A bipartite transcription factor module controlling expression in the bundle sheath of *Arabidopsis thaliana*

Patrick J. Dickinson[1*], Jana Kneřová[1*], Marek Szecówka[1], Sean S. Stevenson[1], Steven J. Burgess[1], Hugh Mulvey[1], Anne-Maarit Bågman[2], Allison Gaudinier[2], Siobhan M. Brady[2] and Julian M. Hibberd[1]

[1]Department of Plant Sciences, Downing Street, University of Cambridge, Cambridge CB2 3EA, UK.

[2]Department of Plant Biology and Genome Center, University of California, Davis, CA 95616, USA

PJD - pd373@cam.ac.uk

JK - j.knerova@gmail.com

MS - szecowka@protonmail.com

SRS - srs62@cam.ac.uk

SJB - sburgess011@gmail.com

HM - hughmulvey@gmail.com

AMB - ambagman@ucdavis.edu

AG - agaudinier@berkeley.edu

SMB - sbrady@ucdavis.edu

JMH (corresponding) - jmh65@cam.ac.uk

* These authors contributed equally

Abstract

C_4 photosynthesis evolved repeatedly from the ancestral C_3 state, improving photosynthetic efficiency by ~50%. In most C_4 lineages photosynthesis is compartmented between mesophyll and bundle sheath cells but how gene expression is restricted to these cell types is poorly understood. Using the C_3 model *Arabidopsis thaliana* we identified *cis*-elements and transcription factors driving expression in bundle sheath strands. Upstream of the bundle sheath preferentially expressed *MYB76* gene we identified a region necessary and sufficient for expression containing two *cis*-elements associated with the MYC and MYB families of transcription factors. *MYB76* expression is reduced in mutant alleles for each. Moreover, down-regulated genes shared by both mutants are preferentially expressed in the bundle sheath. Our findings are broadly relevant for understanding the spatial patterning of gene expression, provide specific insights into mechanisms associated with evolution of C_4 photosynthesis and identify a short tuneable sequence for manipulating gene expression in the bundle sheath.

42 **Introduction**

43 A fundamental characteristic of multicellular eukaryotes is the ability to carry out diverse
44 and specialised functions in distinct tissues. Diversification in tissue function is associated
45 with variation in protein content that is to a large extent determined by patterns of gene
46 expression. One striking example of metabolic compartmentalisation is represented by C_4
47 photosynthesis where carbon is initially fixed in the mesophyll but is then released and re-
48 fixed in bundle sheath cells. The C_4 pathway is more efficient than C_3 photosynthesis under
49 warm, dry conditions, and as a consequence it has been proposed that engineering the C_4
50 pathway into C_3 crops such as rice could lead to increased yields (Hibberd et al., 2008; Von
51 Caemmerer et al., 2012). Understanding mechanisms directing expression to bundle sheath
52 or mesophyll cells is crucial to this effort and previous work has shown that multiple
53 mechanisms can drive cell type-preferential expression in C_4 species (Hibberd & Covshoff,
54 2010). For example, expression of the *Glycine decarboxylase P subunit (GLDPA)* in the
55 bundle sheath and veins of C_4 *Flaveria bidentis* is due to interplay between multiple
56 regulatory regions (Wiludda et al., 2012). Sequence in the distal promoter generates strong
57 expression but is not tissue-specific, however in the presence of proximal promoter
58 elements, expression in the bundle sheath is brought about by transcripts derived from the
59 distal promoter being degraded in mesophyll cells through nonsense-mediated RNA decay
60 of incompletely spliced transcripts (Engelmann et al., 2008; Wiludda et al., 2012). Similarly,
61 for the *Phosphoenolpyruvate carboxylaseA1 (PpcA1)* gene from C_4 *Flaveria trinervia*, two
62 submodules in a distal region that are enhanced by interaction with sequence in the proximal
63 promoter are sufficient to confer mesophyll specificity (Gowik et al., 2004; Akyildiz et al.,
64 2007). In addition to promoter sequences, other genic regions contain *cis*-elements that
65 generate tissue-specific gene expression. For example, preferential expression of the
66 *CARBONIC ANHYDRASE2, CARBONIC ANHYDRASE4* and
67 *PYRUVATE,ORTHOPHOSPHATE DIKINASE* genes in mesophyll cells of the C_4 species
68 *Gynandropsis gynandra* is mediated by a nine base pair motif present in both 5' and 3'
69 untranslated regions (Williams et al., 2016). Moreover, preferential expression of *NAD-*
70 *ME1&2* genes in the bundle sheath of *G. gynandra* is associated with two motifs known as
71 Bundle Sheath Modules (BSM) 1a and 1b that co-operatively restrict gene expression to this
72 tissue. BSM1a and BSM1b represent duons because they are located in coding sequence
73 and so determine amino acid composition as well as gene expression (Brown et al., 2011;
74 Reyna-Llorens et al., 2018). In summary, tissue-specific expression can be generated
75 through multiple mechanisms, but factors in *trans* that interact with the *cis*-elements
76 controlling tissue specific patterning of gene expression have not yet been identified.

77 As the C_4 pathway appears to have evolved repeatedly from the ancestral C_3 state by co-
78 opting existing molecular mechanisms from C_3 leaves (Matsuoka et al., 1994: Kajala et al.,
79 2012; Reyna-Llorens & Hibberd, 2017) we sought to leverage the C_3 model *Arabidopsis*
80 *thaliana* (hereafter Arabidopsis) to better understand mechanisms allowing cell type-specific
81 gene expression. The bundle sheath represents about 15% of cells in leaves of Arabidopsis
82 (Kinsman & Pyke, 1998) and has been proposed to play important roles in hydraulic
83 conductance (Shatil-Cohen et al., 2011), transport of metabolites (Leegood, 2008), as well
84 as storage of carbohydrates (Koroleva et al., 1997), ions (Williams et al., 2018) and water
85 (Sage, 2001; Griffiths et al., 2013). A number of findings are consistent with bundle sheath
86 cells also being involved in sulphur metabolism and glucosinolate biosynthesis. First, the
87 promoter of *SULPHUR TRANSPORTER2.2* generates preferential expression in the bundle
88 sheath (Takahashi et al., 2000; Kirschner et al., 2018) and secondly, compared with the
89 whole leaf, transcripts encoding enzymes of sulphur metabolism are more abundant on
90 bundle sheath ribosomes (Aubry et al., 2014). Transcripts of *MYB76* and other MYB domain
91 transcription factors involved in glucosinolate biosynthesis are also preferentially associated
92 with bundle sheath ribosomes (Aubry et al., 2014) but how this patterning of gene expression
93 is achieved is not known.
94 To better understand the regulation of cell type-specific gene expression in Arabidopsis
95 we focussed on the *MYB76* gene that is preferentially expressed in the bundle sheath. A
96 classical truncation analysis was combined with computational interrogation of transcription
97 factor binding sites to identify a 256-nucleotide region necessary and sufficient for
98 expression in the Arabidopsis bundle sheath. Within this region we identified MYC and MYB
99 transcription factor binding sites. We show MYC and MYB transcription factors are
100 necessary for *MYB76* expression as well as the expression of at least forty-seven additional
101 genes that are preferentially expressed in the bundle sheath. We propose that the MYC-
102 MYB module previously associated with expression of glucosinolate biosynthetic genes
103 (Schweizer et al., 2013) acts as a driver of bundle sheath preferential expression in
104 Arabidopsis. To our knowledge, this work provides the first example of a regulatory system
105 governing the spatial control of gene expression in leaves.

106 **Results**

107 **A short region in the *MYB76* promoter necessary for expression in the bundle sheath**

108 Gene Ontology (GO) analysis of publicly available data (Aubry et al., 2014) indicated that
109 in addition to transcripts encoding proteins important for amino acid export, those encoding
110 glucosinolate (GLS) biosynthesis proteins are strongly enriched in the bundle sheath of
111 Arabidopsis (Figure 1A). In fact, all but five of the thirty genes reported to be involved in GLS
112 biosynthesis (Li et al., 2014) showed more than two-fold higher expression in the bundle
113 sheath compared with the whole leaf (Figure 1B). The expression of genes involved in
114 aliphatic GLS metabolism is mostly controlled by MYC and MYB transcription factors
115 (Gigolashvili et al., 2008; Schweizer et al., 2013) including MYC2, MYC3, MYC4 and
116 MYB28, MYB29 and MYB76. Notably, genes encoding MYB28, MYB29 and MYB76 were
117 strongly expressed in the bundle sheath compared with the whole leaf (Figure 1B). We
118 sought to use these transcription factors to better understand how gene expression is
119 restricted to the bundle sheath of C_3 plants.

120 To test whether regions upstream of *MYB28, MYB29* and *MYB76* were sufficient to drive
121 expression in the bundle sheath of Arabidopsis, each was fused to the *uidA* reporter gene
122 encoding GUS and multiple transgenic lines were generated. We were unable to detect GUS
123 staining in leaves from the promoter of *MYB28* alone (Supplemental Figure 1) and whilst the
124 promoter of *MYB29* did mark veins and bundle sheath cells, it also led to some GUS
125 accumulating in mesophyll cells (Supplemental Figure 1). In contrast, a construct containing
126 the promoter and 279bp after the translational start site of *MYB76* generated clearly
127 detectable GUS in the Arabidopsis bundle sheath with no GUS detected in the mesophyll
128 (Figure 1C, Supplemental Figure 2). We conclude that regulatory elements sufficient for
129 bundle sheath expression of *MYB76* (Figure 1B) are contained in this sequence but the
130 preferential accumulation of transcripts from *MYB28* and *MYB29* (Supplemental Figure 1)
131 is likely mediated by *cis*-elements located outside of the sequence tested.

132 As expression patterns can be determined by *cis*-elements in promoters, untranslated
133 regions, exons, introns or downstream 3' regions (Ali & Taylor, 2001; Brown et al., 2011;
134 Kajala et al., 2011; Williams et al., 2016; Gallegos & Rose, 2017) a translational fusion
135 between the *MYB76* genomic sequence and *uidA* driven by the *MYB76* promoter was
136 generated to confirm that the strong expression in the bundle sheath mediated by the
137 *MYB76* promoter reflected the pattern of expression associated with the intact genomic
138 sequence. Transgenic lines harbouring this genomic fusion showed preferential
139 accumulation of GUS in the bundle sheath (Figure 1D, Supplemental Figure 3) mirroring the
140 pattern found from the promoter alone. Use of the fluorometric 4-MethylUmbelliferyl β-D-

141 Glucuronide (MUG) assay showed that GUS accumulation was lower when nucleotides
142 +280 to +1254 relative to the translational start site were included (Supplemental Figure 6)
143 suggesting that the full genomic sequence of *MYB76* contains regulators that quantitatively
144 repress expression. To confirm that the GUS reporter generated a reliable read-out of spatial
145 expression patterns, a nuclear localised *pMYB76::H2B::GFP* line was produced. Imaging of
146 GFP in deep tissue of leaves such as the bundle sheath is challenging, but consistent with
147 data from the GUS reporter (Figure 1C), GFP was detectable in nuclei of the vasculature
148 and bundle sheath cells but was absent from the mesophyll (Figure 1E, Supplemental Figure
149 4). These results show that the promoter of *MYB76* generates bundle sheath preferential
150 expression.

151 To further investigate the elements driving *MYB76* expression in the bundle sheath, 5'
152 deletions of the promoter were generated. Removal of nucleotides -1725 to -1264 relative
153 to the translational start site did not impact on GUS localisation, however once nucleotides
154 -1264 to -796 were removed GUS was no longer detectable in the bundle sheath. Further
155 removal of another 500 base pairs had no additional impact on the spatial pattern of GUS
156 accumulation (Fig 1F, Supplemental Figure 5). These findings are supported by
157 quantification *via* MUG assays (Supplemental Figure 6) that showed removal of nucleotides
158 -1725 to -1264 reduced accumulation of the reporter, and MUG was no longer detectable
159 once sequence upstream of nucleotide -796 was absent. Overall, these data indicate that
160 the *MYB76* promoter contains a region between nucleotides -1264 and -796 upstream of
161 the translational start site that directs expression to the bundle sheath.
162

163 **A DNaseI Hypersensitive Site in the *MYB76* promoter is necessary and sufficient for**
164 **bundle sheath preferential expression**

165 The DNaseI enzyme preferentially cuts accessible DNA and so can be used to define
166 sequences available for transcription factor binding and thus the location of regulatory DNA
167 (Hesselberth et al., 2009). To complement our truncation analysis, an existing dataset that
168 defined DNaseI Hypersensitive Sites (DHS) in Arabidopsis (Zhang et al., 2012) was
169 interrogated. Two DHS were detected upstream of *MYB76* in both flower tissue and leaves
170 (Figure 2A) with a DHS encompassing nucleotides -909 to -654 upstream of the translational
171 start site overlapping with the region required for expression in the bundle sheath (Figure
172 1F). Consistent with the DHS data, *MYB76* has been reported to be expressed in both
173 leaves and flowers (Gigolashvili et al. 2008). Although the DHS had a lower DHS score in
174 the leaf than the flower (mean DHS score of 1.3 versus 2.6 (Zhang et al., 2016)), this may
175 be due to the fact that bundle sheath cells make up a small proportion of cells in a C_3 leaf

176 (Kinsman & Pyke, 1998) such that transcription factor binding upstream of *MYB76* would be
177 diluted in whole leaves.
178 Removing the DHS found between nucleotides -909 to -654 of the *MYB76* promoter
179 abolished accumulation of GUS, and fusing this DHS to the minimal *CaMV35S* promoter
180 was sufficient to generate GUS in the bundle sheath (Figure 2B, Supplemental Figure 7).
181 Furthermore, oligomerizing two copies of the DHS upstream of the minimal *CaMV35S*
182 promoter resulted in very strong accumulation of GUS in the bundle sheath (Figure 2B,
183 Supplemental Figure 7). From these data we conclude that sequence within this DHS is both
184 necessary and sufficient to activate expression preferentially in the bundle sheath.
185 Combined with the truncation analysis indicating that nucleotides -1264 to -796 upstream of
186 the translational start site were required for expression in the bundle sheath (Figure 1F,
187 Supplemental Figure 5), our findings indicate that a positive regulator of bundle sheath
188 expression is located between nucleotides -909 (the start of the DHS) and -796 upstream of
189 the *MYB76* translational start site.
190 Phylogenetic foot-printing identified two motifs (Figure 2C) in the *MYB76* DHS that are
191 shared by *MYB76* and the promoters of *SCR, SULTR2.2* and *GLDP* that have previously
192 been reported to generate expression in the Arabidopsis bundle sheath (Wysocka-Diller et
193 al. 2000; Engelmann et al. 2008; Takahashi et al. 2000; Kirschner et al. 2018). Whilst site
194 directed mutagenesis of motif one (TGGGCA) had no impact on accumulation of GUS in the
195 bundle sheath (Supplemental Figure 8) deletion of motif two (TGCACCG) in the context of
196 the full genomic sequence of *MYB76* abolished GUS accumulation in the bundle sheath
197 (Figure 2D, Supplemental Figure 8). These data indicate that this sequence is necessary to
198 pattern expression from both the promoter of *MYB76* alone, but also the full genic *MYB76*
199 sequence containing exons, introns and UTRs. To test whether this sequence is sufficient
200 to direct expression to the bundle sheath it was combined with ten upstream and ten
201 downstream nucleotides from the endogenous *MYB76* promoter, oligomerized, and fused
202 to *uidA*. Although this construct did not recapitulate the strong bundle sheath expression of
203 the *MYB76* DHS it was able to generate weak expression in the bundle sheath (Figure 2D,
204 Supplemental Figure 10). We conclude that the 27bp sequence is necessary and weakly
205 sufficient to direct expression to the bundle sheath.
206
207 **MYC, MYB and DREB transcription factors control bundle sheath expression of**
208 ***MYB76* from the DHS**
209 To better understand the *cis*-regulatory landscape within the *MYB76* DHS we used the
210 Find Individual Motif Occurrences (FIMO) tool (Grant et al., 2011) to predict transcription

211 factor binding sites. For the majority of Arabidopsis transcription factors, DNA binding sites
212 have not yet been defined, and so to allow us to search for broad consensus sequences
213 associated with groups of transcription factors we clustered the 555 transcription factor
214 binding motifs for Arabidopsis from the JASPAR motif database (Formes et al., 2019) into
215 43 groups based on relatedness of the motif position weight matrices (PWMs)
216 (Supplemental Table 1). Plotting matches for each of these motifs in the *MYB76* DHS
217 (Figure 3A) showed that it contained binding sites for motifs from twelve clusters. Notably
218 the 27bp region necessary for expression in the bundle sheath contained binding sites from
219 clusters 1, 8, 11 and 16 that correspond to the binding sites from ERF (clusters 1 and 11),
220 bHLH (cluster 8), and G2-like (cluster 16) transcription factor families. There are also
221 predicted binding sites for IDD (cluster 9), MADS (cluster 23) and MYB (clusters 10 and 18)
222 families within the DHS. To supplement this *in silico* analysis, Yeast One-Hybrid identified
223 thirteen transcription factors from seven different families that were able to bind the DHS
224 (Figure 3B). Six of these transcription factors including DF1, MYB73 and AIL5 were
225 previously reported to bind the whole *MYB76* promoter (Li et al. 2014) (Figure 3B). Although
226 not identified by Yeast One-Hybrid, MYC2, MYC3, MYC4, MYB28 and MYB29 can control
227 *MYB76* expression (Gigolashvili et al., 2008; Schweizer et al., 2013) and so were
228 incorporated into our list of candidate regulators of *MYB76*. Of these candidates, *MYB28*,
229 *MYB29*, *DF1*, *MYB73* and *AIL5* were strongly preferentially expressed in the bundle sheath
230 (with a log2(bundle sheath/whole leaf) > 0.75 cut off) (Figure 3B).
231 We next tested whether these candidate transcription factors could activate expression
232 from the *MYB76* DHS *in planta.* For the following reasons we chose to test MYC2, MYC3,
233 MYC4, DREB2A, DREB26, DF1 and MYB73. First, MYC2, MYC3, MYC4, DREB2A and
234 DREB26 have binding sites for the clusters described above within the 27bp region and
235 MYC2, MYC3 and MYC4 have previously been reported to affect *MYB76* expression
236 (Schweizer et al., 2013). Second, DREB2A was found in our Yeast One-Hybrid experiment
237 and DREB26 was identified in both Yeast One-Hybrid screens (Figure 3B). Third, DF1 and
238 MYB73 were found in both Yeast One-Hybrid studies (Figure 3B), are preferentially
239 expressed in the bundle sheath (Figure 3B) and DF1 has previously been associated with
240 *GLS* gene expression (Li et al., 2014). Although the DHS contains two additional predicted
241 MYB binding sites (one associated with cluster 10 and one with cluster 18 MYBs) these
242 were outside the region necessary for activating *MYB76* expression in the bundle sheath
243 (Figure 3A). Each transcription factor was used in a *trans*-activation assay for the *MYB76*
244 DHS in *Nicotiana benthamiana.* Infiltration with MYC2, MYC3, MYC4 and DREB2A resulted
245 in significantly more LUCIFERASE (LUC) signal than infiltration with the DHS alone, with

246 MYC2, MYC3 and MYC4 driving higher LUC signal than DREB2A (Figure 3C). LUC signal
247 was not significantly different from the DHS alone for any of the other transcription factors
248 tested (Figure 3C).
249 As the MYCs (Cluster 8) and DREB2A (Cluster 11) have predicted binding sites that
250 overlap the 27bp region necessary for expression in the bundle sheath (Figure 3A) and were
251 able to activate expression from the DHS in the *trans*-activation assays (Figure 3C) we
252 tested whether expression of *MYB76* was perturbed in mutant alleles of each. qRT-PCR on
253 *MYB76* was performed on *myc2/3/4* and *dreb2a* mutants. *MYB76* expression was reduced
254 by approximately half in the *dreb2a* mutant, and in the *myc2/3/4* triple mutant by about 19
255 times (Figure 3D). These data are consistent with the *trans*-activation results which showed
256 a strong increase in expression driven by MYC transcription factors and a weaker increase
257 driven by DREB2A (Figure 3C). Taken together this indicates that under the conditions we
258 used MYC2, MYC3 and MYC4 have a major role in controlling *MYB76* expression and
259 DREB2A has a smaller effect. Previous work has shown that MYC transcription factors
260 interact with MYB28, MYB29 and MYB76 to activate the expression of genes involved in
261 GLS metabolism (Schweizer et al., 2013). Therefore, despite them not appearing in either
262 Yeast One-Hybrid screen, we asked whether they are involved in controlling *MYB76*
263 expression. Re-analysis of publicly available data (Burow et al., 2015) showed that *MYB76*
264 expression was substantially reduced in a *myb28/29* mutant (Figure 3E). This is consistent
265 with previous reports of MYB28, 29 and 76 being able to activate each other's expression
266 (Gigolashvili et al., 2008). *MYB76* is expressed at similar levels to wild type in both *myb28*
267 and *myb29* single mutants (Sønderby et al., 2007) suggesting that there is redundancy
268 between *MYB28* and *MYB29* in the control of *MYB76* expression. Although *MYB29* and
269 *MYB76* are tandem duplicates on chromosome five, because *MYB76* is expressed similarly
270 to wild type in a *myb29* mutant, the reduction of *MYB76* expression in the *myb28/29* mutant
271 is unlikely to be a result of the proximity of the *myb29-1* T-DNA insertion to *MYB76.* MYB28,
272 29 and 76 do not have defined transcription factor binding motifs in publicly available
273 databases, however mapping motif clusters to a phylogenetic reconstruction of MYB
274 transcription factors (Supplemental Figure 11) showed that MYB28, MYB29 and MYB76
275 were found in the cluster 18 clade, strongly suggesting that their binding preference is similar
276 to those of cluster 18 MYBs (Supplemental Table 1). Taken together these data suggest
277 that MYC2/MYC3/MYC4 as well as MYB28/MYB29 are involved in controlling *MYB76*
278 expression from the DHS.
279

