUS assays were allowed to proceed for 4 h. Scale bar represents
495 50 µm.

496 **Supplemental Figure 1: Additional deletion series generated for *SULTRT2;2*. (A)** Schematics
497 illustrating each deletion (left) and quantitative analysis of expression from each construct based on
498 the GUS activity assay (right). The series was designed to remove around 400 bp each time and to
499 avoid *cis*-regulatory elements predicted by the software PLACE (Higo *et al.*, 1999). Deletion of region
500 II resulted in a decline in GUSactivity and deletion of region III lead to a total loss of expression. **(B-**
501 **H)** Representative images after histological staining for GUS indicating that deleting region III led to
502 loss of GUS in the bundle sheath. Data from GUS activity assays include the median (M) indicated
503 by red lines and the number (n) of independent lines. Histological GUS assays were allowed to
504 proceed for 6 h. Scale bars each 50 µm.
505
506 **Supplemental Figure 2:** Alignment of 5' ends of cDNAs obtained via 5' Rapid amplification of cDNA
507 ends. No distinct transcription start site was found. The top line shows the template used for
508 alignment. Predicted translational start sites are marked in blue (annotated in TAIR Accession
509 1009028759) and red (from Takahashi *et al.* (2000)). A TATA box motif found by the PLACE
510 database (Higo *et al.*, 1999) is underlined. The black arrowhead marks the TAIR annotated
511 transcription start site (TAIR Accession 1009028759).


A
−2815
+123
GUS
B
C
D
−3418
−1
GUS
E
F


G
M = 0.399 n = 26
−3418 to −1::GUS
M = 3.961 n = 9
−2185 to +123::GUS
0
5
10
nmol MU/(mg protein*min)

