

The bundle sheath of rice is conditioned to play an active role in water transport as well as sulfur assimilation and jasmonic acid synthesis

Lei Hua¹, Sean R. Stevenson¹, Ivan Reyna-Llorens¹, Haiyan Xiong¹, Stanislav Kopriva² and Julian M. Hibberd¹

¹Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, United Kingdom

²Institute for Plant Sciences, Cluster of Excellence on Plant Sciences (CEPLAS), University of Cologne, Zùlpicher Str. 47b, 50674 Cologne, Germany

Email addresses:

LH - lh556@cam.ac.uk

SRS - srs62@cam.ac.uk

IRL - suallorems@gmail.com

HX - hx253@cam.ac.uk

SK - skopriva@uni-koeln.de

JMH - jmh65@cam.ac.uk

Running title: Gene expression in the rice bundle sheath

Keywords: *Oryza sativa*; Bundle Sheath; Water Transport; Sulphur Metabolism; Deep sequencing; Laser Capture Microdissection

Abstract

Leaves comprise multiple cell types but our knowledge of the patterns of gene expression that underpin their functional specialization is fragmentary. Our understanding and ability to undertake rational redesign of these cells is therefore limited. We aimed to identify genes associated with the incompletely understood bundle sheath of C_3 plants, which represents a key target associated with engineering traits such as C_4 photosynthesis into rice. To better understand veins, bundle sheath and mesophyll cells of rice we used laser capture microdissection followed by deep sequencing. Gene expression of the mesophyll is conditioned to allow coenzyme metabolism and redox homeostasis as well as photosynthesis. In contrast, the bundle sheath is specialized in water transport, sulphur assimilation and jasmonic acid biosynthesis. Despite the small chloroplast compartment of bundle sheath cells, substantial photosynthesis gene expression was detected. These patterns of gene expression were not associated with presence/absence of particular transcription factors in each cell type, but rather gradients in expression across the leaf. Comparative analysis with C_3 *Arabidopsis* identified a small gene-set preferentially expressed in bundle sheath cells of both species. This included genes encoding transcription factors from fourteen orthogroups, and proteins allowing water transport, sulphate assimilation and jasmonic acid synthesis. The most parsimonious explanation for our findings is that bundle sheath cells from the last common ancestor of rice and *Arabidopsis* was specialized in this manner, and since the species diverged these patterns of gene expression have been maintained.

Significance statement: The role of bundle sheath cells in C_4 species have been studied intensively but this is not the case in leaves that use the ancestral C_3 pathway. Here, we show that gene expression in the bundle sheath of rice is specialized to allow sulphate and nitrate reduction, water transport and jasmonate synthesis, and comparative analysis with *Arabidopsis* indicates ancient roles for bundle sheath cells in water transport, sulphur and jasmonate synthesis.

Introduction

Although the major cell types of a leaf were described in the 19th century (Haberlandt, 1884) we have an incomplete understanding of the role that each plays (Aubry et al., 2014b; Mustroph et al., 2009). This lack of knowledge hinders our understanding of how basic processes are organized but is also likely to limit the rational re-design of leaves for crop improvement. One example is associated with attempts to engineer C₄ photosynthesis into species such as C₃ rice to increase yields (von Caemmerer et al., 2012; Hibberd et al., 2008; Langdale, 2011). As C₄ photosynthesis typically requires metabolic compartmentation between mesophyll and bundle sheath cells, a better understanding of these tissues in rice may facilitate such a project.

The C₄ pathway is thought to have evolved over 60 times independently in monocotyledons and dicotyledons in response to reduced water availability and CO₂ supply (Sage, 2004). These environmental factors reduce the ratio of carboxylation to oxygenation events at the active site of RuBisCO, and so lead to higher rates of the photorespiration (Tipple and Pagani, 2007; Sage et al., 2012). Whilst in some C₄ species, a carbon concentrating mechanism is established within large cells (Jurić et al., 2016; von Caemmerer et al., 2014; Voznesenskaya et al., 2001) in the majority of known lineages this takes place across distinct cell-types. In these two-celled C₄ species, RuBisCO activity is replaced with phosphoenolpyruvate carboxylase in mesophyll cells to allow bicarbonate to be fixed into oxaloacetate. High concentrations of C₄ acids derived from oxaloacetate build up in mesophyll cells and drive diffusion to the bundle sheath where C₄ acid decarboxylases resupply CO₂ to RuBisCO. This reorganization of photosynthesis is thus enabled by the presence of distinct cell types such as the mesophyll and bundle sheath (Furbank, 2016). Not only do bundle sheath cells of C₄ plants undertake a key role in photosynthesis, but they are also the primary location of starch synthesis (Lunn and Furbank, 1997) and sulphur metabolism (Gerwick et al., 1980; Passera and Ghisi, 1982; Schmutz and Brunold, 1984; Burnell, 1984; Burgener et al., 1998). Distinct classes of bundle sheath cells have been reported in maize, with abaxial bundle sheath cells of rank-2 intermediate being specialised for phloem loading (Bezruczyk et al., 2021). The importance of bundle sheath cells in the C₄ leaf, and the discovery that their thickened cell walls allow them to be separated from the