280 **MYC2/3/4 and MYB28/29 control the expression of many other bundle sheath**
281 **preferential genes in Arabidopsis**
282 The findings above are consistent with the MYC-MYB module, previously reported to
283 activate GLS metabolism genes in response to herbivory (Schweizer et al., 2013) activating
284 *MYB76* in the bundle sheath in the absence of herbivory. As this module is involved in
285 activating the expression of multiple GLS metabolism genes in Arabidopsis (Schweizer et
286 al., 2013) and most are preferentially expressed in the bundle sheath (Figure 1B), we wished
287 to test whether the MYC-MYB system might also be responsible for their bundle sheath
288 preferential expression. We therefore re-analysed publicly available transcriptome data for
289 *myc2/3/4* (Major et al., 2017) and *myb28/29* (Burow et al., 2015) mutants. We identified 207
290 genes that were down-regulated (log2 vs. WT < -0.75) in *myc2/3/4* (Figure 4A), 729 genes
291 that were down-regulated in *myb28/29* (Figure 4B) and 76 genes that were down-regulated
292 in both *myc2/3/4* and *myb28/29* (Figure 4C). Next, we used a published dataset (Aubry et
293 al., 2014) to test whether any of these gene sets were preferentially associated with the
294 Arabidopsis bundle sheath. Genes down-regulated only in *myc2/3/4* (Figure 4D) or
295 *myb28/*29 (Figure 4E) were not preferentially expressed in either the bundle sheath or whole
296 leaf. However, of the 54 genes down-regulated in both *myc2/3/4* and *myb28/29* and present
297 in the Aubry et al., (2014) dataset, 47 were strongly bundle sheath preferential (log2(bundle
298 sheath/whole leaf) > 0.7) (Figure 4F). Only four genes were strongly depleted in the bundle
299 sheath (log2(bundle sheath/whole leaf) < -0.7) (Figure 4F). Consistent with these down-
300 regulated genes being directly regulated by the MYC-MYB system, motif enrichment
301 analysis showed that the MYC2/MYC3/MYC4 (cluster 8) and MYB28/MYB29 (cluster 18)
302 motifs were strongly enriched in promoters of genes down-regulated in both *myc2/3/4* and
303 *myb28/29* mutants (Figure 4I). This was not the case for the *myc2/3/4* mutant (Figure 4G)
304 or the *myb28/29* mutant alone (Figure 4H).
305 Previous work suggested that MYBs and MYCs bind to adjacent regions of promoters to
306 activate expression of *GLS* genes (Schweizer et al., 2013). This was also the case in the
307 *MYB76* DHS (Figure 3A). To test if this was also true for genes down-regulated in both
308 *myc2/3/4* and *myb28/29* we investigated minimum distance between cluster 8 and 18 motifs
309 in each set of promoters and compared this with random sets of genes. This showed that
310 as well as being more enriched in the promoters of down-regulated genes in both mutants
311 (Figure 4I) where cluster 8 and 18 motifs were both present, they were closer together in
312 the down-regulated genes common to *myc2/3/4* and *myb28/29* compared with those down-
313 regulated in only one mutant background, or in any of the random sets (Figure 4J). In
314 summary, genes down-regulated in both *myc2/3/4* and *myb28/29* were generally strongly

315 preferential to the bundle sheath, showed an enrichment of MYC2/3/4 and MYB28/29/76
316 binding sites in their promoters and where these motifs are both present, they were closer
317 together than the other gene sets that we assessed. Taken together these findings suggest
318 that the MYC-MYB module is important in controlling the expression of at least forty-seven
319 genes in the bundle sheath of Arabidopsis.

320 **Discussion**

321 **A positive regulator patterning gene expression to the bundle sheath**

322 Although many promoters allowing expression in defined tissues of the shoot or root have
323 been reported, our understanding of how these expression domains are generated is limited.
324 Examples include promoters that drive expression in tissues such as apical meristems
325 (Baima et al., 1995; Levin and Meyerowitz, 1995; Fletcher et al., 1999; Otsuga et al., 2001;
326 Sarkar et al., 2007) the stele (Helariutta et al., 2000; Bonke et al., 2003), endodermis
327 (Wysocka-Diller et al., 2000), cortex (Heidstra et al., 2004; Mustroph et al., 2009) and
328 trichoblasts or atrichoblasts (Ruzicka et al., 2007; Masucci et al., 1996) of the root, as well
329 as guard cells (Nakamura et al., 1995), phloem companion cells (Imlau et al., 1999) and
330 epidermal cells (Thoma et al., 1994) of the shoot. In leaves, perhaps the best characterised
331 of these promoters come from analysis of C_4 species (Gowik et al., 2004; Akyildiz et al.,
332 2007; Wiludda et al., 2012). For example, mesophyll specific expression of *PPCA1* from *F.*
333 *trinervia* is due to two modules in a distal region of the promoter (Gowik et al., 2004; Akyildiz
334 et al., 2007), whilst the *GLDPA* promoter from C_4 *F. bidentis* generates expression in the
335 bundle sheath because proximal promoter sequence leads to transcripts derived from a
336 distal promoter element being degraded in mesophyll cells (Engelmann et al., 2008; Wiludda
337 et al., 2012). In roots, using *SHORTROOT* as a model, it has been proposed that multiple
338 *cis*-elements recognised by a complex network of both activators and repressors confine
339 expression to the root vasculature (Sparks et al., 2016). Analysis of transcription factor
340 binding *in vivo* is consistent with a highly combinatorial mosaic of regulatory DNA
341 underpinning patterns of gene expression (Neph et al., 2012). To our knowledge, and
342 perhaps due to this highly complex regulatory landscape, there are no examples of *cis*-
343 elements and interacting transcription factors that limit gene expression to specific cell types
344 in leaves. In this work, we show that combining DNaseI-SEQ data with functional analysis
345 allowed identification of *cis*-elements and cognate transcription factors that pattern
346 expression to bundle sheath cells of the leaf.

347 The module that generates expression in bundle sheath cells contains two *cis*-elements
348 recognised by MYC and MYB transcription factors. When sequence containing this region
349 is oligomerised it leads to strong and specific expression in the bundle sheath and so
350 represents a short sequence that could be used to mis-express genes in this tissue. Bundle
351 sheath cells link the vasculature to the photosynthetic mesophyll cells. In C_3 species, they
352 play important roles including the control of hydraulic conductance (Sage, 2001; Shatil-
353 Cohen et al., 2011, Griffiths et al., 2013), transport of metabolites in and out of veins
354 (Leegood, 2008), responses to high light episodes (Fryer et al., 2003), and assimilation of

355 sulphur (Aubry et al., 2014). However, there are relatively few promoters available to drive
356 or perturb expression in these cells (Takahashi et al., 2000; Wysocka-Diller et al., 2000).
357 Short synthetic promoters have a number of advantages over the long promoter fragments
358 currently available to direct gene expression to bundle sheath cells. These include reducing
359 the likelihood of homology-based gene silencing if used more than once in any construct
360 (Bhullar et al., 2003) and decreasing the chances of leakiness or off-target gene expression
361 associated with use of full-length promoter fragments (Hernandez-Garcia & Finer, 2014).
362 Oligomerization of *cis*-elements to achieve higher expression levels is a common strategy
363 when creating synthetic promoters (Dey et al., 2015). As the *MYB76* DHS is short and can
364 be oligomerized to tune expression levels, it appears to represent a promising fragment with
365 which to perturb and modify functions including the control of hydraulic conductance,
366 metabolite transport, responses to high light, and assimilation of sulphur in the bundle
367 sheath.
368
369 **Control of bundle sheath expression by a MYC-MYB module**
370 Glucosinolates are a diverse group of nitrogen and sulphur containing secondary
371 metabolites that accumulate preferentially around the mid-vein and outer lamina in
372 Arabidopsis rosette leaves and which are involved in defence against herbivory (Shroff et
373 al., 2008). The methionine-derived aliphatic glucosinolate biosynthetic pathway is largely
374 controlled by MYB28, MYB29 and MYB76 in combination with MYC2, MYC3 and MYC4
375 (Sønderby et al., 2007; Gigolashvili et al., 2008; Malitsky et al., 2008; Sønderby et al., 2010;
376 Schweizer et al., 2013; Li et al., 2014). The *MYB76* promoter has previously been reported
377 to generate expression in the vasculature (Gigoyashvili et al., 2008). Our analysis now
378 shows that its expression domain includes the bundle sheath, and that this is under control
379 of a MYC-MYB module. None of the MYC2*,* MYC3*,* MYC4*,* MYB28*,* MYB29 or MYB76
380 transcription factors were identified in our Yeast One-Hybrid analysis (Figure 3B) or
381 previously published analysis of the whole *MYB76* promoter (Li et al., 2014). This may be
382 due to these transcription factors requiring additional partners to bind DNA, consistent with
383 the model of MYC-MYB dimers being required for the activation of target genes (Schweizer
384 et al., 2013). Although this appears inconsistent with the *trans*-activation assays where
385 infiltration of individual MYC transcription factors activated expression from the *MYB76* DHS
386 (Figure 3C), this may be due to their interaction with endogenous MYBs from *N.*
387 *benthamiana*. Alternatively, and consistent with the MYC binding site alone being able to
388 generate weak expression in the bundle sheath (Figure 2D), the interaction may be below
389 the detection limit of the Yeast One-Hybrid assay.

390 It is theoretically possible that this MYC-MYB module acts as a general activator of
391 transcription across all cells of the leaf and another distinct mechanism represses
392 expression in other cell types. However, for the following reasons, we favour a model in
393 which the MYC-MYB module activates expression preferentially in the bundle sheath. First,
394 the 256bp DHS which contains the MYC-MYB binding sites is sufficient for bundle sheath
395 preferential expression, and the 27bp region containing the potential MYC binding site, but
396 lacking MYB binding sites, is sufficient for weak expression in the bundle sheath. Second,
397 *MYB28* and *MYB29* are preferentially expressed in the bundle sheath (Figure 1B) consistent
398 with them activating expression in this cell-type. Third, other transcription factor binding sites
399 required for directing bundle sheath preferential expression would have to be present in the
400 27bp region. This is possible, however we would expect such binding sites to be enriched
401 in promoters of additional genes that are strongly expressed in the bundle sheath of wild-
402 type plants but down regulated in both *myc2/3/4* and *myb28/29* mutants. No such binding
403 sites were detected (Figure 4I). In summary, we cannot completely rule out a requirement
404 for other factors operating in other cell types such that *MYB76* expression is restricted to the
405 bundle sheath. However, the MYC-MYB module activating expression preferentially in the
406 bundle sheath is a more parsimonious mechanism. To our knowledge, this MYC-MYB
407 module provides the first example of a regulatory system governing the spatial control of
408 gene expression in leaves.
409 At this point, how the MYC-MYB module directs preferential expression in the bundle
410 sheath is not clear. *MYB28* and *29* transcripts accumulate preferentially in the bundle sheath
411 (Figure 1B) but this is not as apparent for transcripts of genes encoding *MYC2, MYC3* and
412 *MYC4* (Figure 3B). MYC transcription factors are regulated by jasmonic acid (JA) signalling
413 thorough interactions with JASMONATE-ZIM DOMAIN (JAZ) proteins (Fernandez-Calvo et
414 al., 2011; Howe et al., 2018). Whether there is a link between JA signalling and bundle
415 sheath preferential gene expression remains to be determined.
416
417 **Roles of additional transcription factors in controlling *MYB76***
418 Many additional transcription factors were identified as binding the *MYB76* DHS in Yeast
419 One-Hybrid analysis (Figure 3B) and some of these have been reported to bind the entire
420 promoter previously (Li et al., 2014). These transcription factors are likely important for
421 controlling other aspects of *MYB76* expression in addition to the spatial patterning
422 determined by the MYC-MYB module. For example, multiple ERF family transcription factors
423 interacted with the *MYB76* DHS (Figure 3B) and there is an ERF family transcription factor
424 binding site in the DHS (Figure 3A). Additionally, DREB2A weakly activates expression from

425 the *MYB76* DHS (Figure 3C) and *MYB76* transcripts were less abundant in a *dreb2a* mutant
426 allele (Figure 3D). Recent work has linked auxin signalling and glucosinolate biosynthesis
427 under drought conditions with DREB2A/B signalling (Salehin et al., 2019). This suggests
428 that DREB2A may have a role in regulating *MYB76* expression in response to environmental
429 stimuli. Because DREB2A transcripts do not accumulate preferentially in the bundle sheath
430 (Figure 3B) it is not clear how activation of *MYB76* outside of the bundle sheath is avoided.
431 Possibilities include DREB2A activity being limited to the bundle sheath by post-
432 transcriptional and/or post-translational mechanisms or a requirement for other binding
433 partners. Post-transcriptional and post-translational regulation has been reported for
434 DREB2A (Agarwal et al., 2017) with post-transcriptional regulation by alternative splicing
435 being reported (Egawa et al., 2006; Qin et al., 2007; Matsukura et al., 2010; Vainonen et al.
436 2012). Whilst overexpression of *DREB2A* in Arabidopsis does not affect the expression of
437 target genes (Liu et al., 1998) an isoform lacking key phosphorylation sites activates the
438 majority of DREB2A targets (Sakuma et al., 2006). Thus, post-translational regulation is
439 essential for DREB2A activity suggesting a mechanism for restricting DREB2A activation of
440 *MYB76* expression to the bundle sheath. The *MYB76* DHS also contains a MADS domain
441 transcription factor binding site (Figure 3A). MADS domain transcription factors are involved
442 in controlling gene expression required for flower development (Theißen et al., 2016) and
443 could potentially be involved in directing *MYB76* expression to flowers where significant
444 levels of GLS accumulate (Sarsby et al., 2012). In summary, as well as the MYC-MYB
445 module generating bundle sheath preferential expression in leaves, multiple other
446 transcription factor binding sites in the DHS may be important for controlling other aspects
447 of *MYB76* expression such as responses to environmental stimuli and expression in different
448 organs.
449
450 **The role of the MYC-MYB module outside glucosinolate biosynthesis**
451 Our data are consistent with the MYC-MYB module patterning the expression of at least
452 forty-seven genes to the Arabidopsis bundle sheath. This represents about 3.6% of the 1316
453 genes previously reported to be preferentially expressed in the bundle sheath (using a
454 log2(bundle sheath/whole leaf) cut off >1) (Aubry et al., 2014) indicating that this module
455 must operate alongside other networks. This notion is supported by previous analysis of
456 promoters controlling bundle sheath preferential expression in Arabidopsis. For example the
457 region identified as controlling the bundle sheath preferential expression of *SULTR2;2*
458 (Kirschner et al., 2018) does not contain the MYC-MYB module.

459 Although extensive glucosinolate biosynthesis is confined to the Brassicaceae (Halkier &
460 Gershenzon, 2006) there are indications that this MYC-MYB module patterns genes
461 unrelated to glucosinolates to the bundle sheath. One possibility is associated with
462 glucosinolate biosynthesis representing a derived pathway that has evolved in the
463 Brassicaceae. It would seem more parsimonious if its patterning to bundle sheath cells was
464 mediated by integration into an existing gene regulatory network associated with this cell
465 type, than through evolution of a network *de novo*. This seems plausible because in addition
466 to several enzymes of primary sulphur metabolism being part of the core glucosinolate
467 biosynthesis pathway (Yatusevich et al., 2010), transcripts encoding many enzymes of
468 sulphur transport and assimilation are more abundant in the bundle sheath compared with
469 whole leaves (Aubry et al., 2014). This raises the possibility that compartmentation of
470 glucosinolate biosynthesis to the bundle sheath may have occurred through acquisition of
471 *cis*-elements that restrict the expression of sulphur assimilation genes to this cell-type.
472 In addition to genes associated with sulphur metabolism, there is also evidence that this
473 MYC-MYB module may pattern genes that are thought to represent some of the first steps
474 towards evolving C_4 photosynthesis. Although the Arabidopsis *GLYCINE*
475 *DECARBOXYLASE P-PROTEIN 1 (GLDP1*) gene is expressed strongly in both the
476 mesophyll cells and the vasculature, deletion of an M-box in the promoter resulted in bundle
477 sheath expression of *GLDP1* (Adwy et al., 2015). The remnant expression in the bundle
478 sheath and vein tissue was associated with a 266bp region named the V-box (Adwy et al.,
479 2015). Re-analysis of these 266bp identified MYC and MYB binding sites within 25bp of
480 each other (Supplemental Figure 12). It is possible that constitutive expression of *GLDP1* in
481 the C_3 Arabidopsis leaf is due to the M-box and MYC-MYB modules driving expression in
482 mesophyll and bundle sheath strands respectively. This is consistent with the MYC-MYB
483 module being an activator of expression in the bundle sheath and therefore its presence not
484 preventing activation in other cell types. This would explain why only some genes containing
485 the MYC-MYB module are preferentially expressed in bundle sheath strands and suggests
486 that bundle sheath preferential expression is partly defined by lack of activation in other cell-
487 types.
488 One of the early events associated with the transition from C_3 to C_4 photosynthesis is
489 thought to be the restriction of the Glycine Decarboxylase complex to the bundle sheath as
490 part of establishing a C_2 photosynthetic cycle (Mallmann et al., 2014). The M-box of *GLDP1*
491 is highly conserved in the Brassicaceae, but is lost in *Moricandia nitens*, a species that uses
492 C_2 photosynthesis and partitions *GLDP* to bundle sheath cells. Conversely the V-box, and
493 predicted MYC and MYB binding sites, is conserved in *M. nitens* (Adwy et al., 2015; Adwy

494 et al., 2019). It is therefore possible that during the evolution of C_2 photosynthesis in the
495 Brassicaceae, the MYC-MYB module in *GLDP* is responsible for bundle sheath expression
496 of the *GLDP* gene once the M-box is lost.

497 In summary we report a MYC-MYB module that directs gene expression to the bundle
498 sheath of Arabidopsis. To our knowledge, this provides the first example of a regulatory
499 system governing the spatial control of gene expression in leaves. In the future it will be
500 interesting to determine if this module has been co-opted during the evolution of C_4
501 photosynthesis to pattern components of the C_4 cycle to this cell type.

502 **Materials and methods**

503 **Plant material, growth conditions and cloning**

504 Seed of Arabidopsis was sterilised by washing in 70% (v/v) ethanol for 3 minutes
505 followed by washing in 100% ethanol for 1 minute. Transformants were selected on 0.5%
506 (w/v) Murashige & Skoog medium (pH 5.8) 1% (w/v) agar with the relevant antibiotics. After
507 2-3 days of stratification in the dark at 4°C, tissue culture plates were transferred to a 16
508 hour photoperiod growth chamber with a light intensity of 200 µmol m^{-2} s^{-1} photon flux
509 density, 65% relative humidity and a temperature cycle of 24°C (day) and 20°C (night).
510 Transformed seedlings were transferred onto 3:1 Levington M3 high nutrient compost and
511 Sinclair fine Vermiculite soil mixture and grown for another 2-3 weeks before analysis. *N.*
512 *benthamiana* plants used for transient assays were grown from seed in pots containing the
513 same soil mixture with a 16 hour photoperiod, 200 µmol m^{-2} s^{-1} photon flux density, 60%
514 relative humidity and 22°C.

515 T-DNA insertion mutants for *dreb2a* (GK-179C04) were obtained from the Nottingham
516 Arabidopsis Stock Centre (NASC). T-DNA insertion lines were genotyped to identify lines
517 homozygous for the required T-DNA insertion and RT-PCR was performed to confirm that
518 the mutation resulted in the loss of *DREB2A* gene expression.

519 The full length *MYB76* gene as well as the promoter alone were amplified from Arabidopsis
520 *Col-0* genomic DNA and then fused to *uidA*. The minimal CaMV35S promoter was
521 synthesised and fused to *MYB76 DHS* by polymerase chain reactions (PCR). Deletion of
522 the DHS within the promoter was achieved by PCR fusion of the 5' end of the promoter with
523 the 3' end of the promoter prior to being cloned into the pENTR/D TOPO vector. Each
524 forward primer contained a CACC overhang to ensure directional cloning. A Gateway LR
525 reaction was performed to transfer the relevant inserts into a modified pGWB3 vector
526 (Nakagawa et al., 2007) that contained an intron within the *uidA* sequence.
527 *MYB76gDNA::uidA*, *2xDHS_CaMV35SMin::uidA* and *2x27_CaMV35SMin::uidA* were
528 constructed using Golden Gate technology (Weber et al., 2011). Motif substitutions were
529 made using the QuikChange Lightning Site-Directed Mutagenesis (Agilent Technologies)
530 and motif deletions were made by overlapping PCR. Constructs were then placed into
531 *Agrobacterium tumefaciens* strain GV3101 and introduced into Arabidopsis Col-0 by floral
532 dipping (Clough & Bent, 1998).