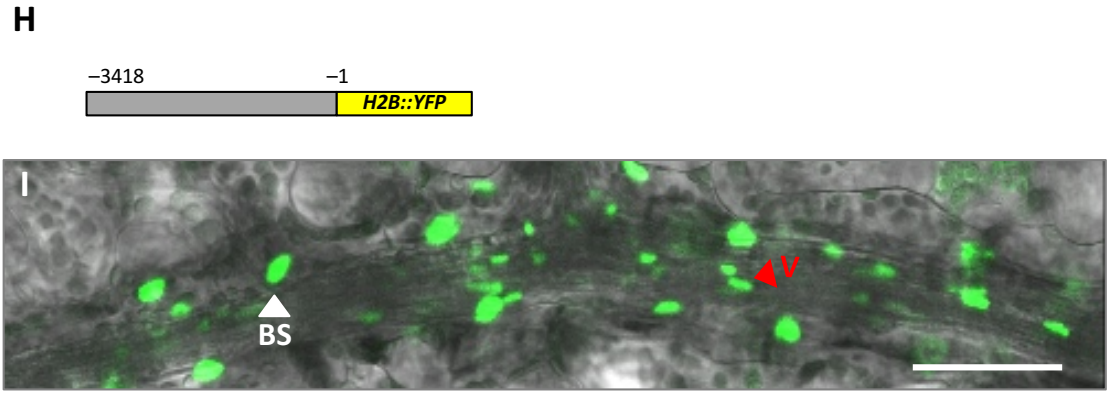

H
−3418
−1
H2B::YFP
I
BS
V

Figure 1

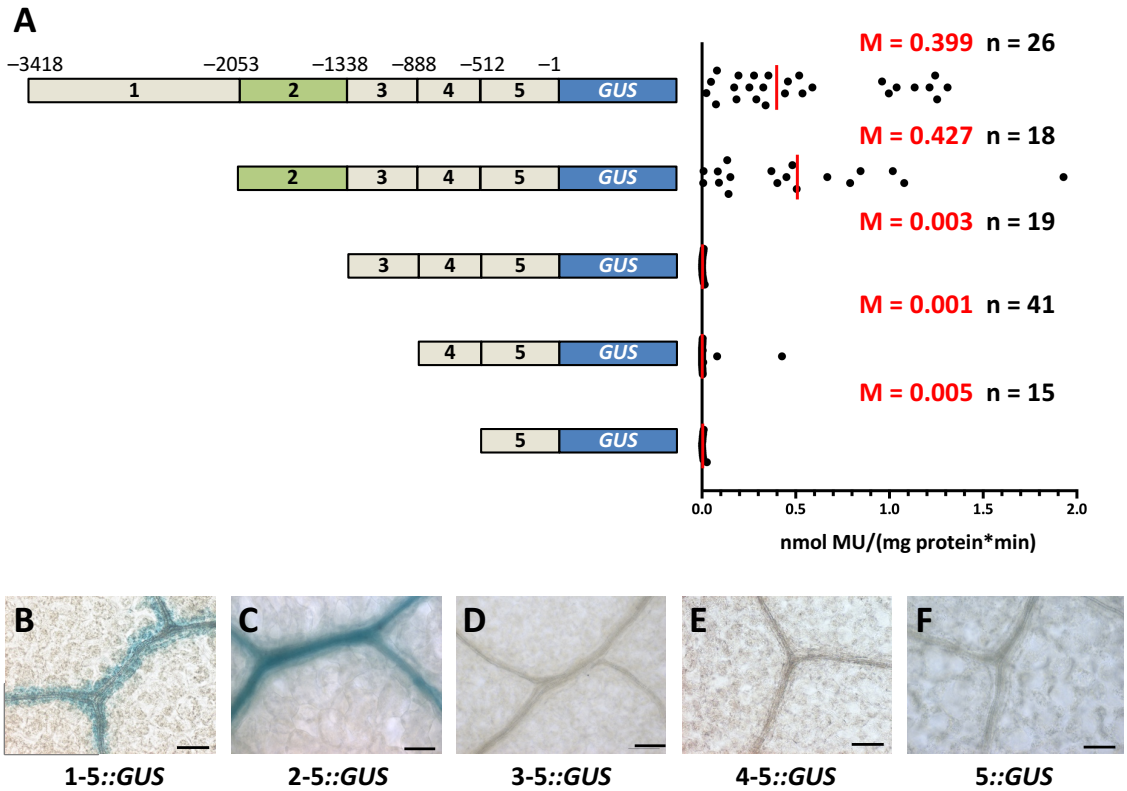

A
−3418
−2053
−1338
−888
−512
−1
1
2
3
4
5
GUS
M = 0.399 n = 26
M = 0.427 n = 18
M = 0.003 n = 19
M = 0.001 n = 41
M = 0.005 n = 15
0.0
0.5
1.0
1.5
2.0
nmol MU/(mg protein*min)
B
C
D
E
F
1-5::GUS
2-5::GUS
3-5::GUS
4-5::GUS
5::GUS