rest of the leaf led to them being studied intensively. In summary, in the C₄ leaf bundle sheath cells are well characterized and carry out a variety of specialized roles.

Although the vast majority of plants use the ancestral C₃ pathway, our understanding of gene expression in bundle sheath cells specifically, and other major cell types more generally of C₃ leaves is poor. In *Arabidopsis thaliana* (hereafter *Arabidopsis*) the bundle sheath represents approximately 15% of chloroplast-containing cells in the leaf (Kinsman and Pyke, 1998). Whilst the photosynthetic apparatus is functional in C₃ bundle sheath cells (Fryer et al., 2002; Williams et al., 1989), the absolute number of chloroplasts per cell is low. Consistent with this, reducing chlorophyll accumulation or increasing the chloroplast compartment in these cells have limited impact on leaf level photosynthesis (Janacek et al., 2009; Wang et al., 2017). Instead, it appears that the bundle sheath of *Arabidopsis* is specialized in sulphur metabolism and glucosinolate synthesis (Aubry et al., 2014b; Koroleva et al., 2010). Stress responsive regulation of aquaporins in bundle-sheath cells are considered important for hydraulic conductance of the whole leaf (Sade et al., 2014; Shatil-Cohen et al., 2011; Attia et al., 2020) and consistent with this, bundle sheath cells have been proposed to help maintain hydraulic integrity of the xylem (Griffiths et al., 2013; Sage, 2001) as well as regulate flux of mineral and metabolites in and out of the leaf (Leegood, 2008; Wigoda et al., 2017). In summary, compared with C₄ plants we have a relatively poor understanding of the role of bundle sheath cells in C₃ species, and this is particularly the case in grasses such as rice.

In roots, one approach that has been used widely to define the function of discrete cell types is to generate lines in which distinct tissues are marked with a fluorescent protein, and after protoplast isolation and cell sorting, the patterns of gene expression defined (Birnbaum et al., 2003; Brady et al., 2007). In leaves, this approach has been less widely adopted, likely due to the longer incubation times typically required to generate protoplasts, and concerns about stress and de-differentiation taking place during protoplasting (Sawers et al., 2007). Recently, rapid protoplasting followed by isolation of bundle sheath cells indicated a key role in transport (Wigoda et al., 2017), single cell RNA sequencing allowed distinct patterns of gene expression to be related to discrete cell types of the *Arabidopsis* leaf, and indicated that bundle sheath protoplasts were most similar to xylem cells (Kim et al., 2021).

121 Alternate technologies that have been applied to this problem include the isolation of
122 ribosomes from specific cell types after they were labelled with an exogenous tag (Aubry et
123 al., 2014b; Mustroph et al., 2009), or the use of laser capture microdissection (Jiao et al.,
124 2009). This latter approach circumvents the need to identify promoters with highly specific
125 expression domains and the production of transgenic lines.