533 Constructs for *trans*-activation assays were made using the Golden Gate system. Coding
534 sequence of candidate transcription factors were cloned from Arabidopsis cDNA,
535 domesticated to remove *Bpi*I and *Bsa*I restriction sites and cloned into level 0 vectors. Level
536 1 constructs were generated to constitutively express candidate transcription factors, to

537 constitutively express a p19 silencing suppressor, to constitutively express a GUS reporter
538 to act as an infiltration control and to fuse the *MYB76* DHS with a *LUCIFERASE* reporter to
539 provide an output of activation from the DHS. These level 1 constructs were then assembled
540 into level 2 modules and transformed into *A. tumefaciens* GV3101. Constructs for the
541 constitutively active LjUBI promoter, the OCS1 terminator, LUC coding region, GUS coding
542 region have been published previously (Feike et al., 2019) and parts were cloned into
543 appropriate Golden Gate vectors (Patron et al., 2015).
544
545 **GUS staining, MUG assays, and GFP imaging**
546 To take into account position effects associated with transgene insertion site, GUS
547 staining was undertaken on at least six randomly selected T1 plants for each *uidA* fusion
548 (Jefferson et al., 1987). The staining solution contained 0.1 M Na_2HPO_4 pH 7.0, 2 mM
549 Potassium ferricyanide, 2 mM Potassium ferrocyanide, 10 mM EDTA pH 8.0, 0.06% (v/v)
550 Triton X-100 and 0.5 mg ml^{-1} X-gluc. Leaves from three-week old plants were vacuum-
551 infiltrated three times in GUS solution for one minute and then incubated at 37°C for between
552 3 and 72 hrs depending on the strength of the promoter being assessed. Next, stained
553 samples were fixed in 3:1 (v/v) ethanol:acetic acid for 30 minutes at room temperature,
554 cleared in 70% (v/v) ethanol at 37°C and then placed in 5 M NaOH for 2 hrs. Samples were
555 stored in 70% (v/v) ethanol at 4°C. To quantify reporter accumulation from each promoter
556 the quantitative assay that assesses the rate of MUG conversion to 4-methylumbelliferone
557 (MU) was performed (Jefferson et al., 1987) on between 10 and 25 lines. Tissue was frozen
558 in liquid nitrogen, homogenised and soluble protein extracted in 5 volumes of Protein
559 extraction buffer (1 mM $MgCl_2$, 100 mM NaCl, 50 mM Tris (Melford) pH7.8). 15 µl of protein
560 extract was incubated with 60 µl of MUG at 37 ºC for one, two, three and four hours
561 respectively. The reaction was stopped after each time point by addition of 75 µl 200 mM
562 anhydrous sodium carbonate. GUS activity was analysed via measurements of fluorescence
563 of MU at 455 nm after excitation at 365 nm. The concentration of MU/unit fluorescence in
564 each sample was interpolated using a concentration gradient of MU over a linear range.
565 GFP imaging was performed on at least seven independent T1 lines of
566 *pMYB76::H2B::GFP* and *2xMYB76_DHS::H2B::GFP.* Rosette leaves of four week old plants
567 were sampled and outer tissue layers were removed by scraping leaves with a razor blade
568 under 1X PBS solution. Samples were imaged on a Leica TCS SP8 confocal microscope,
569 GFP was excited at 488 nm and emission was detected at 500-530 nm. Images were
570 recorded on LAS Image analysis software (Leica) and processed to merge channels and
571 add scale bars in ImageJ v1.52a (https://imagej.nih.gov/ij/).

572

573 **Yeast One-Hybrid screen**

574 Regions screened for transcription factor binding via Yeast One-Hybrid were first inserted
575 into pENTR 5'TOPO TA entry vector (Thermofisher) and subsequently placed into the
576 pMW2 and pMW3 destination vectors containing *HIS3* and *LACZ* marker genes respectively
577 (Deplancke et al., 2006). The enhanced Yeast One-Hybrid screen against a complete
578 collection of 2000 Arabidopsis transcription factors was undertaken as described previously
579 (Gaudinier et al., 2011; Pruneda-Paz et al., 2014; Gaudinier et al., 2017). Details of the bait
580 sequence and list of interactors found in Supplemental table 4.

581

582 ***Trans* activation assays and qRT-PCR**

583 To test interactions between the *MYB76* DHS and transcription factors *in planta* transient
584 infiltration of *N. benthamiana* was performed. Overnight cultures of *A. tumefaciens* were
585 pelleted and re-suspended in infiltration buffer (10mM MES (pH5.6), 10mM $MgCl_2$, 150μM
586 acetosyringone) to an optical density of 0.3. Cultures were then incubated for 2hrs at room
587 temperature and infiltrated into the abaxial side of leaves of four-week old plants with a 1 ml
588 syringe.

589 Leaf discs from infiltrated regions were sampled 48hrs after infiltration and flash frozen in
590 liquid N_2. Protein for MUG and LUC assays was extracted on ice in 1x passive lysis buffer
591 (PLB: Promega). MUG assays were performed by adding 40μl of protein extract to 100μl of
592 MUG assay buffer (2mM MUG, 50mM Na_3PO_4/Na_2PO_4 buffer (pH 7.0), 10 mM EDTA, 0.1%
593 (v/v) Triton X-1000, 0.1% (w/v) Sodium Lauroyl sarcosinate and 10 mM DTT). Stop buffer
594 (200 mM Na_2CO_3) was added at 0 and 30 mins and rate of MUG accumulation was
595 measured in triplicate on a plate reader (CLARIOstar, BMG lab tech) with excitation at 360
596 nm and emission at 465 nm. LUC activity was measured with 20 μl of protein sample and
597 100 μl of LUC assay reagent (Promega). Activation from the DHS was calculated as (LUC
598 luminescence/rate of MUG accumulation) * 100.

599 Single rosette leaves from four-week old col-0, *dreb2a* and *myc2/3/4* plants were sampled
600 six hours after the onset of light and flash frozen in liquid N_2. RNA was extracted with the
601 RNeasy plant mini-kit (Qiagen) as manufacturer's instructions and cDNA was synthesised
602 using the Superscript double stranded cDNA synthesis kit as manufacturer's instructions
603 (Invitrogen) with on-column DNase1 treatment. qRT-PCR was performed using SYBR green
604 master mix (Bio-Rad) on a CFX384 touch Real-Time PCR Detection System (Bio-RAD).
605 Transcripts of *MYB76* were normalised to the expression of *ASCORBATE PEROXIDASE 3*
606 *(APX3: At4g35000), ASPARTIC PROTEINASE A1 (APA1: At1g11910)* and *UBIQUITIN*

607 *CONJUGATING ENZYME 21 (UBC21: AT5G25760).* Relative expression was determined
608 using the single delta Ct method and the data reported are from normalisation against *APA1.*
609 Results were very similar regardless of the reference genes used.
610
611 **Datasets, GO term analysis and *de novo* motif identification**
612 Computational analysis used previously published Arabidopsis datasets for bundle
613 sheath and whole leaf translatomes (Aubry et al., 2014), *myc2/3/4* mutants and col-0 (Major
614 et al., 2017), and *myb28/29* mutants and col-0 (Burow et al., 2015). For Gene Ontology (GO)
615 term enrichment analysis the top 200 most bundle sheath preferential (log2(bundle sheath
616 expression/35S expression) genes in the cell-type specific translatome (Aubry et al., 2014)
617 were used as input. The AgriGO tool (Du et al., 2010) was used with default parameters and
618 all genes annotated in the Arabidopsis TAIR10 genome were sued as background. The
619 MEME tool from The Multiple Em for Motif Elucidation (MEME) suite v.4.8.1. (Bailey et al.,
620 2009) was applied to search for conserved motifs within promoter sequences of genes
621 expressed in the Arabidopsis bundle sheath. Maximum length of the motif was set to eight
622 nucleotides, both strands of the sequence were searched and each motif had to be present
623 in every sequence.
624 To cluster transcription factor binding motifs the RSAT matrix-clustering tool (Castro-
625 Mondrago et al., 2017) was run on all Arabidopsis motifs from the JASPAR motif database
626 (Fornes et al., 2019) using default parameters which generated 43 motif clusters. The Find
627 Individual Motif Occurances (FIMO) (Grant et al., 2011) was used to scan DNA sequences
628 for matches to Arabidopsis transcription factor binding motifs found in the JASPAR motif
629 database (Fornes et al., 2019). To account for input sequence composition a background
630 model was generated using the fasta-get-markov tool from the MEME suite (Bailey et al.,
631 2009). FIMO was then run with default parameters and a p-value cut-off of 1e-04.
632 Motif enrichment in promoters of gene sets was analysed using a custom BASH script.
633 Promoter regions (1500bp) were extracted using the getfasta tool from Bedtools (Quinlan
634 and Hall, 2010). These promoters were scanned for transcription factor binding motifs using
635 FIMO (as above) and counts of motifs in gene sets were recorded. Frequency of a given
636 motif in a gene set was calculated as a proportion of the total motifs and enrichment was
637 calculated as frequency vs background frequency. Background frequency was defined as
638 mean motif frequency in promoters of three random sets of 2000 Arabidopsis genes. Results
639 of motif frequency analysis presented as the log2 of enrichment and motifs sorted by motif
640 cluster. FIMO outputs were sorted to only include matches to cluster 8 and 18 motifs and a

641 custom Python script was used to find the minimum distance between the centres of cluster
642 8 and 18 motif in the same promoter.
643
644 **Statistical analysis**
645 For statistical analysis extreme outliers were identified and removed from analysis.
646 Normality of the data was assessed using the Shapiro-Wilks test. Where data were normally
647 distributed pairwise T-tests were used to assess significance. Where data were not normally
648 distributed, Wilcoxon rank-sum tests were used to assess significance. Levene's test was
649 used to assess equality of variance. Where variance was equal standard deviations were
650 pooled, where variance was not equal variance was not pooled. All tests were two-sided.
651 Pairwise T-tests with pooled SD were used to assess significance in *trans*-activation assays
652 and without pooled SD in qRT-PCR assays. Wilcoxon rank sum tests were used to assess
653 significance for differences in distributions of minimum differences between cluster 8 and 18
654 motifs in different gene sets. All statistical analysis was performed using R (RStudio Team,
655 2015) and plots generated using ggplot2 (Wickham, 2009).
656
657 **Code availability**
658 All code associated with this manuscript is available in the Github repository:
659 https://github.com/hibberdlab/Knerova_Dickinson_Arabidopsis_bundle_sheath_bipartite_
660 module.
661
662 **Data availability**
663 Underlying data required to generate plots are available in the Github repository:
664 https://github.com/hibberdlab/Knerova_Dickinson_Arabidopsis_bundle_sheath_bipartite_
665 module. All other data available on request.

666 **References**

667 Adwy, W., Laxa, M. & Peterhansel, C. A simple mechanism for the establishment of C_2-
668 specific gene expression in Brassicaceae. *Plant Journal* **84**, 1231–1238 (2015).
669 Adwy, W. et al. Loss of the M-box from the glycine decarboxylase P-subunit promoter in C_2
670 Moricandia species. *Plant Gene* **18**, 100176 (2019)
671 Agarwal, P.K. et al. Dehydration responsive element binding transcription factors and their
672 applications for the engineering of stress tolerance. *Journal of Experimental Botany* **68**,
673 2135–2148 (2017).
674 Akyildiz, M. et al. Evolution and function of a cis-regulatory module for mesophyll-specific
675 gene expression in the C_4 dicot Flaveria trinervia. *The Plant Cell* **19**, 3391–3402 (2007).
676 Ali, S. & Taylor, W.C. The 3 ' non-coding region of a C_4 photosynthesis gene increases
677 transgene expression when combined with heterologous promoters. *Plant Molecular Biology*
678 **46**, 325–333 (2001).
679 Aubry, S. et al. Transcript residency on ribosomes reveals a key role for the Arabidopsis
680 thaliana bundle sheath in sulphur and glucosinolate metabolism. *The Plant journal* **78**, 659–
681 673 (2014).
682 Bailey, T.L. et al. MEME Suite: tools for motif discovery and searching. *Nucleic Acids*
683 *Research* **37**, W202–W208 (2009).
684 Baima, S. et al. The expression of the Athb-8 homeobox gene is restricted to provascular
685 cells in Arabidopsis thaliana. *Development* **121**, 4171–4182 (1995).
686 Bhullar, S. et al. Strategies for Development of Functionally Equivalent Promoters with
687 Minimum Sequence Homology for Transgene Expression in Plants: *cis*-Elements in a Novel
688 DNA Context versus Domain Swapping. *Plant Physiology* **132**, 988-998 (2003).
689 Bonke, M. et. al. APL regulates vascular tissue identity in Arabidopsis. *Nature* **426**, 181–186
690 (2003).
691 Brown, N.J. et al. Independent and Parallel Recruitment of Preexisting Mechanisms
692 Underlying C_4 Photosynthesis. *Science* **331**, 1436–1439 (2011).
693 Burow, M. et al. The Glucosinolate Biosynthetic Gene AOP2 Mediates Feed-back
694 Regulation of Jasmonic Acid Signaling in Arabidopsis. *Molecular Plant* **8**, 1201–1212 (2015).
695 Von Caemmerer, S., Quick, W.P. & Furbank, R.T. The development of C_4 rice: Current
696 progress and future challenges. *Science* **336** 671–1672 (2012).
697 Castro-Mondragon J.A. et al. RSAT matrix-clustering: Dynamic exploration and re-
698 dundancy reduction of transcription factor binding motif collections. *Nucleic Acids Res* **45**,
699 e119 (2017).

700 Clough, S.J. & Bent, A.F. Floral dip: a simplified method for Agrobacterium-mediated
701 transformation of Arabidopsis thaliana. *Plant Journal* **16**, 735–743 (1998).
702 Deplancke, B. et al. A Gene-Centered C. elegans Protein-DNA Interaction Network. *Cell*
703 **125**, 1193–1205 (2006).
704 Dey, N. et al. Synthetic promoters in planta. *Planta* **242**, 1077–1094 (2015).
705 Du, Z. et al. agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res*
706 **38**, W64–W70 (2010).
707 Egawa, C. et al. Differential regulation of transcript accumulation and alternative splicing of
708 a *DREB2* homolog under abiotic stress conditions in common wheat. *Genes & Genetic*
709 *Systems* **81**, 77–91 (2006).
710 Engelmann, S. et al. The gene for the P-subunit of glycine decarboxylase from the C_4
711 species *Flaveria trinervia*: analysis of transcriptional control in transgenic *Flaveria bidentis*
712 (C_4) and Arabidopsis (C_3). *Plant Physiology* **146**, 1773–1785 (2008).
713 Feike, D. et al. Characterizing standard genetic parts and establishing common principles
714 for engineering legume and cereal roots. *Plant Biotechnology Journal* **17**, 2234–2245
715 (2019).
716 Fernández-Calvo, P. et al. The Arabidopsis bHLH transcription factors MYC3 and MYC4 are
717 targets of JAZ repressors and act additively with MYC2 in the activation of jasmonate
718 responses. *The Plant Cell* **23**, 701-715 (2011).
719 Fletcher, J.C. et al. Signaling of Cell Fate Decisions by CLAVATA3 in Arabidopsis Shoot
720 Meristems. *Science* **19**, 1911–1914 (1999).
721 Fornes, O. et al. JASPAR 2020: update of the open-access database of transcription factor
722 binding profiles. *Nucleic Acids Res*., doi: 10.1093/nar/gkz1001 (2019).
723 Fryer, M.J. et al. Control of Ascorbate Peroxidase 2 expression by hydrogen peroxide and
724 leaf water status during excess light stress reveals a functional organisation of Arabidopsis
725 leaves. *Plant Journal* **33**, 691–705 (2003).
726 Gallegos, J.E. & Rose, A.B. Intron DNA Sequences Can Be More Important Than the
727 Proximal Promoter in Determining the Site of Transcript Initiation. *The Plant Cell* **29**, 843–
728 853 (2017).
729 Gaudinier, A. et al. Enhanced Y1H assays for Arabidopsis. *Nature methods* **8**, 1053 - 1055
730 (2011).
731 Gaudinier, A. et al. Identification of Protein–DNA Interactions Using Enhanced Yeast One-
732 Hybrid Assays and a Semiautomated Approach. *Methods Mol Biol* **1610**, 187-215 (2017).

733 Gigolashvili, T. et al. HAG2/MYB76 and HAG3/MYB29 exert a specific and coordinated
734 control on the regulation of aliphatic glucosinolate biosynthesis in Arabidopsis thaliana. *New*
735 *Phytologist* **177**, 627–642 (2008).
736 Gowik, U. et al. *cis*-Regulatory elements for mesophyll-specific gene expression in the C_4
737 plant *Flaveria trinervia*, the promoter of the C_4 phospho *enol*pyruvate carboxylase gene. *The*
738 *Plant Cell* **16**, 1077–1090 (2004).
739 Grant, C.E., Bailey, T.L. & Noble, W.S. FIMO: scanning for occurrences of a given motif.
740 *Bioinformatics* **27**, 1017–1018 (2011).
741 Griffiths, H. et al. You're so vein: bundle sheath physiology, phylogeny and evolution in C_3
742 and C_4 plants. *Plant Cell Environ* **36**, 249–261 (2013).
743 Halkier, B.A. & Gershenzon, J. Biology and Biochemistry of Glucosinolates. *Annual Review*
744 *of Plant Biology* **57**, 303–333 (2006).
745 Heidstra, R., Welch, D. & Scheres, B. Mosaic analyses using marked activation and deletion
746 clones dissect Arabidopsis SCARECROW action in asymmetric cell division. *Genes and*
747 *Development* **18**, 1964–1969 (2004).
748 Helariutta, Y. et al. The SHORT-ROOT Gene Controls Radial Patterning of the Arabidopsis
749 Root through Radial Signaling. *Cell* **101**, 555–567 (2000).
750 Hernandez-Garcia, C.M. & Finer, J.J. Identification and validation of promoters and *cis*-
751 acting regulatory elements. *Plant Science* **217–218**, 109–119 (2014).
752 Hesselberth, J.R. et al. Global mapping of protein-DNA interactions in vivo by digital
753 genomic footprinting. *Nature methods* **6**, 283–9 (2009).
754 Hibberd, J.M. & Covshoff, S. The Regulation of Gene Expression Required for C_4
755 Photosynthesis. *Annu. Rev. Plant Biol* **61**, 181–207 (2010).
756 Hibberd, J.M., Sheehy, J.E. & Langdale, J.A. Using C_4 photosynthesis to increase the yield
757 of rice - rationale and feasibility. *Current Opinion in Plant Biology* **11**, 228–231 (2008).
758 Howe, G.A., Major, I.T. & Koo, A.J. Modularity in jasmonate signaling for multistress
759 resilience. *Annu. Rev. Plant Bio*l. **69**, 387–415 (2018).
760 Imlau, A., Truernit, E. & Sauer, N. Cell-to-Cell and Long-Distance Trafficking of the Green
761 Fluorescent Protein in the Phloem and Symplastic Unloading of the Protein into Sink
762 Tissues. *The Plant Cell* **11**, 309–322 (1999).
763 Jefferson, R.A., Kavanagh, T.A. & Bevan, M.W. GUS fusions: ß-Glucuronidase as a
764 sensitive and versatile gene fusion marker in higher plants. *EMBO Journal* **6**, 3901–3907
765 (1987).
766 Kajala, K. et al. Multiple Arabidopsis genes primed for direct recruitment into C_4
767 photosynthesis. *Plant Journal* **69**, 47–56 (2012).

768 Kinsman, E.A. & Pyke, K.A. Bundle sheath cells and cell-specific plastid development in
769 Arabidopsis leaves. *Development* **125**, 1815–1822 (1998).
770 Kirschner, S. et al. Expression of SULTR2;2, encoding a low-affinity sulphur transporter, in
771 the Arabidopsis bundle sheath and vein cells is mediated by a positive regulator. *Journal of*
772 *Experimental Botany* **69**, 4897-4906 (2018).
773 Koroleva, O. et al. Patterns of solute in individual mesophyll, bundle sheath and epidermal
774 cells of barley leaves induced to accumulate carbohydrate. *New Phytol*. **136**, 97–104 (1997).
775 Leegood, R.C. Roles of the bundle sheath cells in leaves of C_3 plants. *Journal of*
776 *Experimental Botany*, **59**, pp.1663–1673 (2008).
777 Levin, J.Z. & Meyerowitz, E.M. UFO - an Arabidopsis Gene Involved In Both Floral Meristem
778 and Floral Organ Development. *Plant Cell* **7**, 529–548 (1995).
779 Li, B. et al. Promoter-Based Integration in Plant Defense Regulation. *Plant Physiology* **166**,
780 1803–1820 (2014).
781 Li, B. et al. Network-Guided Discovery of Extensive Epistasis between Transcription Factors
782 Involved in Aliphatic Glucosinolate Biosynthesis. *The Plant Cell* **30**, 178–195 (2018).
783 Liu, Q. et al. Two Transcription Factors, DREB1 and DREB2, with an EREBP/AP2 DNA
784 Binding Domain Separate Two Cellular Signal Transduction Pathways in Drought- and Low-
785 Temperature-Responsive Gene Expression, Respectively, in Arabidopsis. *The Plant Cell*
786 **10**, 1391-406. (1998)
787 Major, I.T. et al. Regulation of growth–defense balance by the JASMONATE ZIM-DOMAIN
788 (JAZ)-MYC transcriptional module. *New Phytologist* **215**, 1533–1547 (2017).
789 Malitsky, S. et al. The Transcript and Metabolite Networks Affected by the Two Clades of
790 Arabidopsis Glucosinolate Biosynthesis Regulators. *Plant Physiology* **148**, 2021-2049
791 (2008).
792 Mallmann, J. et al. The role of photorespiration during the evolution of C_4 photosynthesis in
793 the genus Flaveria. *eLife* **3**, e02478 (2014).
794 Masucci, J.D. et al. The homeobox gene GLABRA2 is required for position-dependent cell
795 differentiation in the root epidermis of Arabidopsis thaliana. *Development* **122**, 1253–1260
796 (1996).
797 Matsukura, S. et al. Comprehensive analysis of rice DREB2-type genes that encode
798 transcription factors involved in the expression of abiotic stress-responsive genes. *Molecular*
799 *Genetics and Genomics* **283**, 185–196 (2010).
800 Matsuoka, M. et al. The promoters of two carboxylases in a C_4 plant (maize) direct cell-
801 specific, light-regulated expression in a C_3 plant (rice). *The Plant Journal* **6**, 311–319 (1994).