Figure 2

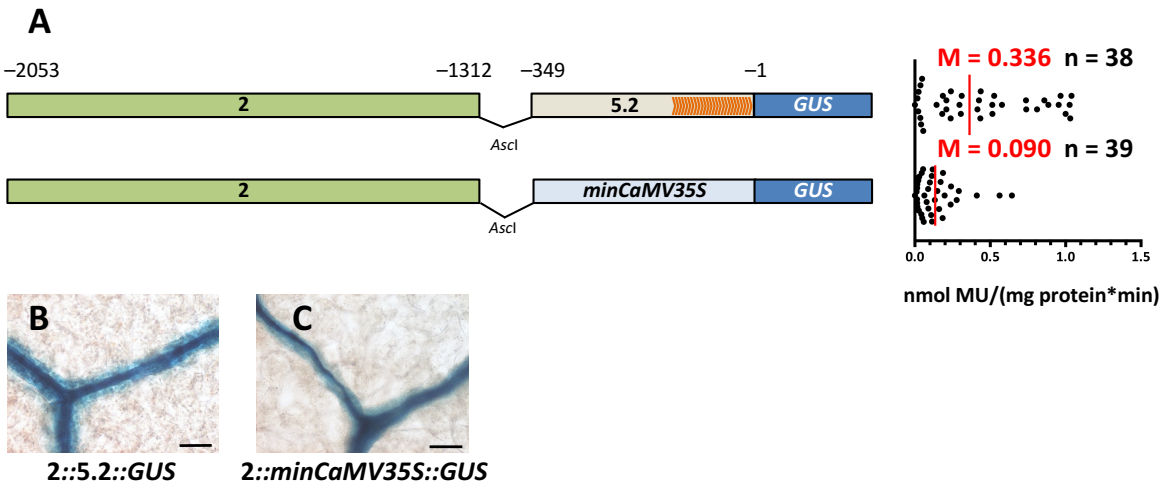
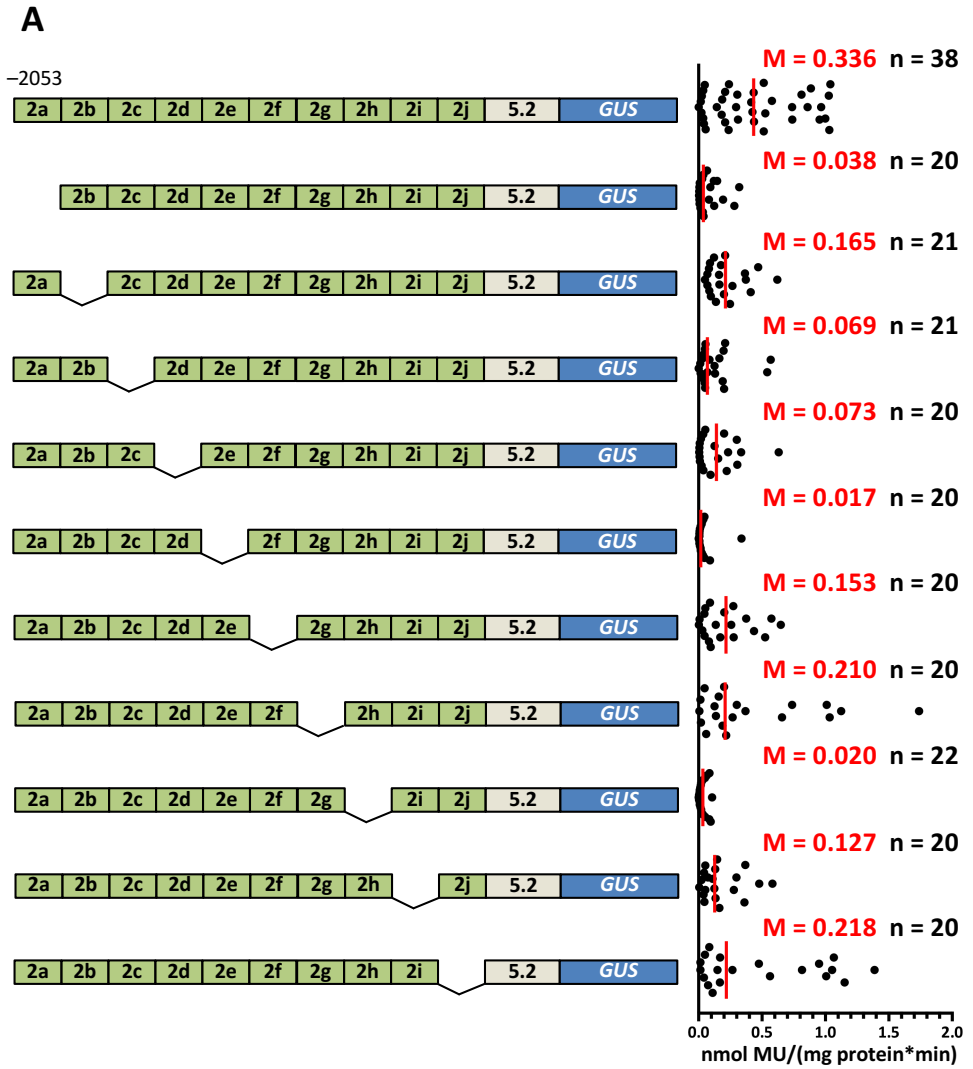

A
–2053
–1312
–349
–1
2
5.2
GUS
AscI
2
minCaMV35S
GUS
AscI
M = 0.336 n = 38
M = 0.090 n = 39
0.0
0.5
1.0
1.5
nmol MU/(mg protein*min)
B
C
2::5.2::GUS
2::minCaMV35S::GUS

Figure 3


A
−2053
2a 2b 2c 2d 2e 2f 2g 2h 2i 2j 5.2 GUS
M = 0.336 n = 38
M = 0.038 n = 20
M = 0.165 n = 21
M = 0.069 n = 21
M = 0.073 n = 20
M = 0.017 n = 20
M = 0.153 n = 20
M = 0.210 n = 20
M = 0.020 n = 22
M = 0.127 n = 20
M = 0.218 n = 20
0.0 0.5 1.0 1.5 2.0
nmol MU/(mg protein*min)