126 Here we used an optimised laser capture microdissection protocol for RNA isolation from
127 leaves (Hua and Hibberd 2019) to study the patterns of gene expression in bundle sheath,
128 veinal (including phloem and xylem parenchyma, xylem, as well as sieve elements and
129 companion cells), and mesophyll cells of rice. We had three main hypotheses. First, that as
130 the rice bundle sheath contains a small chloroplast compartment, gene expression would
131 be poorly set up for photosynthesis. Second, as in other species, gene expression in the
132 rice bundle sheath would favour water transport. Third, in contrast to Arabidopsis, as rice
133 does not make glucosinolates there has been no selection pressure to restrict sulphur
134 assimilation to the bundle sheath. Mesophyll and bundle sheath cells were distinguished by
135 over-representation of terms including photosynthesis and co-enzyme metabolism in the
136 former, and solute transport and protein synthesis in the latter. Transcripts encoding the
137 majority of aquaporins were more abundant in bundle sheath cells, and this was also the
138 case for transcripts encoding proteins associated with sulphur assimilation and nitrate
139 reduction. Transcription factors that were preferential to each cell type were identified, but
140 in most cases a gradient from veins to bundle sheath to mesophyll cells, or *vice versa*, was
141 detected. Direct comparison to publicly available data from Arabidopsis identified groups of
142 genes encoding aquaporins, proteins allowing sulphur assimilation, jasmonic acid synthesis
143 and also a small number of transcription factor families that showed preferential expression
144 in the bundle sheath of both species. Whilst these findings could be due to evolutionary
145 convergence, the most parsimonious explanation is that the bundle sheath of the last
146 common ancestor of monocotyledons and dicotyledons was specialised in water transport,
147 sulphur assimilation as well as jasmonic acid synthesis, and that members of the Basic
148 Leucine Zipper, C₂H₂-type Zinc Finger, DNA-binding with One Finger, Ethylene Responsive
149 Factor, Hairy-Related Transcription-Factor, GRAS, MYB, Nuclear Factor-YB and Vascular
150 Plant One-Zinc Finger Protein transcription factor families play ancient and conserved roles
151 in this cell type.

152 Results

153 The rice bundle sheath is specialized for transport but also photosynthesis

154 To gain insight into the genetic basis for functional specialization associated with
 155 mesophyll, bundle sheath and veinal cells of rice, laser capture microdissection was used
 156 to isolate RNA from these tissues. Paradermal sections allowed the unambiguous
 157 identification of mesophyll cells (Figure 1A), bundle sheath cells containing large vacuoles
 158 and fewer chloroplasts (Figure 1C) and veins (Figure 1E). To isolate each cell-type with
 159 minimal cross contamination, mesophyll cells were first dissected and captured (Figure 1A,
 160 1B). Sequential capture of bundle sheath cells (Figure 1C, 1D) followed by veinal cells was
 161 then possible (Figure 1E, 1F). RNA was extracted from each tissue and RNA Integrity
 162 Numbers (RIN) ranging from 6.0-7.3 indicated good quality. Strong peaks associated with
 163 ribosomal RNAs of the chloroplast were evident in mesophyll cells (Figure 1G), whereas in
 164 bundle sheath cells they were less abundant (Figure 1H) and in veins they were not
 165 discernable (Figure 1I).

166 Three prime mRNA sequencing was performed and from each cell type, 24-36 million
 167 reads from four or five biological replicates obtained. After processing to remove low quality
 168 reads 13-23 million were quantified against the rice cDNA reference (MSU v7)
 169 (Supplemental Table 1). An average of 10,097, 10,083 and 13,648 transcripts were detected
 170 in each cell type (Supplemental Table 1). Spearman ranked correlation coefficients for gene
 171 expression showed little variation between replicates from each cell type, and that each cell
 172 type exhibited distinct patterns of gene expression (Figure 1J). Principle components
 173 analysis also showed close grouping of biological replicates, and that 46.1% of variance was
 174 associated with the three cell types whilst that a second component separated the bundle
 175 sheath from mesophyll and veins (Figure 1K). To assess the purity of the tissues sampled,
 176 we examined transcript abundance of genes previously reported to be associated with each
 177 cell type. Consistent with these studies, *SUCROSE PHOSPHATE SYNTHASE (SPS1*,
 178 *LOC_Os01g69030)* and aquaporin *PIP2;7 (LOC_Os09g36930)* were preferentially
 179 expressed in mesophyll cells (Chávez-Bárcenas et al., 2000; Li et al., 2008), transcripts
 180 derived from two Tonoplast Monosaccharide Transporters *TMT1* and *TMT2* were detected
 181 in mesophyll, bundle sheath cells and vascular bundles (Cho et al., 2010),
 182 *PHOSPHOENOLPYRUVATE CARBOXYKINASE (PCK1)* was expressed most strongly in

183 bundle sheath and veins (Nomura et al., 2005), and the Sucrose Transporter *SUT1*
184 (LOC_Os03g07480) (Ibraheem et al., 2013; Scofield et al., 2007) was predominately
185 expressed in vascular tissue (Supplemental Figure 1). The strong clustering of each cell-
186 type combined with congruence to previous studies are consistent with the notion that these
187 samples obtained by laser capture microdissection contained relatively little cross-
188 contamination.