802 Mustroph, A. et al. Profiling translatomes of discrete cell populations resolves altered cellular
803 priorities during hypoxia in Arabidopsis. *Proceedings of the National Academy of Sciences*
804 **106**, 18843–18848 (2009).
805 Nakagawa, T. et al. Development of series of gateway binary vectors, pGWBs, for realizing
806 efficient construction of fusion genes for plant transformation. *J Biosci Bioeng* **104**, 34–41
807 (2007).
808 Nakamura, R.L. et al. Expression of an Arabidopsis Potassium Channel Gene in Guard
809 Cells. *Plant Physiology* **109**, 371–374 (1995).
810 Neph, S. et al. An expansive human regulatory lexicon encoded in transcription factor
811 footprints. *Nature* **489**, 83–90 (2012).
812 Otsuga, D. et al. REVOLUTA regulates meristem initiation at lateral positions. *The Plant*
813 *Journal* **25**, 223–236 (2001).
814 Patron, N.J. et al. Standards for plant synthetic biology: a common syntax for exchange of
815 DNA parts. *New Phytologist* **208**, 13–19 (2015).
816 Pruneda-Paz, J.L. et al. A Genome-Scale Resource for the Functional Characterization of
817 Arabidopsis Transcription Factors. *Cell Reports* **8**, 622–632 (2014).
818 Qin, F. et al. Regulation and functional analysis of ZmDREB2A in response to drought and
819 heat stresses in Zea mays L. *The Plant Journal* **50**, 54–69 (2007).
820 Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic
821 features. *Bioinformatics* **26**, 841–842 (2010).
822 RStudio Team. RStudio: Integrated Development Environment for R
823 http://www.rstudio.com/ (2015).
824 Reyna-Llorens, I. & Hibberd, J.M. Recruitment of pre-existing networks during the evolution
825 of C_4 photosynthesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*
826 **372**, 2–7 (2017).
827 Reyna-Llorens, I. et al. Ancient duons may underpin spatial patterning of gene expression
828 in C_4 leaves. *Proc. Natl Acad. Sci. USA* **115**, 1931–1936 (2018).
829 Robinson, J.T. et al. Integrative genomics viewer. *Nature Biotechnology* **29**, 24 – 26 (2011).
830 Ruzicka, D.R. et al. The ancient subclasses of Arabidopsis ACTIN DEPOLYMERIZING
831 FACTOR genes exhibit novel and differential expression. *The Plant Journal* **52**, 460–472
832 (2007).
833 Sage, R. Environmental and evolutionary preconditions for the origin and diversification of
834 the C_4 photosynthetic syndrome. *Plant Biology* **3**, 202–213 (2001).
835 Sakuma, Y. et al. Functional Analysis of an Arabidopsis Transcription Factor, DREB2A,
836 Involved in Drought-Responsive Gene Expression. *The Plant Cell* **18**, 1292–1309 (2006).

837 Salehin, M. et al. Auxin-sensitive Aux/IAA proteins mediate drought tolerance in Arabidopsis
838 by regulating glucosinolate levels. *Nature Communications* **10,** 4021 (2019).
839 Sarkar, A.K. et al. Conserved factors regulate signalling in Arabidopsis thaliana shoot and
840 root stem cell organizers. *Nature* **446**, 811–814 (2007).
841 Sarsby, J. et al. Mass spectrometry imaging of glucosinolates in Arabidopsis flowers and
842 siliques. *Phytochemistry* **77**, 110–118 (2012).
843 Schweizer, F. et al.. Arabidopsis basic helix-loop-helix transcription factors MYC2, MYC3,
844 and MYC4 regulate glucosinolate biosynthesis, insect performance, and feeding behavior.
845 *Plant Cell* **25**, 3117–3132 (2013).
846 Shatil-Cohen, A., Attia, Z. & Moshelion, M. Bundle-sheath cell regulation of xylem-mesophyll
847 water transport via aquaporins under drought stress: a target of xylem-borne ABA? *The*
848 *Plant Journal* **67**, 72–80 (2011).
849 Sparks, E.E. et al. Establishment of expression in the SHORTROOT-SCARECROW
850 transcriptional cascade through opposing activities of both activators and repressors.
851 *Developmental Cell* **39**, 585–596 (2016).
852 Sønderby I.E. et al. A systems biology approach identifies a R2R3MYB gene subfamily with
853 distinct and overlapping functions in regulation of aliphatic glucosinolates. *PLoS ONE* **2**,
854 e1322 (2007).
855 Sønderby I.E. et al. A complex interplay of three R2R3 MYB transcription factors determines
856 the profile of aliphatic glucosinolates in Arabidopsis. *Plant Physiology* **153**, 348–363 (2010).
857 Takahashi, H. et al. The roles of three functional sulphate transporters involved in uptake
858 and translocation of sulphate in Arabidopsis thaliana. *Plant Journal* **23**, 171–182 (2000).
859 Theißen, G., Melzer, R. & Ruümpler, F. MADS-domain transcription factors and the floral
860 quartet model of flower development: Linking plant development and evolution.
861 *Development* **143**, 3259–3271 (2016).
862 Thoma, S. et al. Tissue-specific expression of a gene encoding a cell wall-localized lipid
863 transfer protein from Arabidopsis. *Plant Physiology* **105**, 35–45 (1994).
864 Vainonen, J.P. et al. RCD1–DREB2A interaction in leaf senescence and stress responses
865 in Arabidopsis *thaliana. Biochemical Journal* **442**, 573-581 (2012).
866 Weber, E. et al. A Modular Cloning System for Standardized Assembly of Multigene
867 Constructs. *PLOS ONE* **6**, p.e16765 (2011).
868 Wickham, H. ggplot2: Elegant Graphics for Data Analysis. http://ggplot2.org/book/ (Springer,
869 2009).
870 Williams, B.P. et al. An untranslated *cis*-element regulates the accumulation of multiple C_4
871 enzymes in *Gynandropsis gynandra* mesophyll cells. *The Plant Cell* **28**, 454–465 (2016).

Williams, M. et al. Visualizing the distribution of elements within barley leaves by energy dispersive X-ray image maps (EDX maps). New Phytol. **125**: 367–372 (2018).

Wiludda, C. et al. Regulation of the photorespiratory *GLDPA* gene in C_4 *Flaveria*: an intricate interplay of transcriptional and posttranscriptional processes. *Plant Cell* **24**, 137–151 (2012).

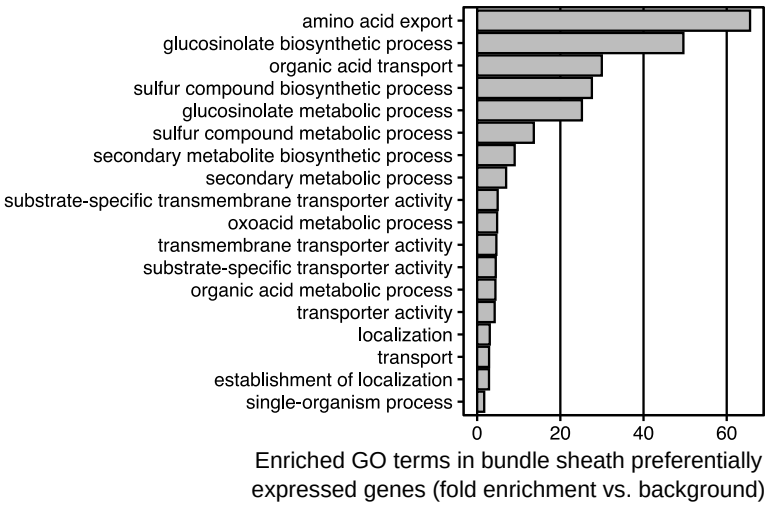
Wysocka-Diller, J.W. et al., Molecular analysis of SCARECROW function reveals a radial patterning mechanism common to root and shoot. *Development* **127**, 595–603 (2000).

Yatusevich, R. et al. Genes of primary sulfate assimilation are part of the glucosinolate biosynthetic network in Arabidopsis thaliana. *Plant Journal* **62**, 1–11 (2010).

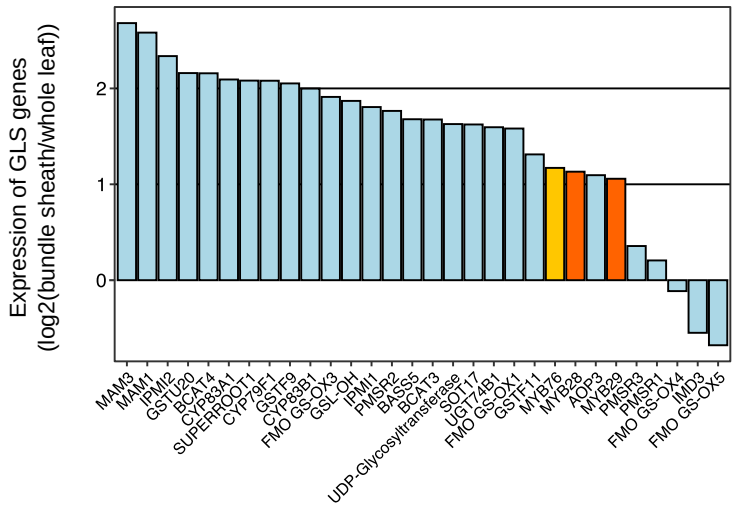
Zhang, W. et al. Genome-Wide Identification of Regulatory DNA Elements and Protein-Binding Footprints Using Signatures of Open Chromatin in Arabidopsis. *The Plant Cell* **24**, 2719–2731 (2012).

Zhang, T., Marand, A.P. & Jiang, J. PlantDHS: a databse for DNase1 hypersensitive sites in plants. *Nucleic Acids Research* **44**, D1148–D1153 (2016).

a


amino acid export
glucosinolate biosynthetic process
organic acid transport
sulfur compound biosynthetic process
glucosinolate metabolic process
sulfur compound metabolic process
secondary metabolite biosynthetic process
secondary metabolic process
substrate-specific transmembrane transporter activity
oxoacid metabolic process
transmembrane transporter activity
substrate-specific transporter activity
organic acid metabolic process
transporter activity
localization
transport
establishment of localization
single-organism process
0
20
40
60
Enriched GO terms in bundle sheath preferentially expressed genes (fold enrichment vs. background)

b


Expression of GLS genes (log2(bundle sheath/whole leaf))
2
1
0
MAM3
MAM1
IPMI2
GSTU20
BCAT4
CYP83A1
SUPERROOT1
CYP79F1
GSTF9
CYP83B1
FMO GS-OX3
GSL-OH
IPMI1
PMSR2
BASS5
BCAT3
UDP-Glycosyltransferase
SOT17
UGT74B1
FMO GS-OX1
GSTF11
MYB76
MYB28
AOP3
MYB29
PMSR3
PMSR1
FMO GS-OX4
IMD3
FMO GS-OX5

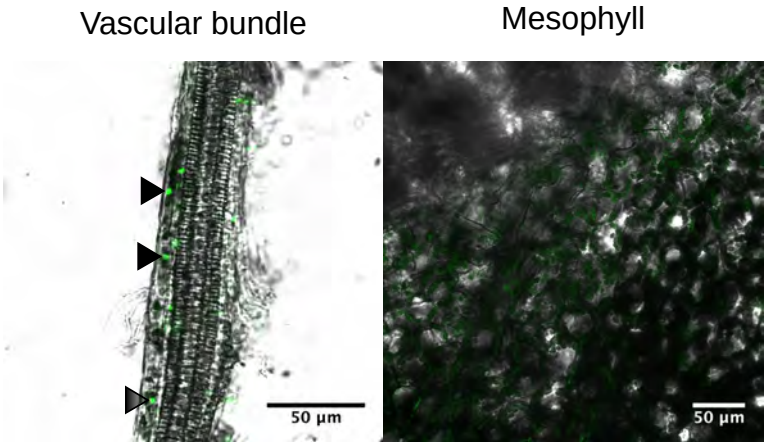
c

-1725
ATG
+279
30hrs

d

-1725
ATG
+1254
48hrs

e


Vascular bundle
Mesophyll
50 µm
50 µm

F

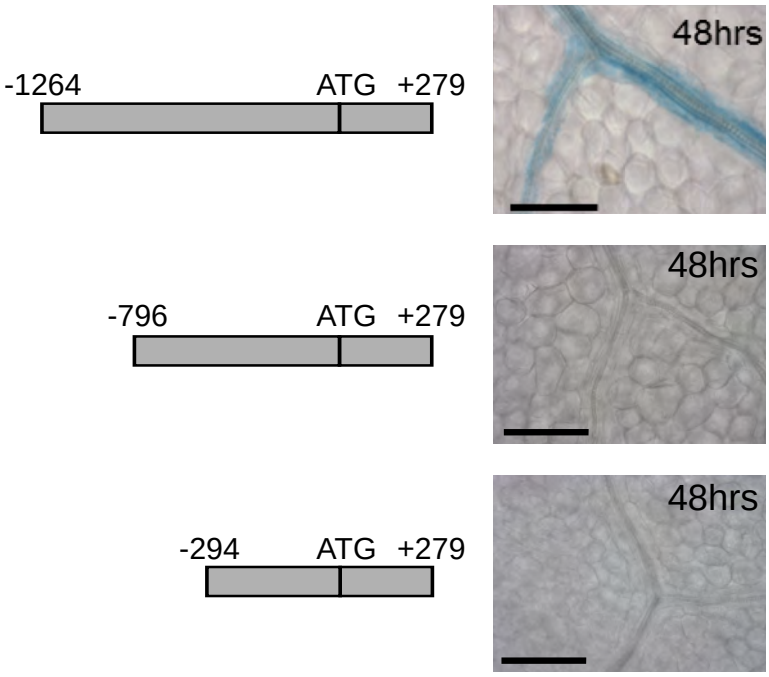
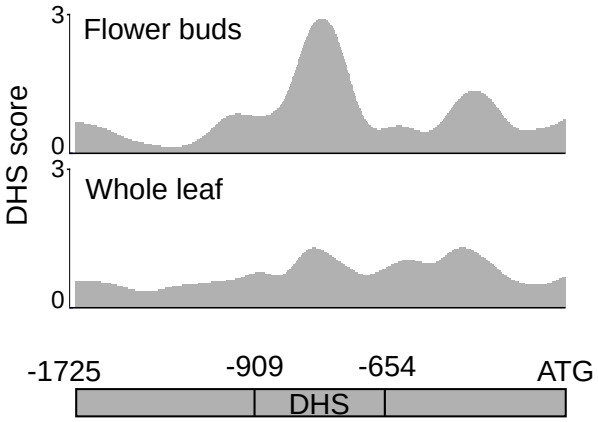
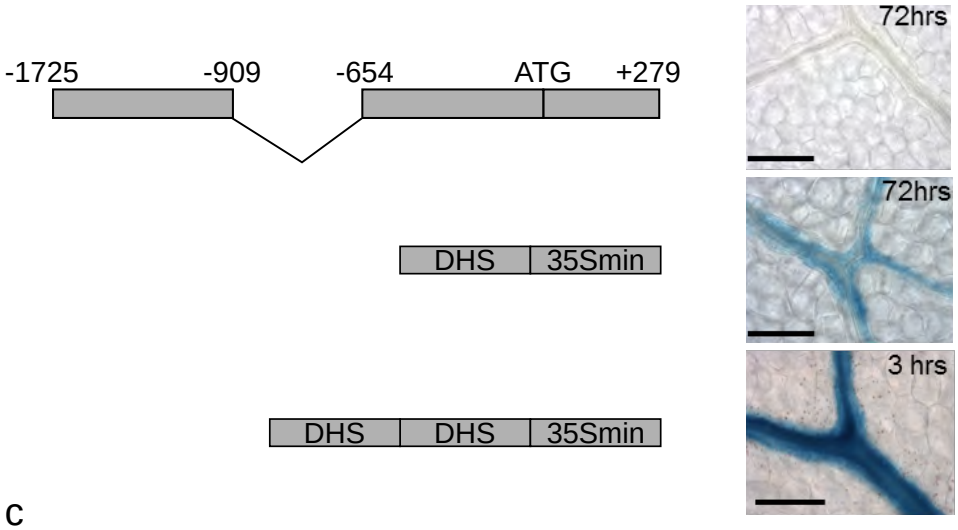

-1264
ATG
+279
48hrs
-796
ATG
+279
48hrs
-294
ATG
+279
48hrs

Figure 1: A 468bp region from the *MYB76* promoter necessary for bundle-sheath expression. A) Gene Ontology term enrichment of the 200 most bundle sheath preferential genes in Arabidopsis. Enrichment shown as fold enrichment compared with background. B) Expression of glucosinolate biosynthesis genes in bundle sheath compared with whole leaves. *MYB76* marked in gold, *MYB28* and *MYB29* marked in orange. C) Schematic and representative image of the *MYB76* promoter plus 279bp of genomic sequence fused to GUS. D) Schematic and representative image of the *MYB76* promoter and full genomic sequence fused to GUS. For C and D staining times are given in the top right corner of each image and scale bars represent 100µm. E) Representative images of the *MYB76* promoter driving expression of a histone GFP fusion. Images of the vascular bundle (left) and mesophyll (right). Black arrowheads indicate nuclei expressing GFP. F) Schematics and representative GUS images of each *MYB76* deletion. Staining times are given in the top right corner of each image. Scale bars represent 100µm.

a



b



c

Motif 1

Motif 2

27bp region CCGAAAAACGTGCACCGTCCAAGATTC

d

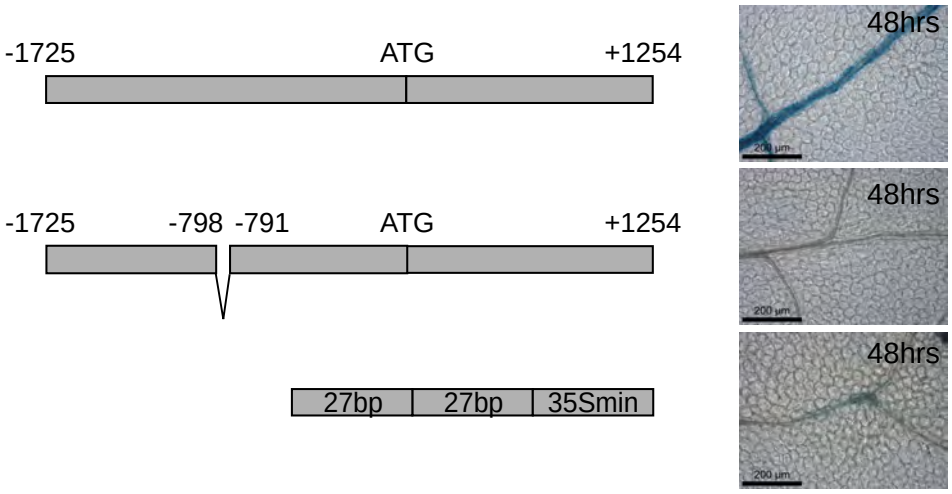
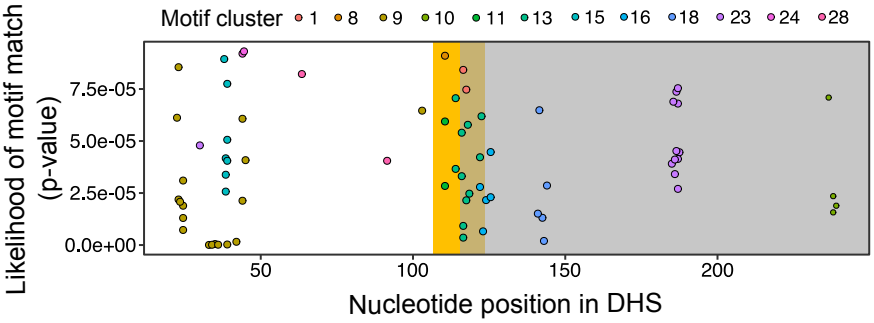


Figure 2: A DNaseI Hypersensitive Site in the *MYB76* promoter is necessary and sufficient for expression in the bundle-sheath. A) The *MYB76* promoter contains a DNaseI Hypersensitive Site (DHS) located between nucleotides -909 to -654. The y-axis shows the DHS score (Zhang et al., 2016) from flower buds (top) and leaves (bottom). Data are from Zhang et al., (2012) visualised with the IGV browser (Robinson et al., 2011). B) Schematics and representative GUS staining images of the *MYB76* promoter with the DHS deleted, the DHS fused to the minimal *CaMV35S* promoter and two copies of the DHS fused to the minimal *CaMV35S* promoter. Staining times are given in the top right corner of each image. Scale bars represent 100µm. C) Position Weight Matrices (PWMs) of two motifs (1 and 2) found in the *MYB76* DHS as well as other promoters driving bundle sheath expression in Arabidopsis and a 27bp region of the *MYB76* DHS containing motif 2 (TGCACCG and highlighted in colours matching the PWM). D) Schematics and representative images of *MYB76* promoter and gDNA (top), *MYB76* promoter and genomic DNA with the *TGCACCG* motif deleted (middle), and oligomerisation of the 27bp region containing TGCACCG (bottom) fused to GUS. Staining times are given in the top right corner of each image. Scale bars represent 200µm.