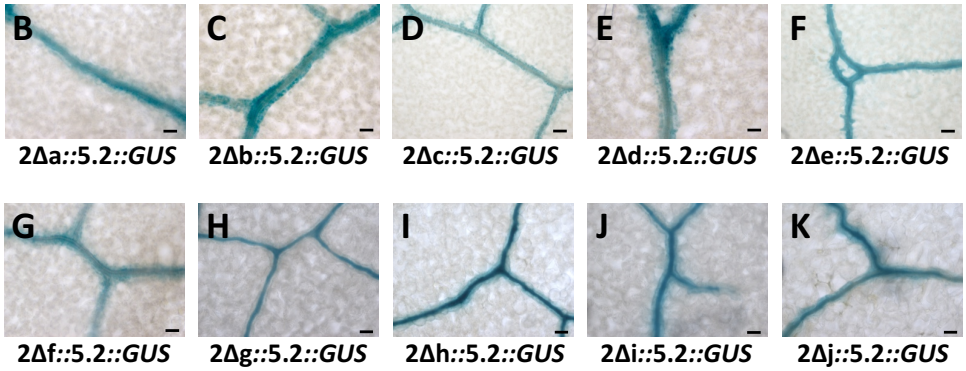
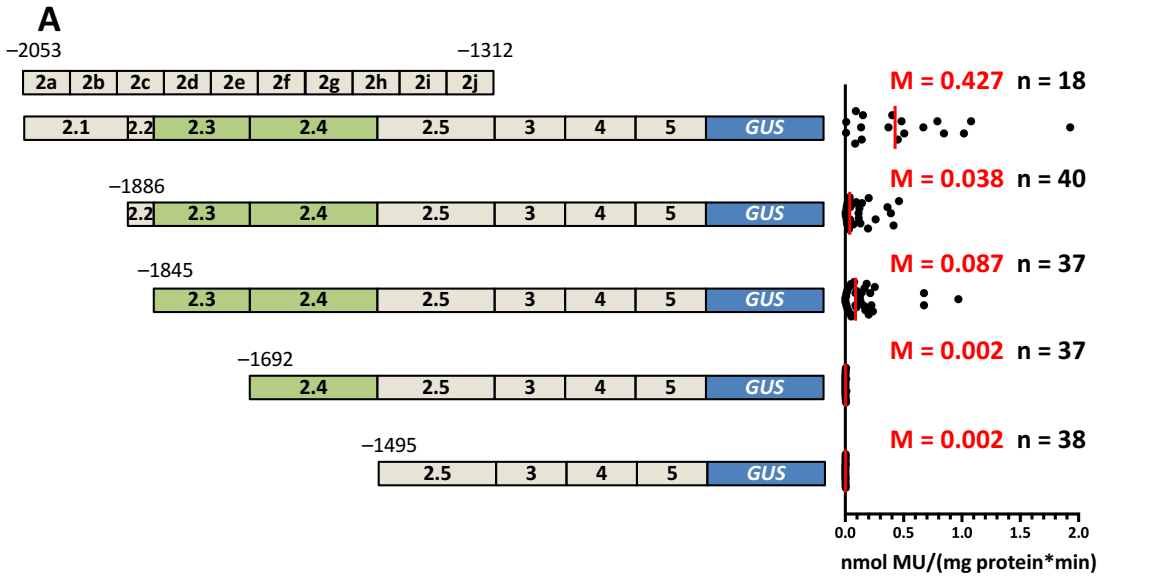
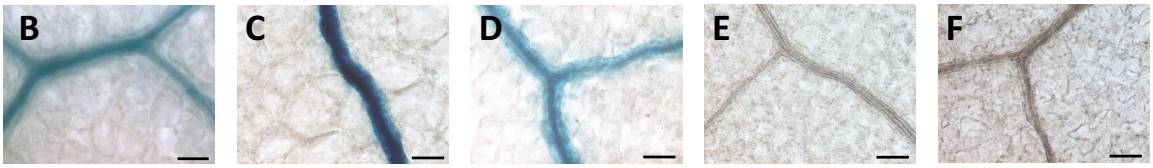
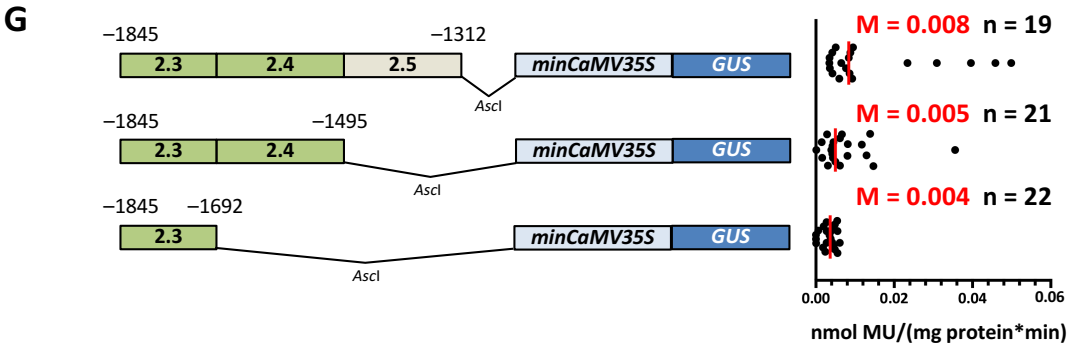