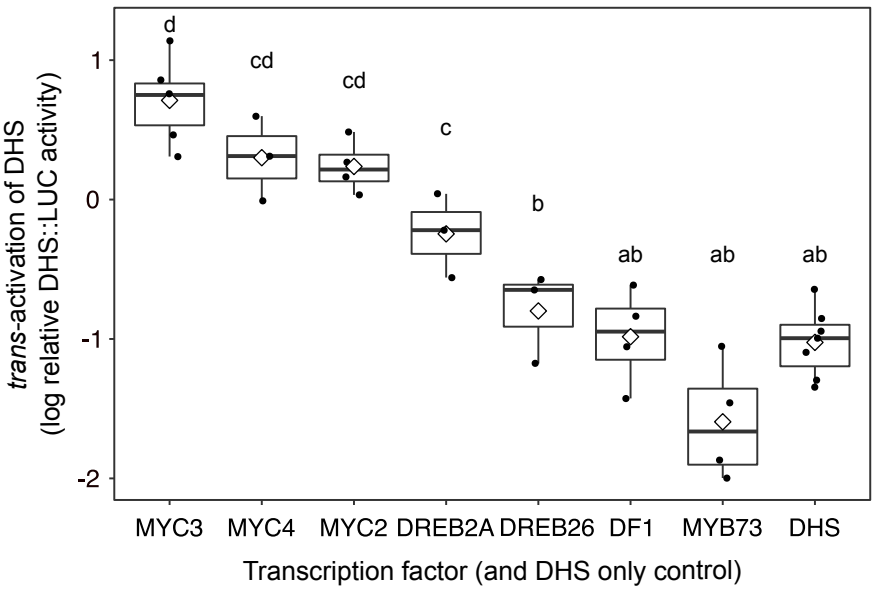
a



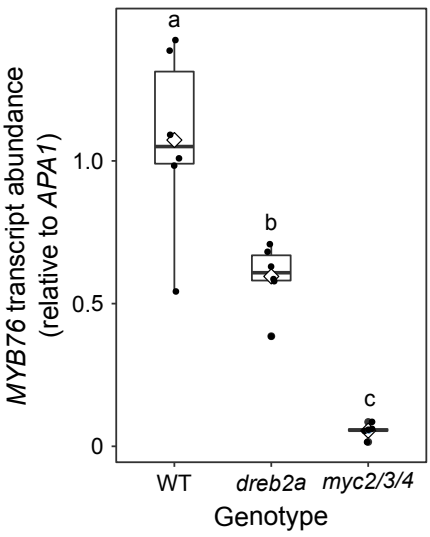
b

Gene id	Name	Family	log2(BS/35S)	Y1H this study	Y1H (Li et al., 2014)	Previously associated with *MYB76*
AT1G32640	MYC2	bHLH		N	N	Y
AT5G46760	MYC3	bHLH		N	N	Y
AT4G17880	MYC4	bHLH		N	N	Y
AT5G61420	MYB28	MYB		N	N	Y
AT5G07690	MYB29	MYB		N	N	Y
AT5G07700	MYB76	MYB		N	N	Y
AT1G21910	DREB26	AP2-EREBP		Y	Y	N
AT1G54060	ASIL1	Trihelix		Y	Y	N
AT1G76880	DF1	Trihelix		Y	Y	N
AT2G22840	GRF1	GRF		Y	Y	N
AT4G37260	MYB73	MYB		Y	Y	N
AT5G52020	RAP2.10	AP2-EREBP		Y	Y	N
AT5G63790	NAC102	NAC		Y	Y	N
AT1G43700	VIP1	bZIP		Y	N	N
AT2G31230	ERF15	AP2-EREBP		Y	N	N
AT2G46270	GBF3	bZIP		Y	N	N
AT3G61830	ARF18	ARF		Y	N	N
AT5G05410	DREB2A	AP2-EREBP		Y	N	N
AT5G57390	AIL5	AP2-EREBP		Y	N	N

c



d



e

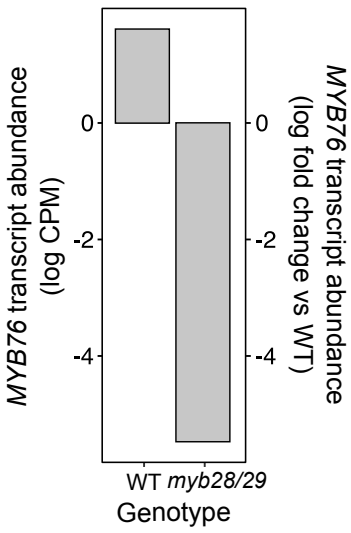


Figure 3: MYC, MYB and DREB transcription factors control *MYB76* expression from the DHS. A) Transcription factor binding motifs within the *MYB76* DHS. Position in the DHS (bp) is on the x-axis, and predicted binding affinity (p-values) on the y-axis. Motifs are coloured by motif cluster (Supplemental table 1). The gold region represents the 27bp region necessary for expression and the grey region indicates sequence unable to generate bundle sheath expression (Figure 1F). B) Summary of candidate transcription factors binding to the *MYB76* DHS. Information provided includes gene identifier, gene name, family, expression in bundle sheath compared with whole leaves (from Aubry et al., 2014), whether they interacted with the DHS in Yeast One-Hybrid, whether they were previously identified as binding the entire *MYB76* promoter (from Li et al., 2014), and if they have previously been associated with controlling *MYB76* expression (Gigolashvili et al., 2008; Schweizer et al., 2013). C) *Trans*-activation assays of candidate transcription factors and the *MYB76* DHS. Values shown represent the log of LUCIFERASE (LUC) signal driven by *MYB76DHS::LUC* normalised to a constitutively expressed GUS infiltration control. Box-plots show inter-quartile range as upper and lower confines of the box, median as a solid black line, mean as a white diamond and whiskers as maximum and minimum values excluding outliers. All individual data points are plotted. a, b, c and d represent statistically significant differences ($p<0.05$) as determined by pairwise T-tests. n = independent biological samples with n=3 for DREB2A, DREB26 and MYC4, n=4 for MYC2, DF1 and MYB73, n=6 for MYC3 and n=7 for DHS. D) qRT-PCR of *MYB76* in WT, *dreb2a* and *myc2/3/4*. Expression shown relative to *APA1,* n=6 independent biological samples for all genotypes. Box-plots show inter-quartile range as upper and lower confines of the box, median as a solid black line, mean as a white diamond and whiskers as maximum and minimum values excluding outliers. All individual data points are plotted. All individual data points are plotted. a, b and c represent statistically significantly differences ($p<0.001$) determined by pairwise T-tests. E) *MYB76* expression from a publically available *myb28/29* transcriptome (Burow et al., 2015). Expression in WT (left) shown as log counts per million. Expression in *myb28/29* (right) shown as log fold change relative to wild type.

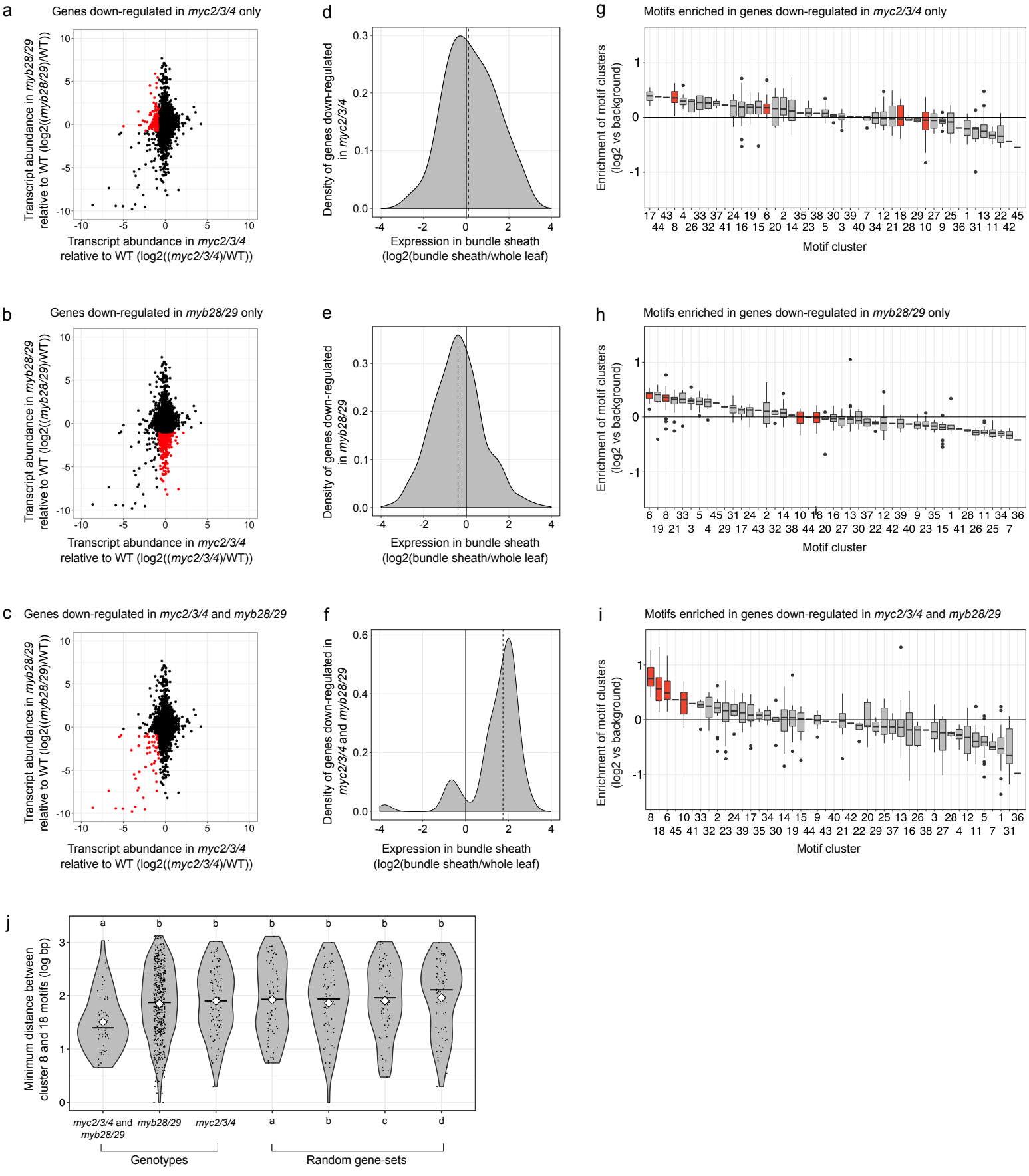
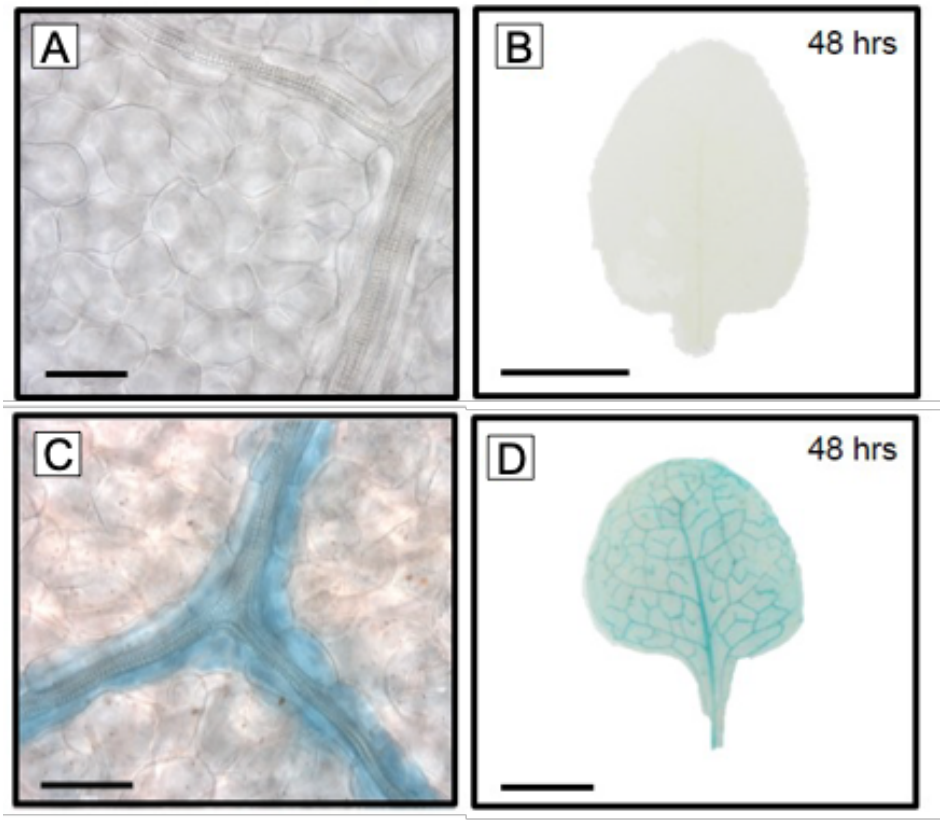
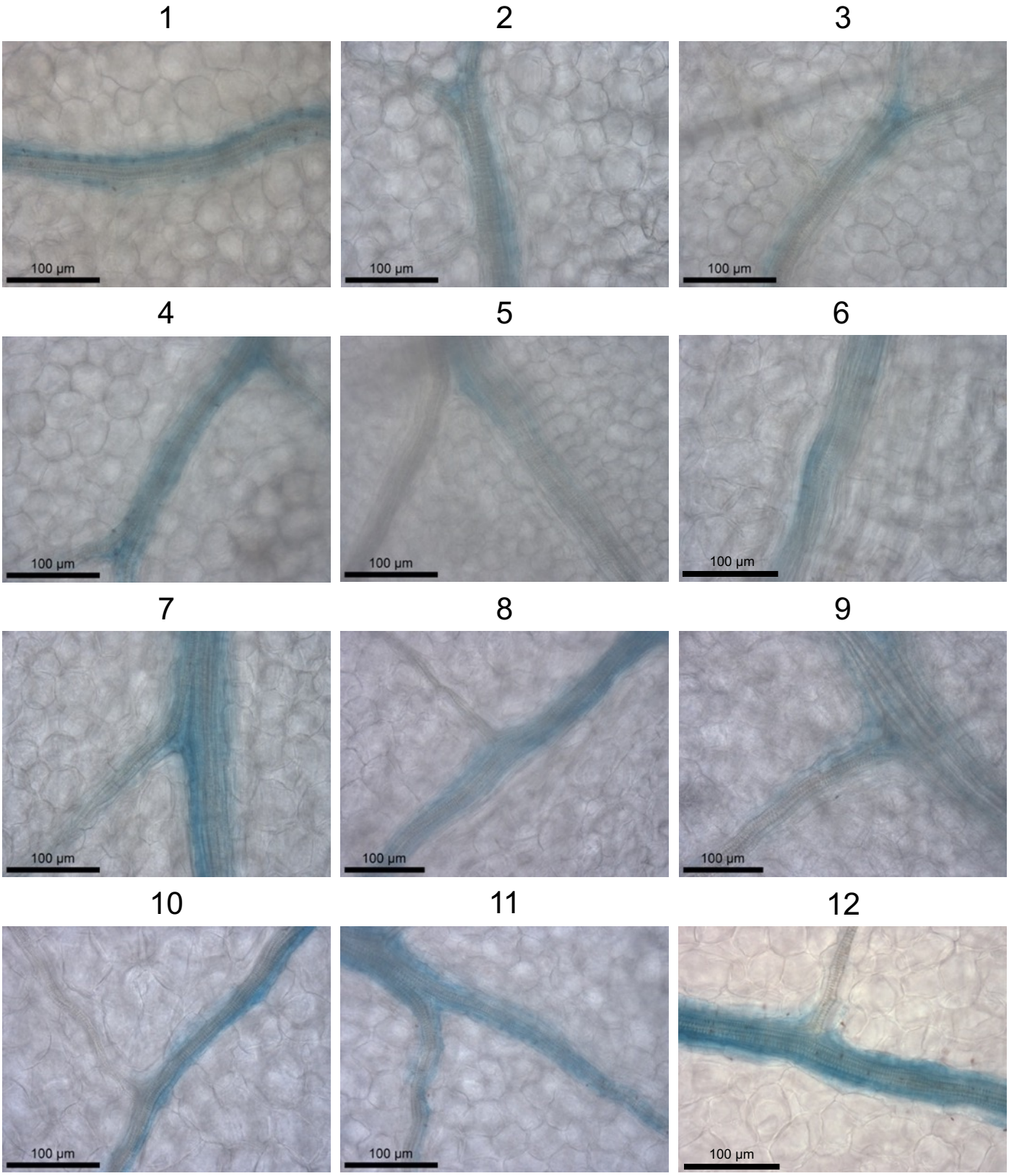

a
Genes down-regulated in myc2/3/4 only
Transcript abundance in myb28/29 relative to WT (log2((myb28/29)/WT))
Transcript abundance in myc2/3/4 relative to WT (log2((myc2/3/4)/WT))
b
Genes down-regulated in myb28/29 only
Transcript abundance in myb28/29 relative to WT (log2((myb28/29)/WT))
Transcript abundance in myc2/3/4 relative to WT (log2((myc2/3/4)/WT))
c
Genes down-regulated in myc2/3/4 and myb28/29
Transcript abundance in myb28/29 relative to WT (log2((myb28/29)/WT))
Transcript abundance in myc2/3/4 relative to WT (log2((myc2/3/4)/WT))
d
Density of genes down-regulated in myc2/3/4
Expression in bundle sheath (log2(bundle sheath/whole leaf)
e
Density of genes down-regulated in myb28/29
Expression in bundle sheath (log2(bundle sheath/whole leaf)
f
Density of genes down-regulated in myc2/3/4 and myb28/29
Expression in bundle sheath (log2(bundle sheath/whole leaf)
g
Motifs enriched in genes down-regulated in myc2/3/4 only
Enrichment of motif clusters (log2 vs background)
Motif cluster
17 43 4 33 37 24 19 6 2 35 38 30 39 7 12 18 29 27 25 1 13 22 45
44 8 26 32 41 16 15 20 14 23 5 3 40 34 21 28 10 9 36 31 11 42
h
Motifs enriched in genes down-regulated in myb28/29 only
Enrichment of motif clusters (log2 vs background)
Motif cluster
6 8 33 5 45 31 24 2 14 10 18 16 13 37 12 39 9 35 1 28 11 34 36
19 21 3 4 29 17 43 32 38 44 20 27 30 22 42 40 23 15 41 26 25 7
i
Motifs enriched in genes down-regulated in myc2/3/4 and myb28/29
Enrichment of motif clusters (log2 vs background)
Motif cluster
8 6 10 33 2 24 17 34 14 15 9 40 42 20 25 13 26 3 28 12 5 1 36
18 45 41 32 23 39 35 30 19 44 43 21 22 29 37 16 38 27 4 11 7 31
j
Minimum distance between cluster 8 and 18 motifs (log bp)
a b b b b b b
myc2/3/4 and myb28/29
myb28/29
myc2/3/4
a
b
c
d
Genotypes
Random gene-sets

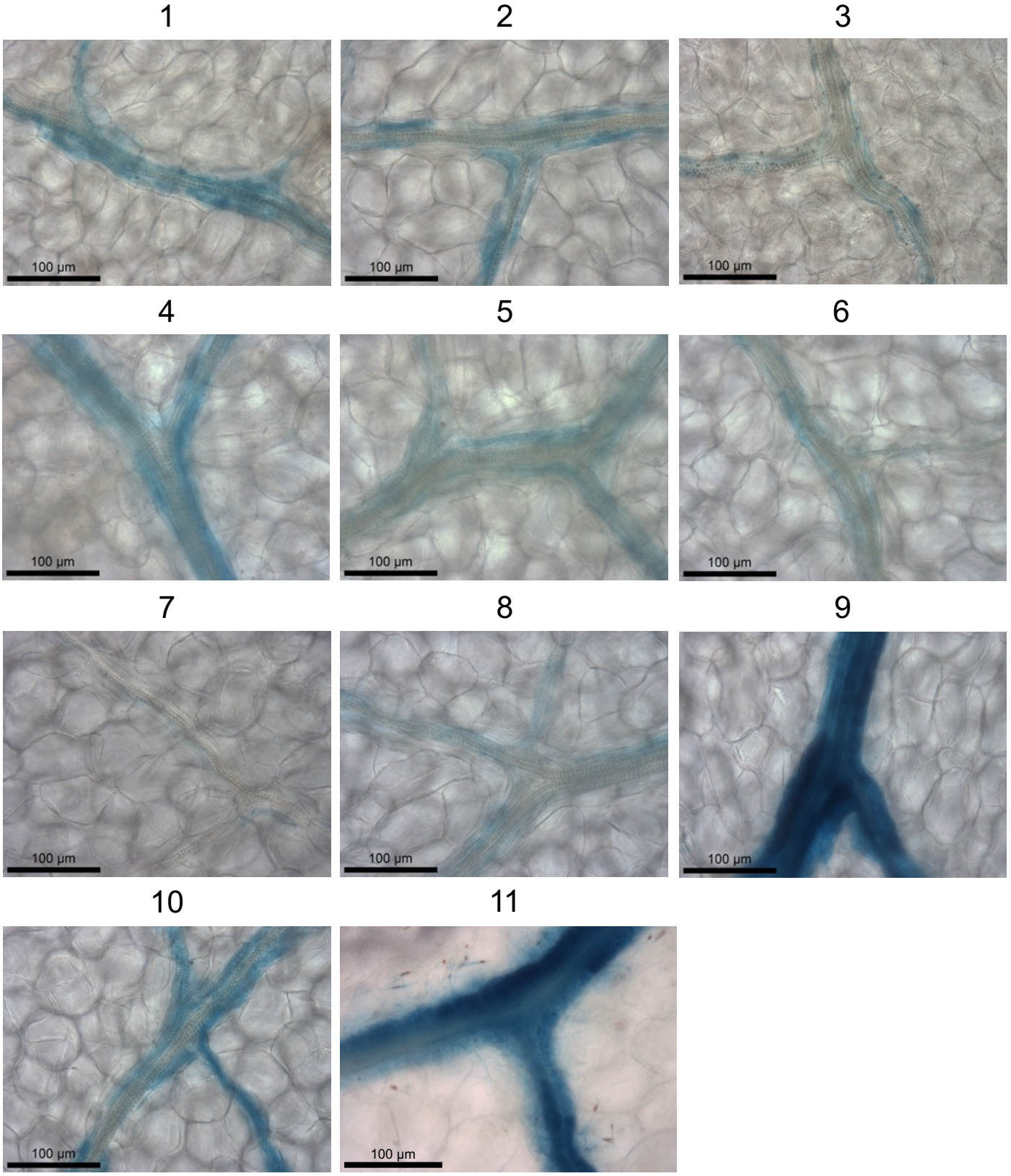
Figure 4: The MYC-MYB module controls bundle sheath expression of multiple genes. A, B, C) Change in expression in *myc2/3/4* compared with wild type (Major et al., 2017) plotted against that of *myb28/29* compared with wild type (Burow et al., 2015). Down-regulated genes (log2 < -0.75) in *myc2/3/4* only (A), *myb28/29* only (B) and in both *myc2/3/4* and *myb28/29* (C) are marked in red. D, E, F) Density plots for down-regulated genes highlighted in A, B and C indicating their expression in bundle sheath cells compared with whole leaves (Aubry et al., 2014). G, H, I) Enrichment analysis of motif clusters found in promoters of down-regulated genes highlighted in A, B and C. Clusters containing possible MYC binding sites (G-boxes) (Clusters 6 and 8) and MYB binding sites (Clusters 10 and 18) are highlighted in red. Note that Clusters 6, 8, 10 and 18 are strongly enriched in genes down-regulated in both *myc2/3/4* and *myb28/29*. Box-plots show inter-quartile range as upper and lower confines of the box, median as a solid black line and whiskers as maximum and minimum values excluding outliers. J) Violin plots depicting minimum distance (log bp) between cluster 8 and 18 motifs in promoters of genes highlighted in A, B, C and in four random sets of genes from Arabidopsis, ordered by median from smallest to largest. The median is shown as a horizontal black line, the mean as a white diamond. a and b indicate statistically significant differences ($p < 0.05$) determined by pairwise Wilcoxon rank-sum tests. n = individual genes tested. n=53 for genes down-regulated in *myc2/3/4* and *myb28/29*, n=506 for genes down=regulate only in *myb28/29*, n=101 in genes down-regulated only in *myc2/3/4* and n=66 for each of the random sets of Arabidopsis genes.