B
2Δa::5.2::GUS
C
2Δb::5.2::GUS
D
2Δc::5.2::GUS
E
2Δd::5.2::GUS
F
2Δe::5.2::GUS
G
2Δf::5.2::GUS
H
2Δg::5.2::GUS
I
2Δh::5.2::GUS
J
2Δi::5.2::GUS
K
2Δj::5.2::GUS

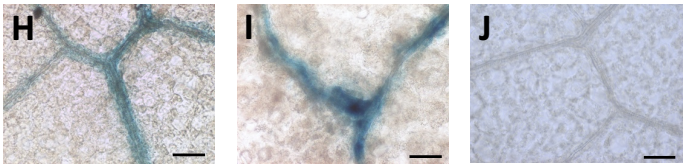
Figure 4




2.1-2.5::*GUS* 2.2-2.5::*GUS* 2.3-2.5::*GUS* 2.4-2.5::*GUS* 2.5-2.5::*GUS*




2.3-2.5::*min CaMV35S ::GUS* 2.3-2.4::*min CaMV35S ::GUS* 2.3::*min CaMV35S::GUS*

Figure 5

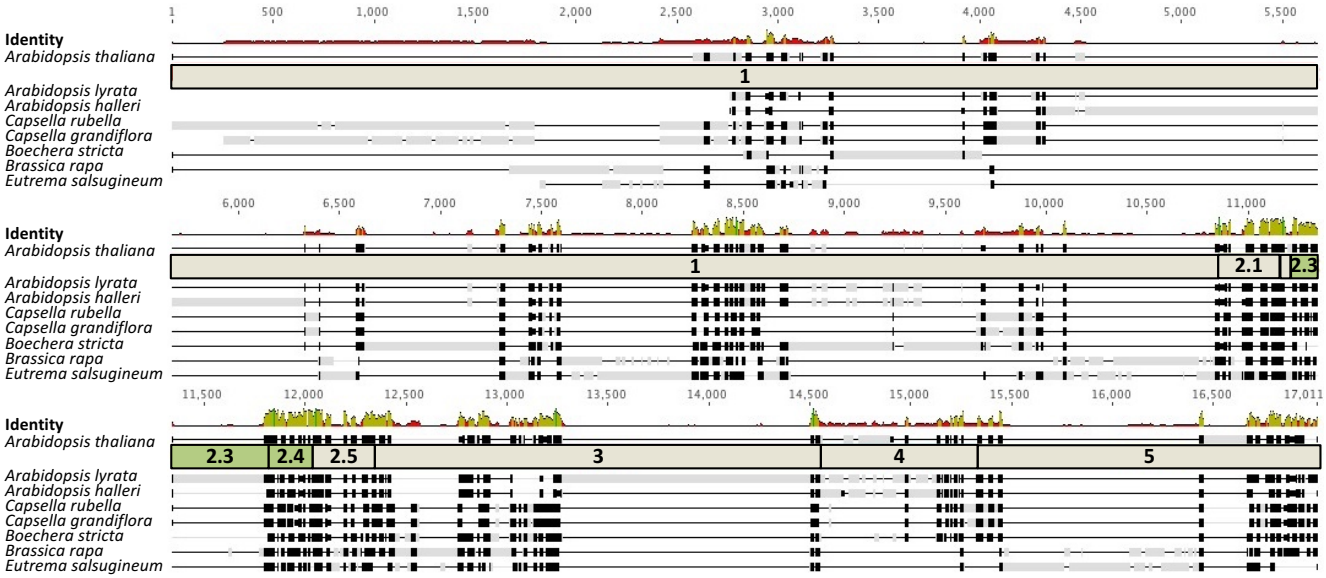

1
500
1,000
1,500
2,000
2,500
3,000
3,500
4,000
4,500
5,000
5,500
Identity
Arabidopsis thaliana
1
Arabidopsis lyrata
Arabidopsis halleri
Capsella rubella
Capsella grandiflora
Boechera stricta
Brassica rapa
Eutrema salsugineum
6,000
6,500