Supplemental Figure 1: Representative images of *proMYB*28 and *proMYB29* GUS. Staining performed for 48hrs and scale bars represent 0.5 cm (a and c) and 50 µm (b and d).

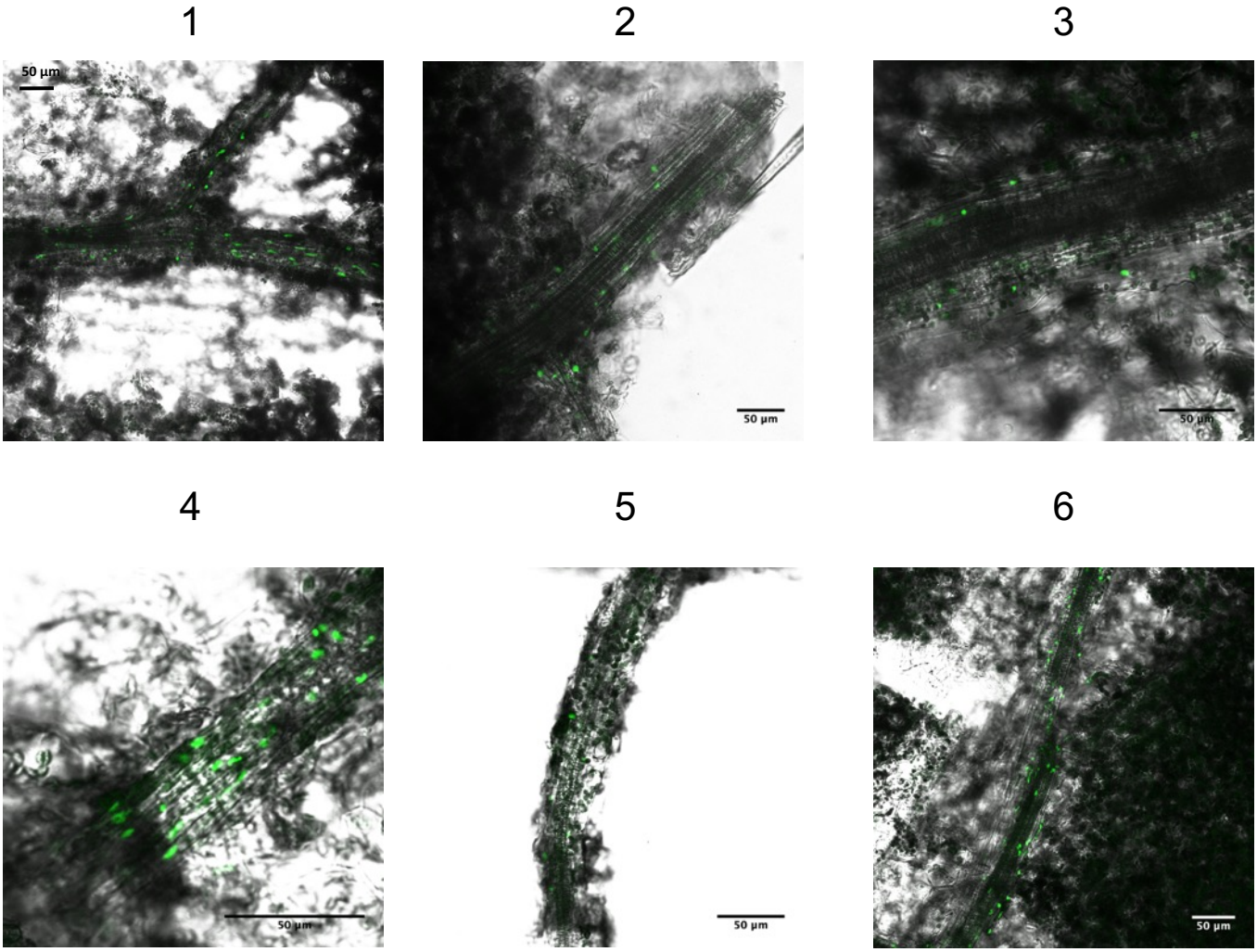


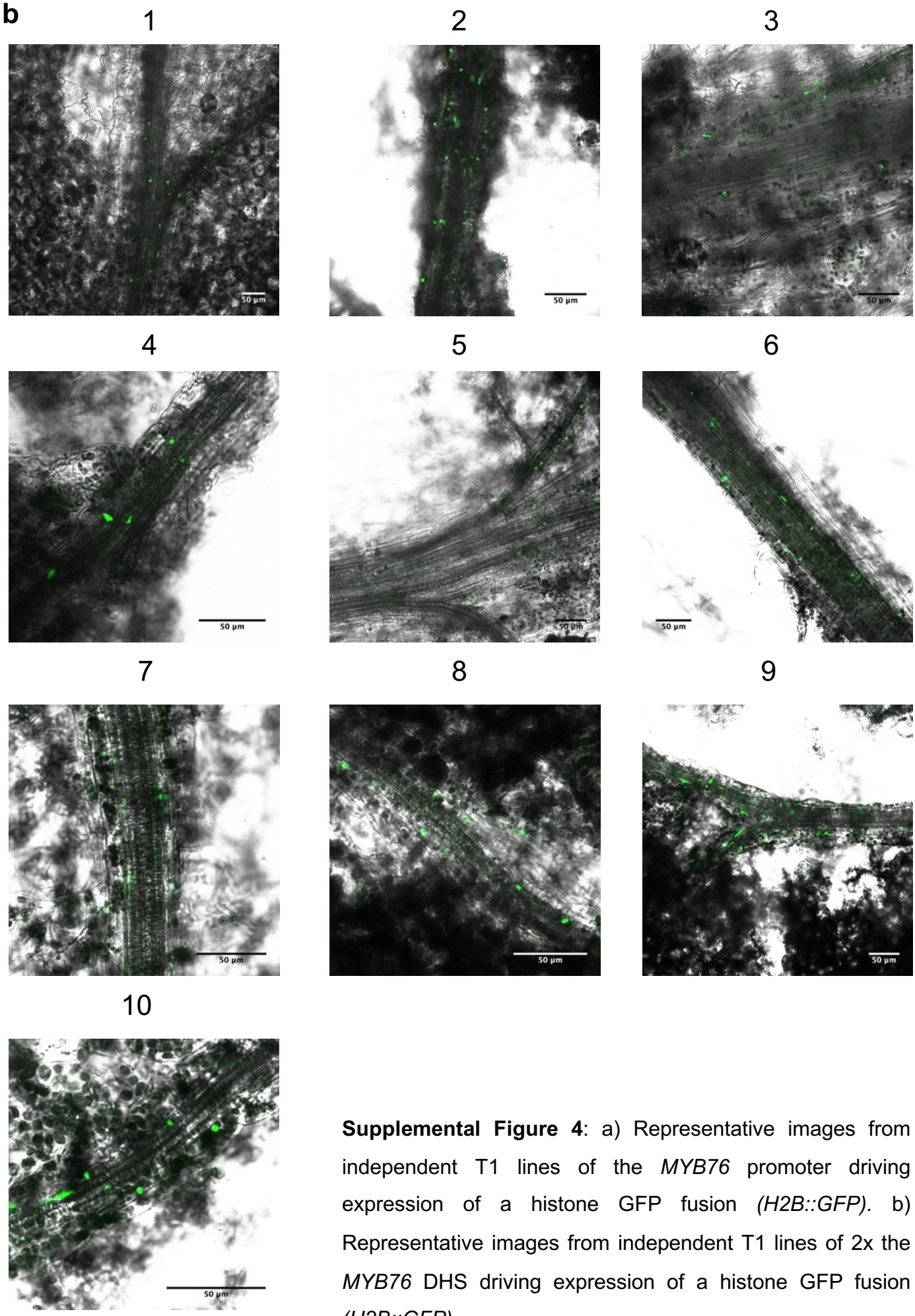
Supplemental Figure 2: Nucleotides -1725 to +279 relative to the predicted translational start site of *MYB76* generate preferential expression in the bundle sheath. Images from twelve independent transgenic lines. Leaves were stained for 30hrs. Scale bars represent 100 µm.



Supplemental Figure 3: The genomic *MYB76* sequence fused to GUS generates preferential expression in the bundle sheath. Images from eleven independent transgenic lines Leaves were stained for 72hrs except line 11 which was stained for 48hrs. Scale bars represent 100 µm.

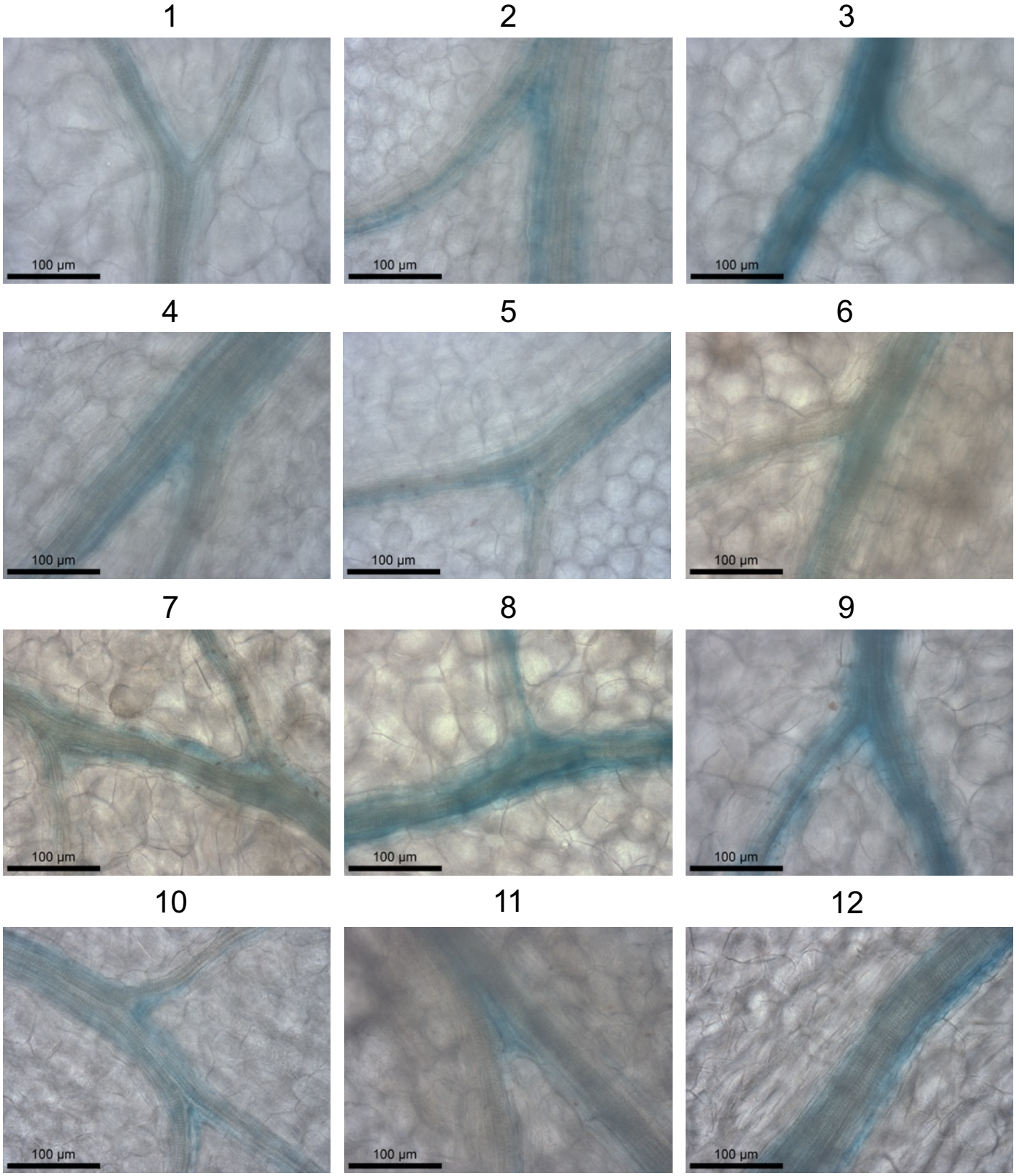
a



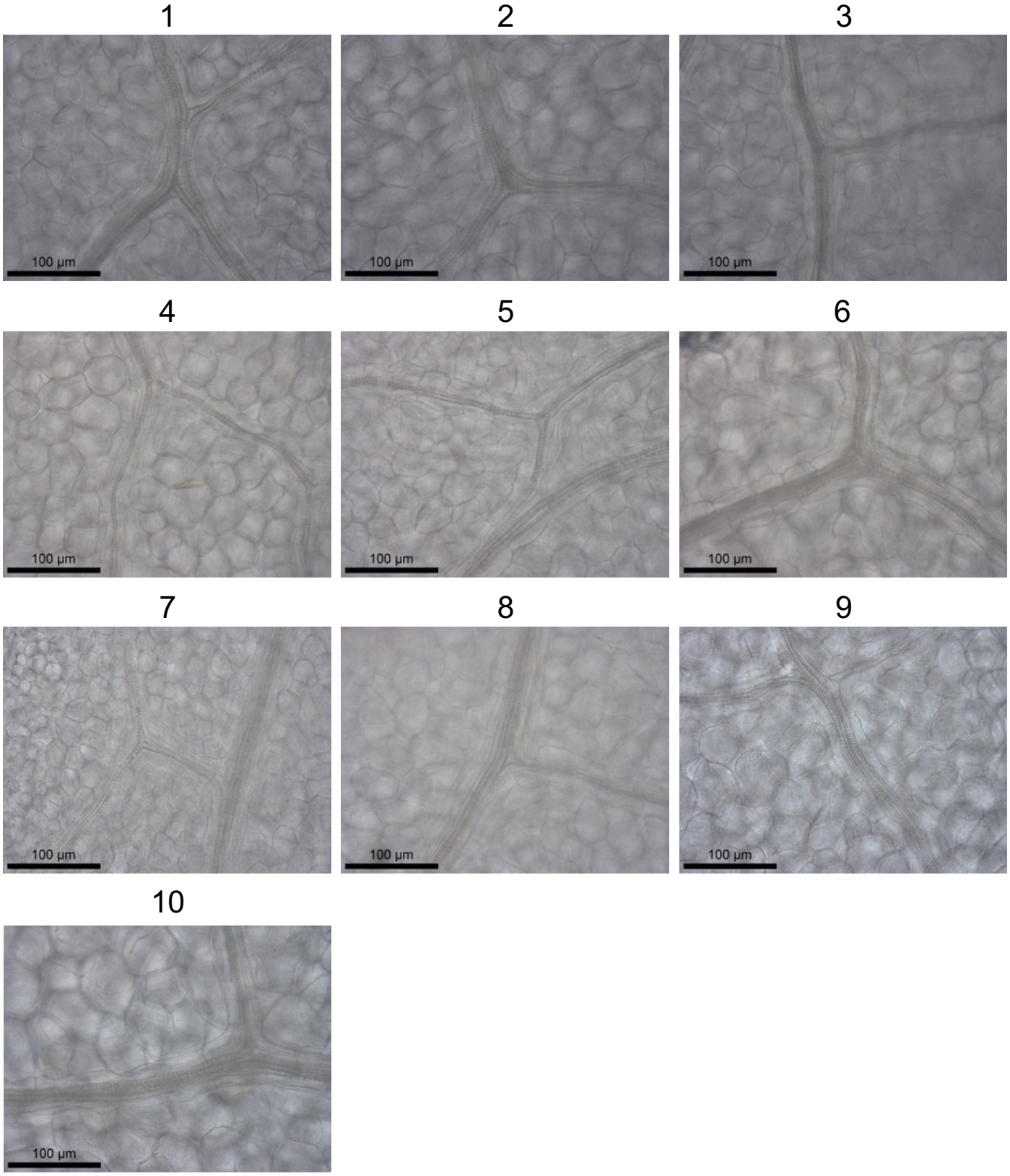


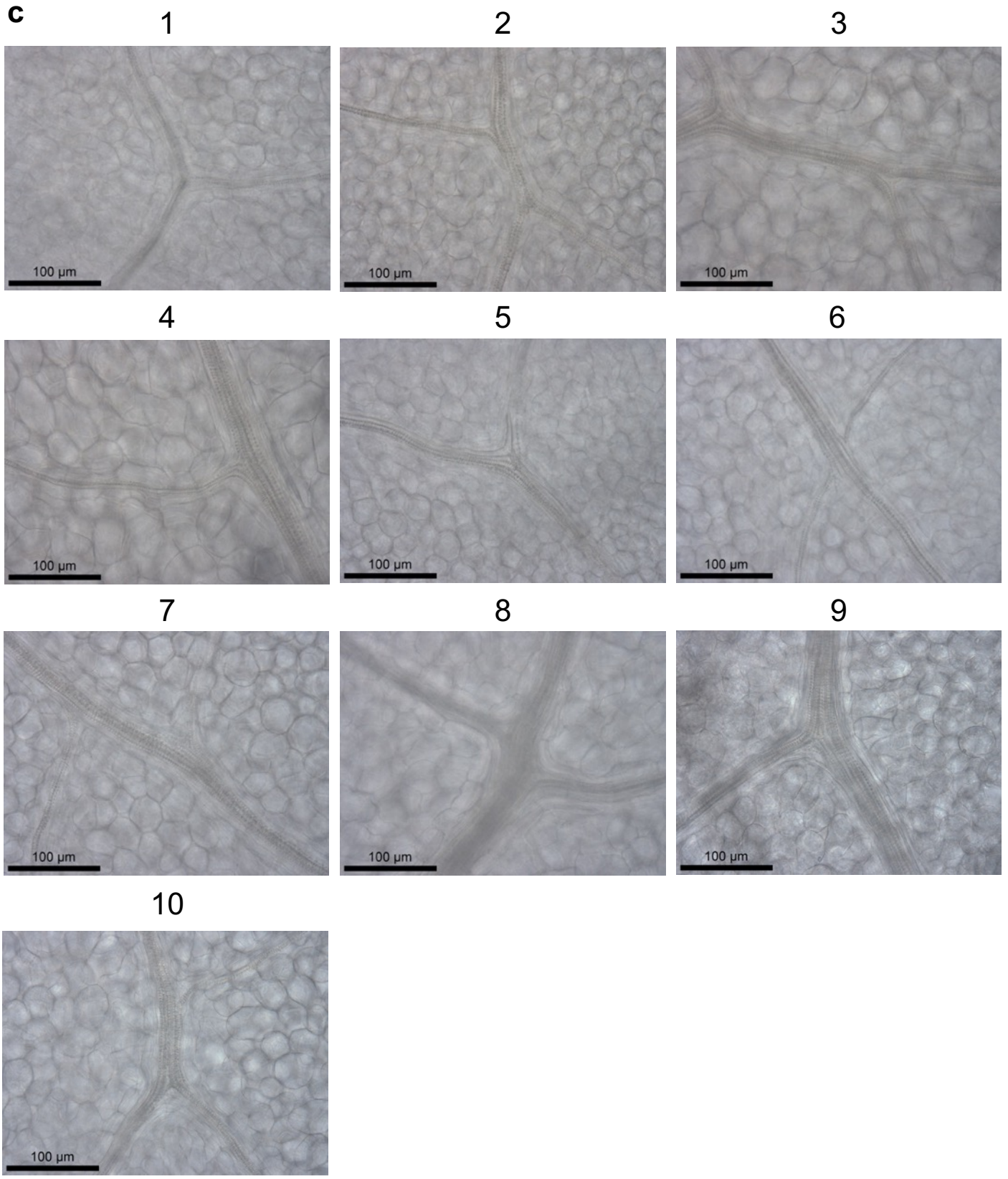
Supplemental Figure 4: a) Representative images from independent T1 lines of the *MYB76* promoter driving expression of a histone GFP fusion *(H2B::GFP).* b) Representative images from independent T1 lines of 2x the *MYB76* DHS driving expression of a histone GFP fusion *(H2B::GFP).*

a

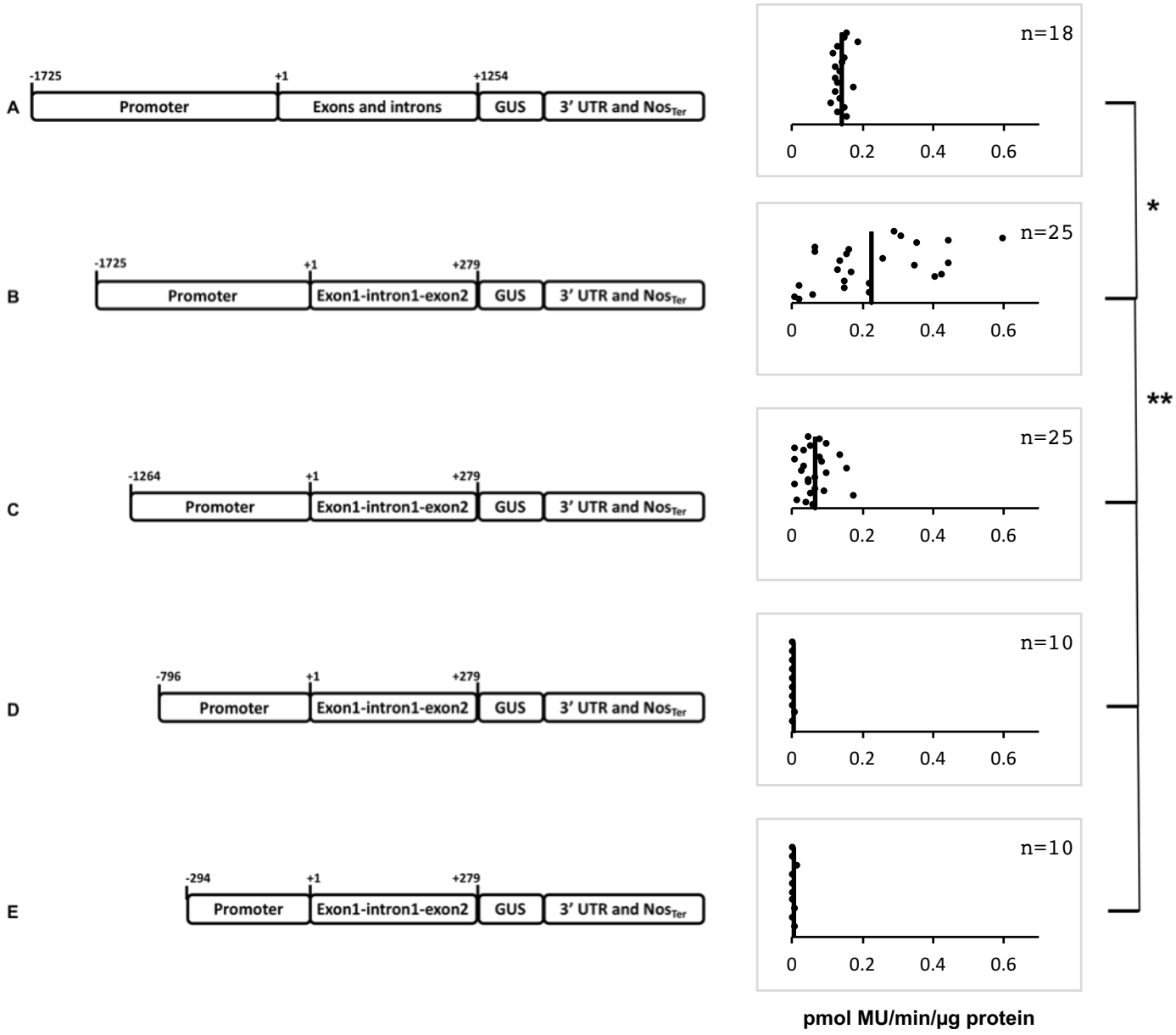


b



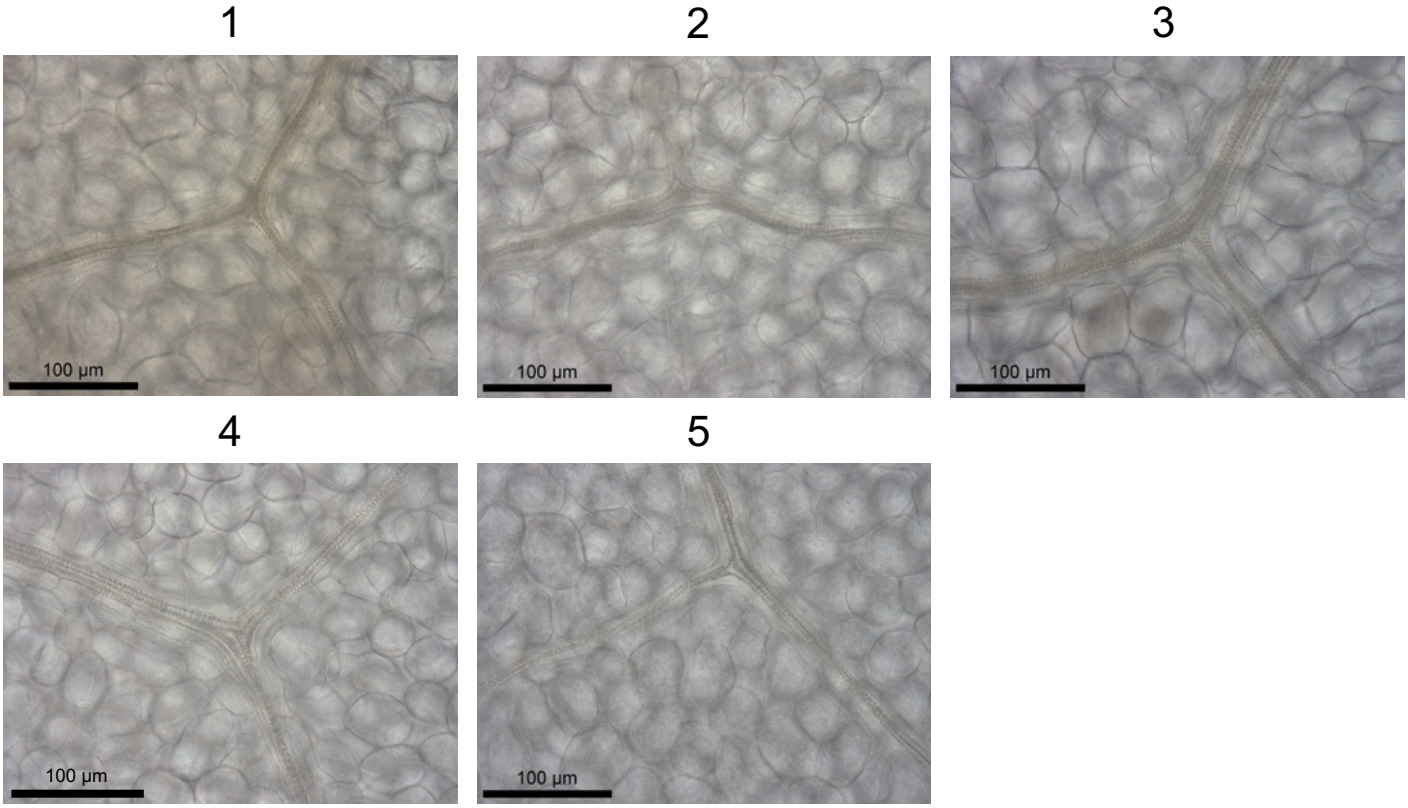


Supplemental Figure 5: a) Nucleotides -1264 to +279 relative to the predicted translational start site of MYB76 generate preferential expression in the bundle sheath. Images from twelve independent transgenic lines. b) 796bp of the promoter combined with the first exon and intron of *MYB76* does not generate BS preferential expression. Images from ten independent transgenic lines. c) 294bp of the promoter combined with the first exon and intron of *MYB76* does not generate BS preferential expression. Images from ten independent transgenic lines. Leaves were stained for 48hrs. Scale bars represent 100 µm.

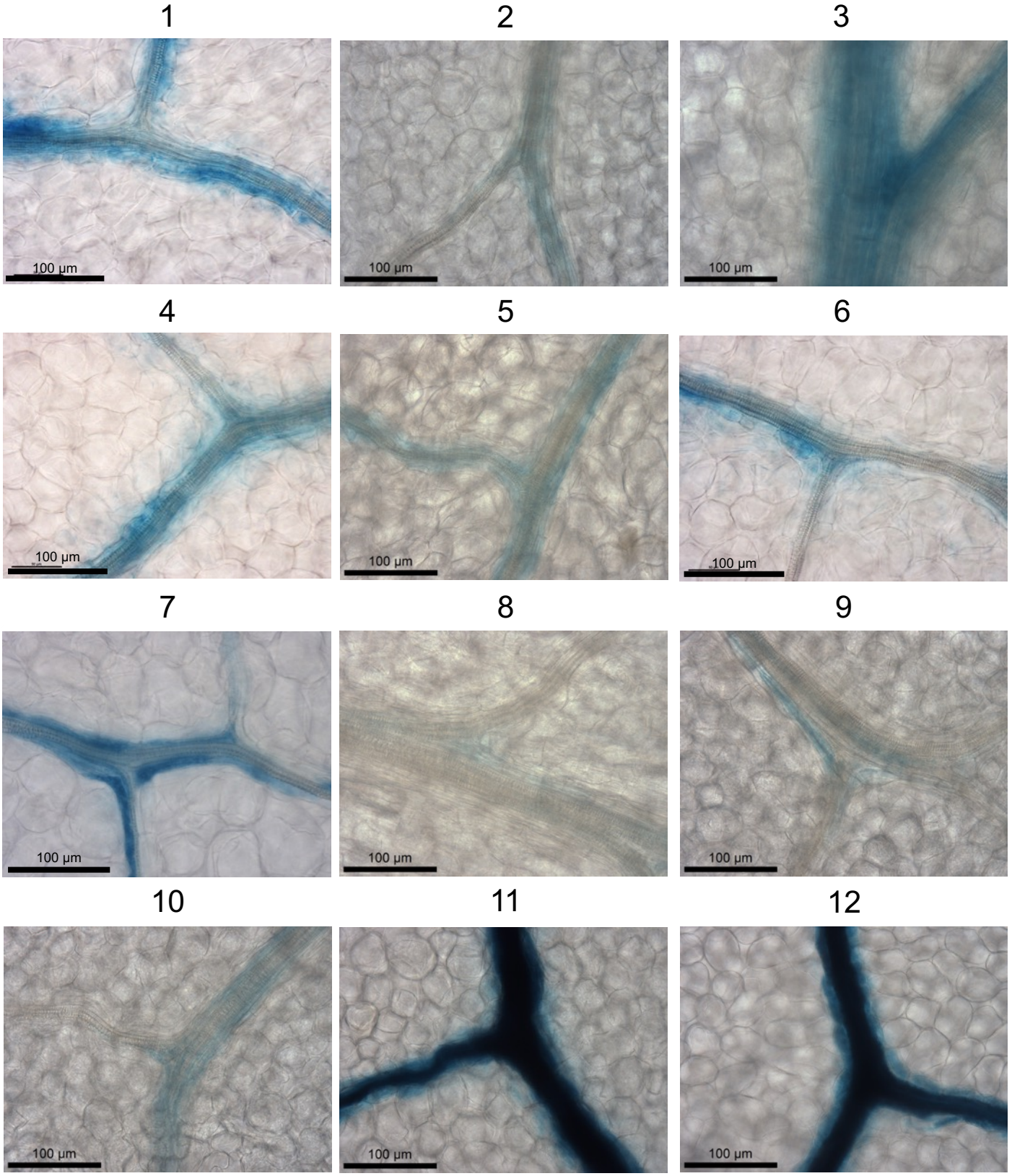


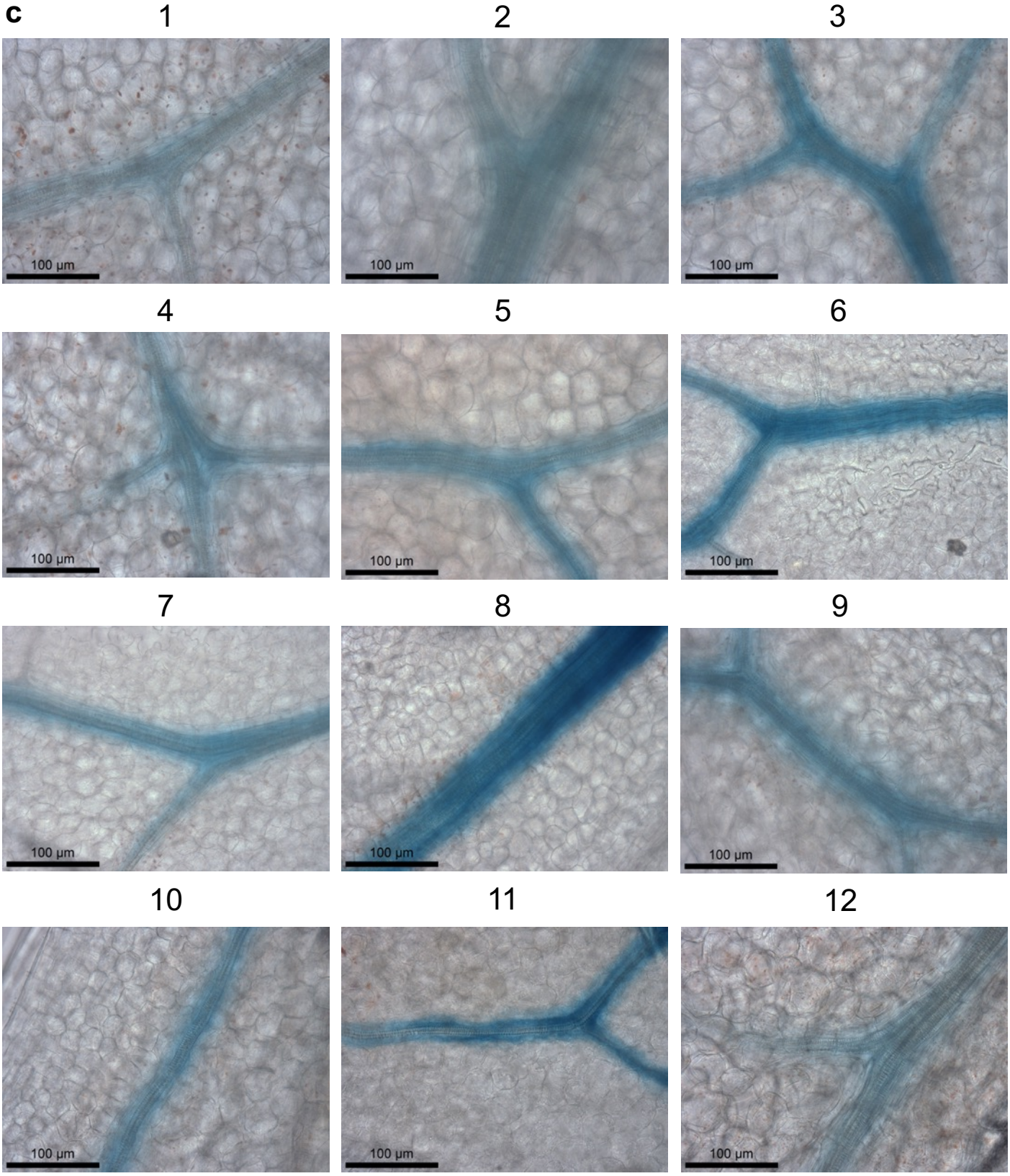
Supplemental Figure 6: Schematics of deletion series (left) and quantification of MUG activity via the flourometric assay (right) for multiple independent transformants from each construct. (A) The *MYB76* genomic sequence. (B) The *MYB76* promoter including the first two exons and the first intron fused to the *uidA*. (C-E) A region between -1264 and -796bp is responsible for GUS accumulation in the bundle sheath. The fluorometric MUG assay shows quantitative repressors and enhancers are located in the gene and in the promoter respectively. X-axis indicates GUS activity and individual biological replicates are ordered randomly on the y axis. n, number of multiple independent transformants; *, p value < 0.05; ** p value < 0.001 determined by T-tests.