7,000
7,500
8,000
8,500
9,000
9,500
10,000
10,500
11,000
Identity
Arabidopsis thaliana
1
2.1
2.3
Arabidopsis lyrata
Arabidopsis halleri
Capsella rubella
Capsella grandiflora
Boechera stricta
Brassica rapa
Eutrema salsugineum
11,500
12,000
12,500
13,000
13,500
14,000
14,500
15,000
15,500
16,000
16,500
17,011
Identity
Arabidopsis thaliana
2.3
2.4
2.5
3
4
5
Arabidopsis lyrata
Arabidopsis halleri
Capsella rubella
Capsella grandiflora
Boechera stricta
Brassica rapa
Eutrema salsugineum

Figure 6

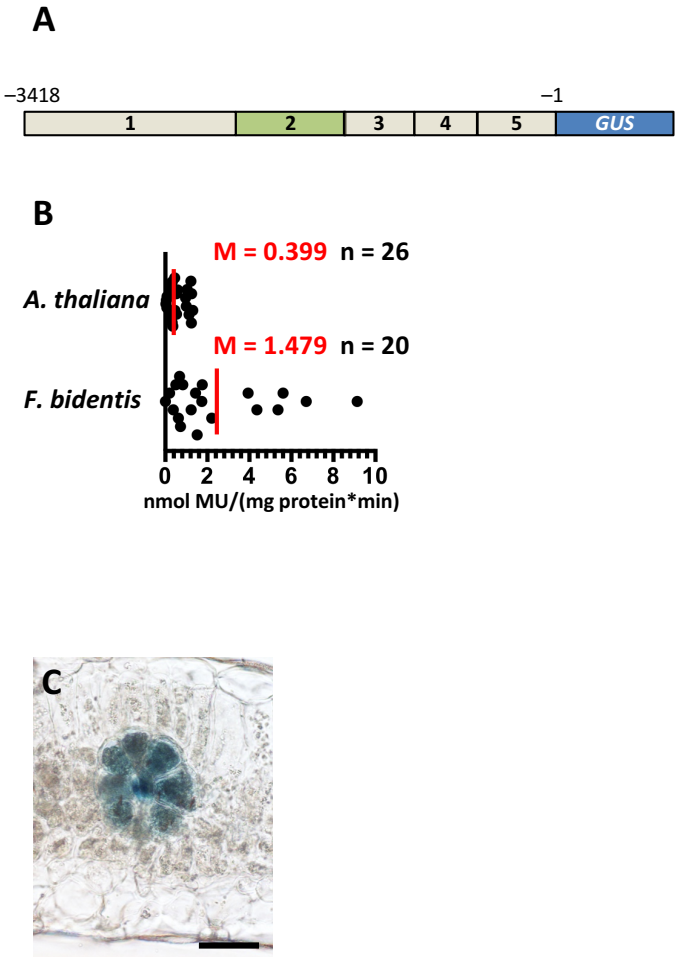

A
−3418
−1
1
2
3
4
5
GUS
B
M = 0.399 n = 26
A. thaliana
M = 1.479 n = 20
F. bidentis
0 2 4 6 8 10
nmol MU/(mg protein*min)
C

Figure 7