a



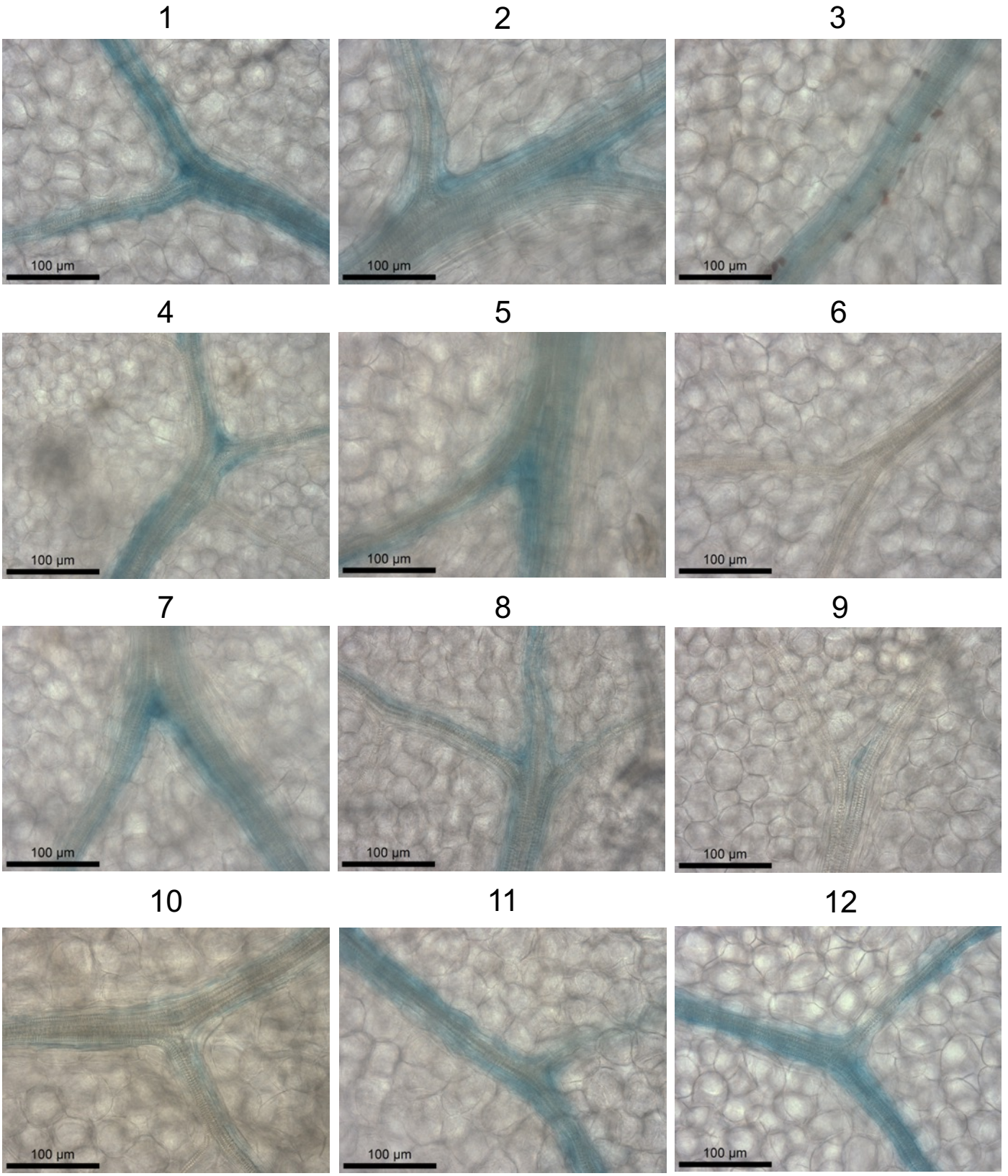
b



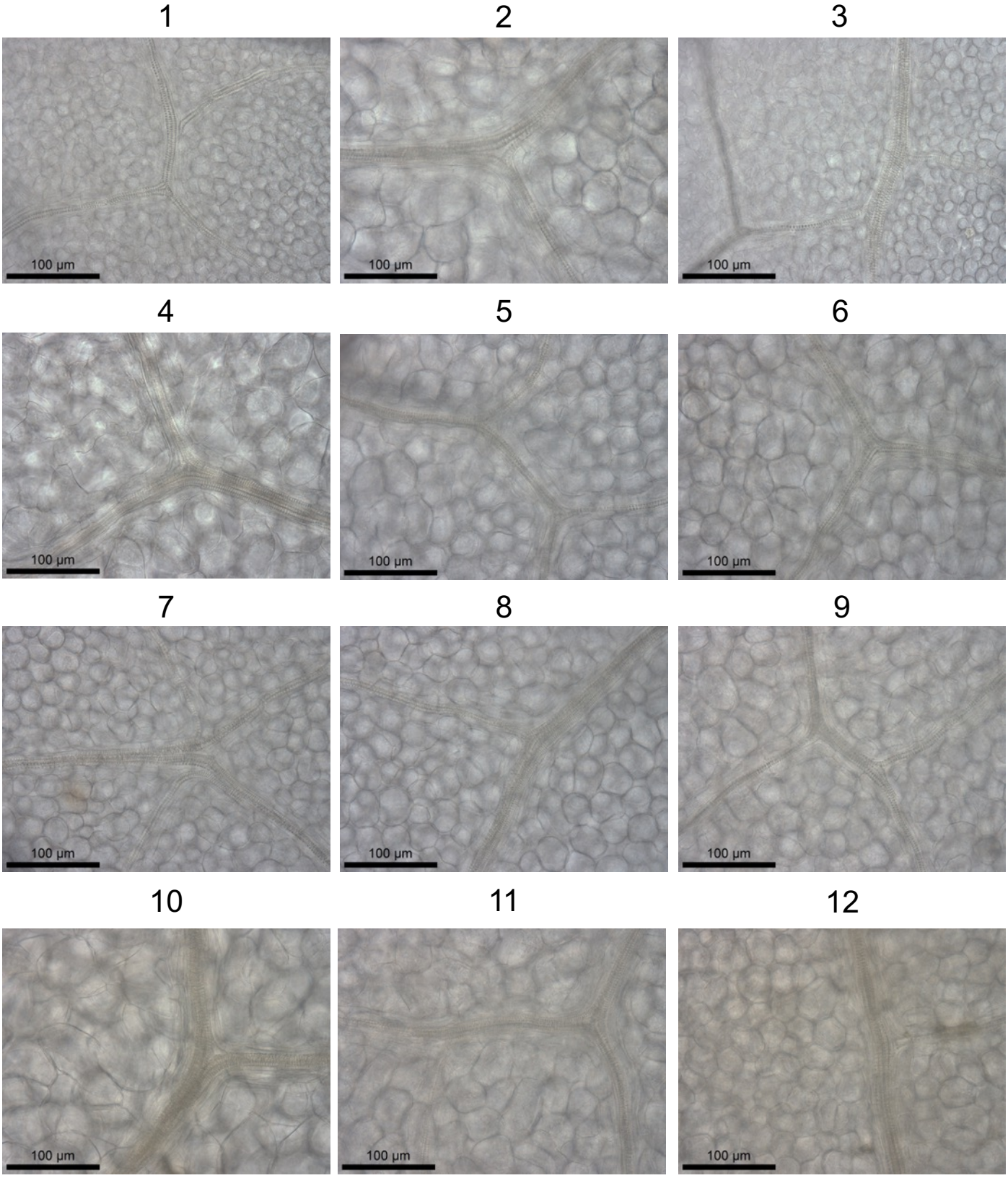


Supplemental Figure 7: a) Deleting the MYB76 DHS leads to loss of GUS in the BS. Images from five independent transgenic lines, leaves were stained for 48 hrs. b) The *MYB76* DHS combined with the minimal 35SCaMV promoter generates preferential expression in the bundle sheath. Images from twelve independent transgenic lines. Leaves were stained for 72hrs. c) Oligomerizing the *MYB76* DHS combined with the minimal 35SCaMV promoter generates strong preferential expression in the bundle sheath. Images from twelve independent transgenic lines. Leaves were stained for 3hrs. Scale bars represent 100 µm.

a

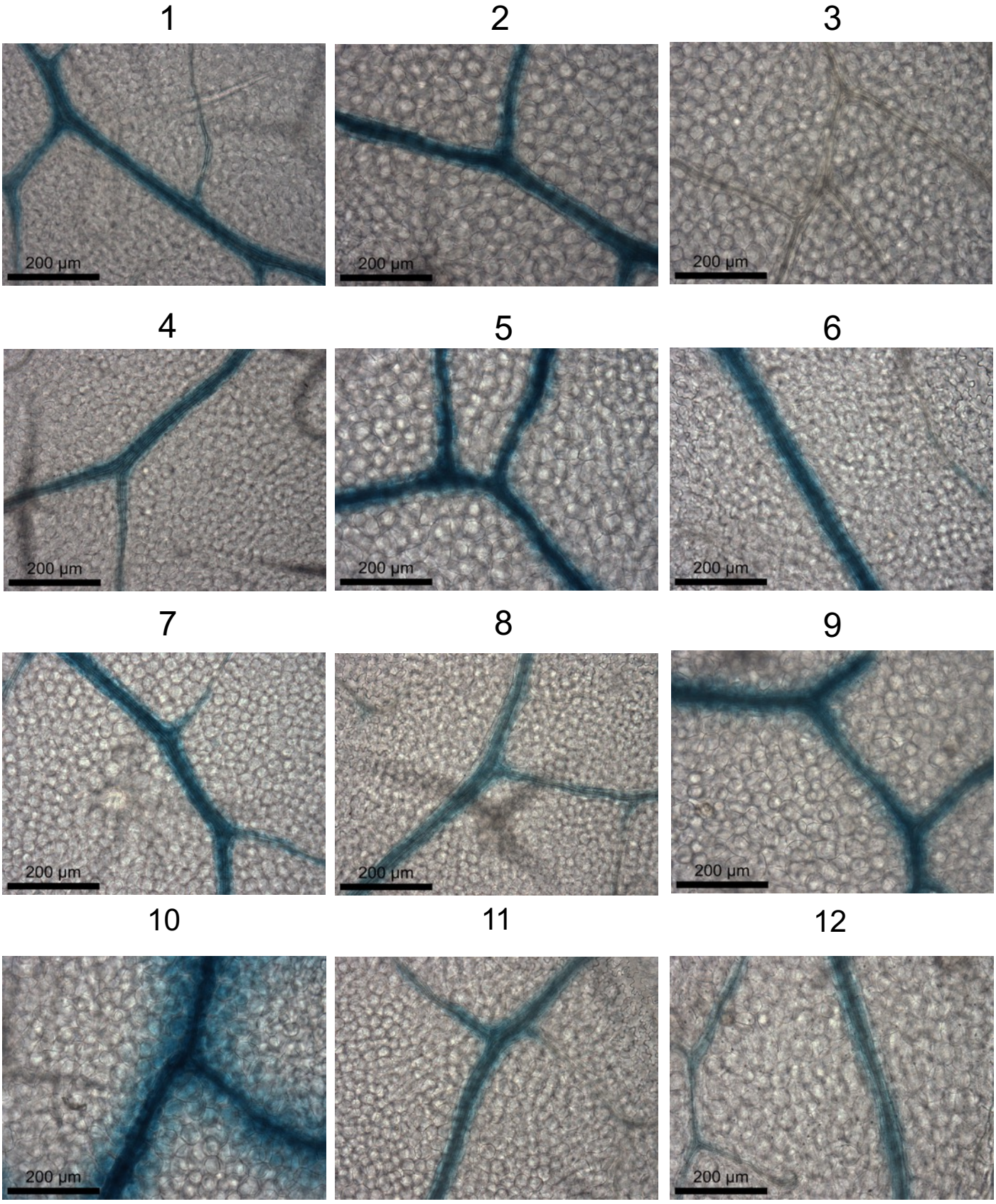


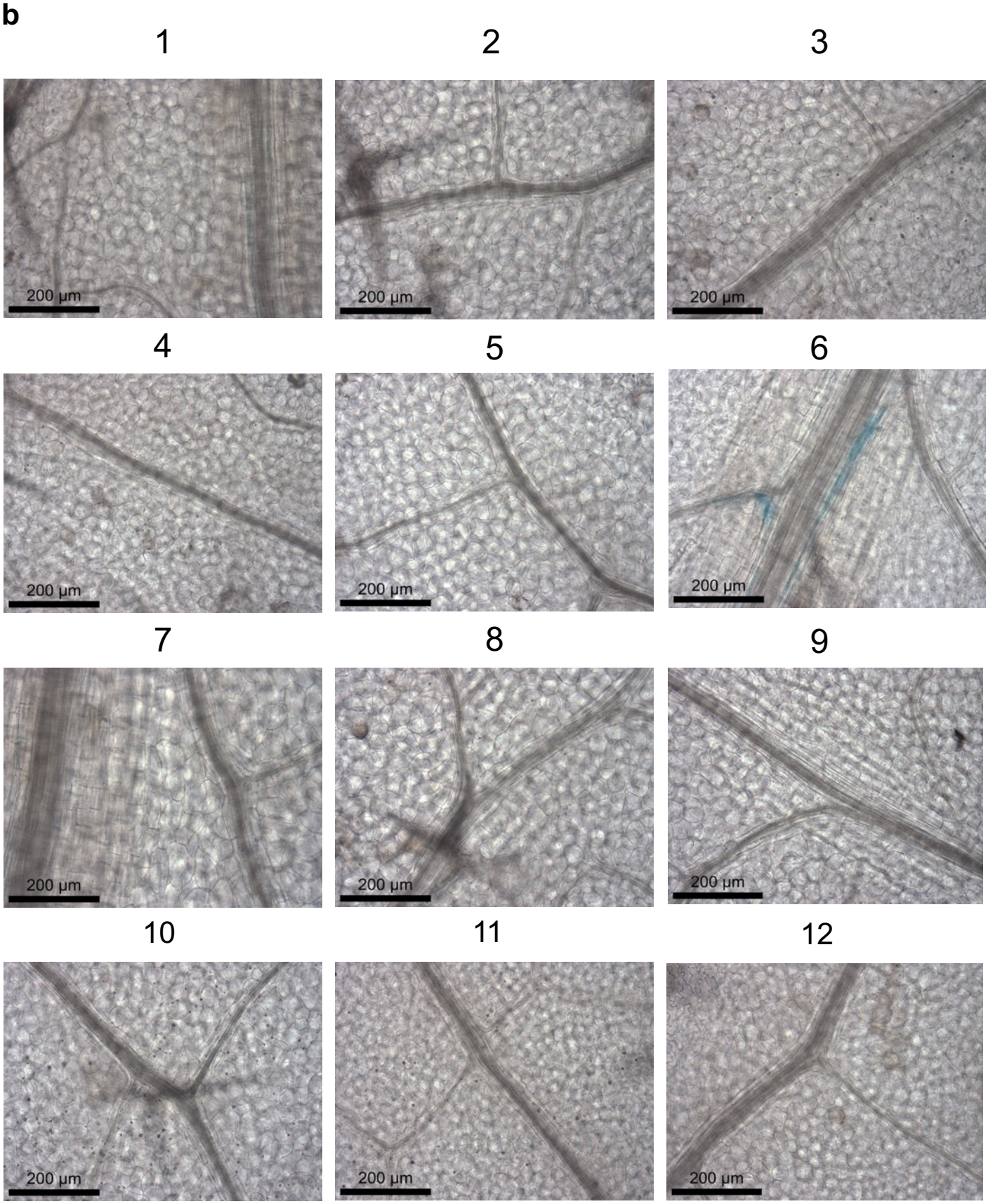
b



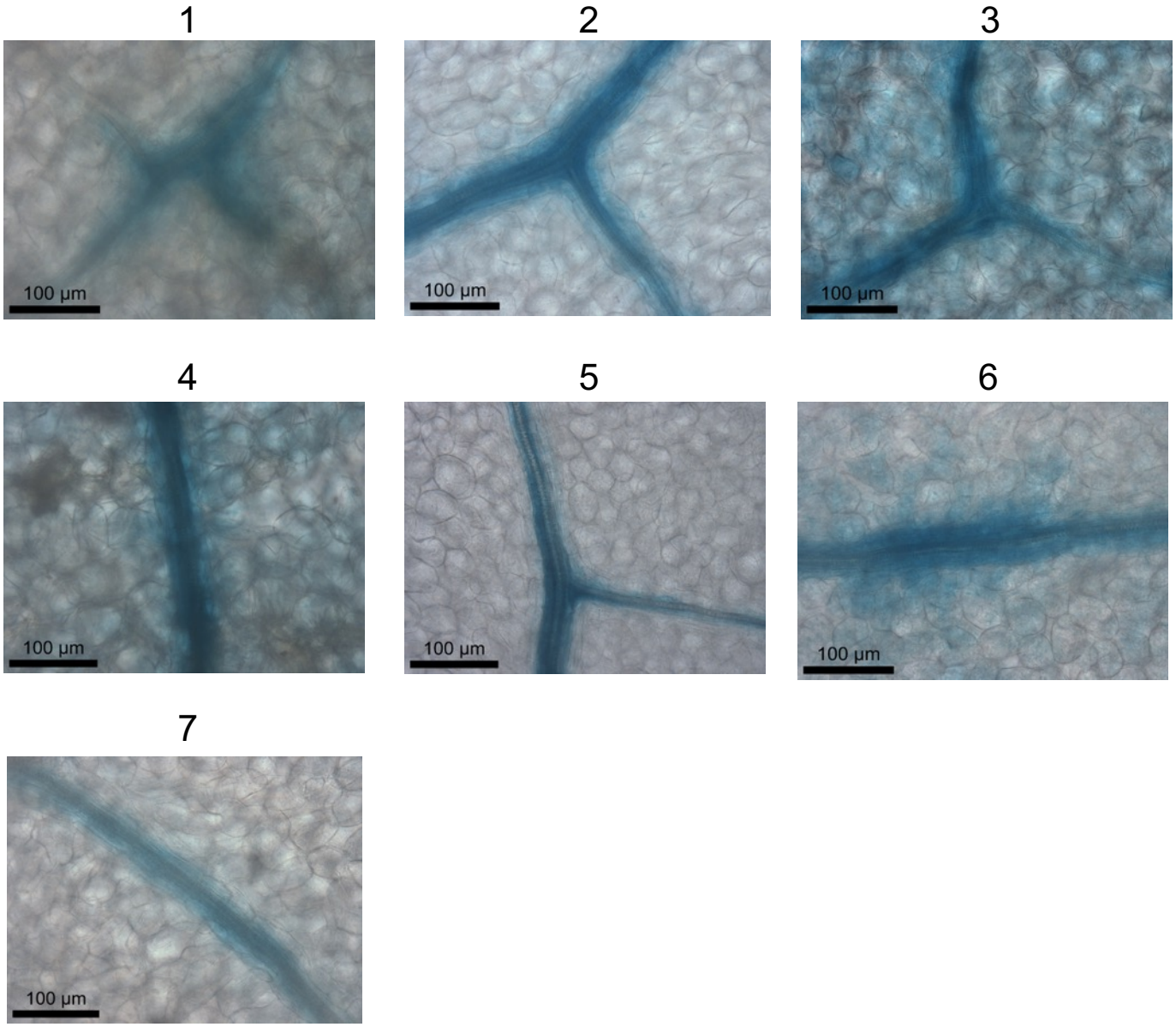
Supplementary Figure 8: a) Mutation of motif 1 (TGGGCA) from the *MYB76* promoter does not abolish accumulation of GUS from the bundle sheath. Images from twelve independent transgenic lines. Leaves were stained for 48hrs. b) Mutation of motif 2 (TGCACCG) from the *MYB76* promoter motif leads to loss of GUS in the BS. Images from twelve independent transgenic lines. Leaves were stained for 48hrs. Scale bars represent 100 µm.

a

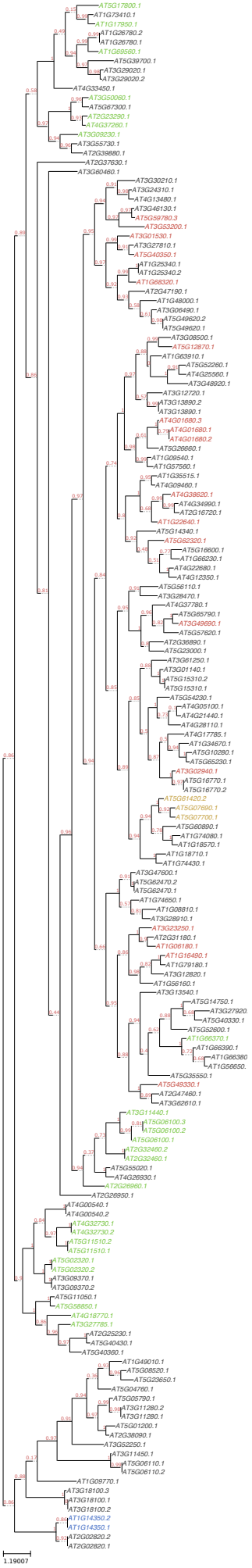




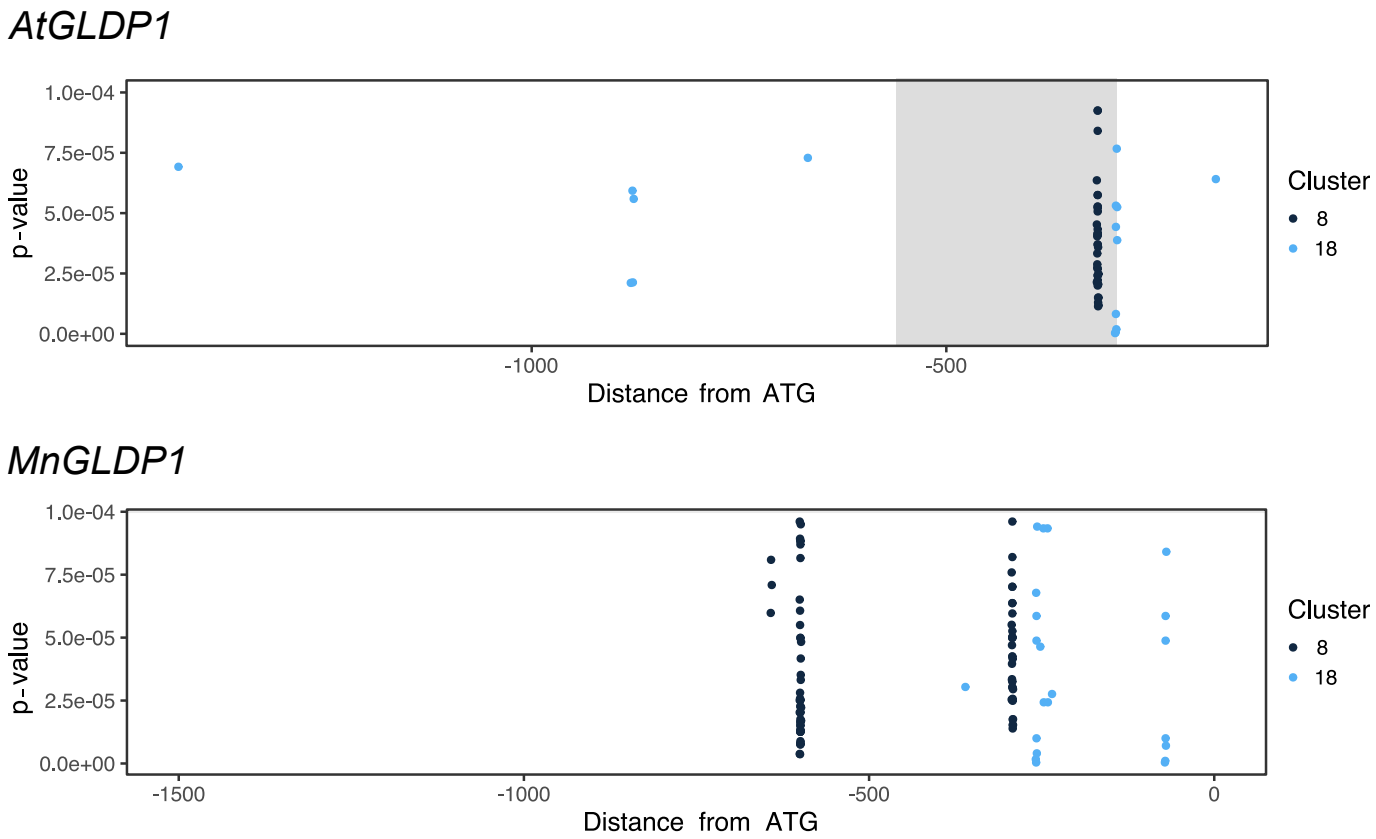
Supplementary Figure 9: a) *MYB76* promoter and *gDNA* fused to GUS generated using Golden Gate cloning generates preferential expression in the bundle sheath. Images from twelve independent transgenic lines. b) Mutation of motif 2 (TGCACCG) in a full length *MYB76* promoter and gDNA translational fusion abolishes GUS accumulation. Images from twelve independent transgenic lines. Leaves were stained for 24 hrs. Scale bars represent 200 µm.



Supplemental Figure 10: Two copies of the TGCACCG motif combined with ten upstream and ten downstream nucleotides within the context of the native *MYB76* promoter fused to the minimal 35SCaMV minimal promoter generate preferential expression in the bundle sheath. Images from seven independent transgenic lines. Leaves were stained for 86hrs. Scale bars represent 100 µm.



Supplemental Figure 11. Phylogenetic tree of MYB transcription factors in *A. thaliana* based on amino acid sequence of whole proteins. Cluster 10 MYBs are coloured in green, cluster 18 MYBs are coloured in red and cluster 31 MYBs are coloured in blue. MYB transcription factors without defined binding motifs are in black and MYB28, MYB29 and MYB76 are coloured in gold.



Supplemental Figure 12. Cluster 8 and 18 transcription factor binding motifs within the promoters of *A. thaliana* and *M. nitens GLDP1* genes. The y axis shows p-values of matches between DHS sequence and motif PWMs and the x axis shows position of the motif centre relative to the translational start site. The grey box in the *AtGLDP1* promoter represents the V-box (Adwy et al., 2015).