189 To quantify the extent to which transcript abundance differed between bundle sheath,
190 mesophyll and veinal cells, we performed differential gene expression analysis using
191 DESeq2 and edgeR (Love et al., 2014; Robinson et al., 2010). This identified 1,919
192 differentially expressed genes between bundle sheath and mesophyll cells, the majority of
193 which (1,173) were more abundant in bundle sheath cells (FDR and adjusted $P < 5\%$)
194 (Supplemental Table 2). Functional enrichment analysis identified three categories over-
195 represented in mesophyll cells containing transcripts linked to photosynthesis, coenzyme
196 metabolism and redox homeostasis (Figure 1L). Eight categories were associated with
197 bundle sheath cells, including solute transport, protein biosynthesis, phytohormone action,
198 nutrient uptake, polyamine metabolism and cellular respiration (Figure 1L). To provide an
199 overview of metabolic specialization in mesophyll and bundle sheath cells, Mapman (Thimm
200 et al., 2004; Schwacke et al., 2019) was used. Consistent with mesophyll cells being
201 specialized for photosynthesis, this indicated that transcripts encoding components of the
202 light reactions, Calvin Benson Bassham cycle and tetrapyrrole biosynthesis were
203 upregulated in mesophyll cells, whilst those associated with many other metabolic
204 processes including cell wall, minor carbohydrate, fatty acid, amino acid and secondary
205 metabolism were more abundant in the bundle sheath (Figure 1M). Quantitative comparison
206 of bundle sheath and veinal cells showed 1,258 and 660 genes were significantly up and
207 down regulated in bundle sheath cells respectively (Supplemental Table 2) and indicated
208 categories associated with the bundle sheath included photosynthesis, carbohydrate
209 metabolism, redox homeostasis, secondary metabolism and solute transport (Supplemental
210 Figure 2A). Mapman outputs confirmed transcripts encoding components of the light-
211 dependent reactions of photosynthesis, as well as the Calvin-Benson-Bassham cycle and
212 photorespiratory pathway were more abundant in bundle sheath than veinal cells
213 (Supplemental Figure 2B). In contrast, transcripts preferential to veins were associated with

processes that included RNA biosynthesis, protein homeostasis, lipid metabolism and solute transport (Supplemental Figure 2A). When mesophyll and veins were compared, a greater number of differentially expressed genes were identified with 1,728 and 2,038 transcripts being more abundant in mesophyll and veins respectively (Supplemental Table 2). The expected preferential expression of photosynthesis-related genes in mesophyll cells was detected, and transcripts associated with protein biosynthesis and cellular respiration were more abundant in veins (Supplemental Figure 2C&D). The greater number of differentially expressed genes between mesophyll and veins is consistent with the correlation and PCA analysis (Figure 1J&K).

To further assess patterns of transcript abundance across all three cell types we clustered genes based on expression. 4155 genes defined as being differentially expressed in the pairwise comparisons above were partitioned into six clusters associated with the cell types in which they were preferentially expressed (Figure 1N). Veins (C_V) had the largest (972) whilst bundle sheath cells (C_{BS}) had the fewest (285) number of preferentially expressed genes. Functional enrichment analysis showed that genes in the mesophyll (C_M) were over-represented in photosynthesis, coenzyme metabolism and solute transport, whilst C_{BS} were enriched in solute transport, enzyme classification, amino acid metabolism and polyamine metabolism (Figure 1O). Genes in C_V were involved in cellular respiration, polyamine metabolism, carbohydrate metabolism and phytohormone action (Figure 1O). $C_{BS\&M}$ contained genes highly expressed in both mesophyll and bundle sheath, and was over-represented in processes including photosynthesis, coenzyme metabolism, carbohydrate metabolism, redox homeostasis, secondary metabolism (Figure 1O). $C_{BS\&V}$ contained genes associated with protein biosynthesis, cellular respiration, and solute transport (Figure 1O). No enriched categories were associated with both the mesophyll and vein cells ($C_{M\&V}$). Consistent with their distinct function, vein and mesophyll clusters showed low overlap, but the most abundant transcripts in bundle sheath cells were also either expressed in veins or mesophyll cells. Overall, and associated with their morphology, the data reveal a gradient in photosynthesis-related transcripts from low in veins to high in mesophyll cells.

Patterning of gene expression in the rice bundle sheath conditions the cells for water transport

To understand the gene classes responsible for enrichment of the transport term in the bundle sheath, we examined expression of major transporter families in each of the six clusters. The family most enriched in C_{BS} was the major intrinsic protein (MIP) group, but multiple genes belonging to the cation diffusion facilitators (CDF) superfamily, major facilitator superfamily (MFS), ion transporter (IT) superfamily, multidrug/oligosaccharidyl-lipid/polysaccharide (MOP) flippase superfamily and amino acid-polyamine-organocation (APC) superfamily were also present (Supplemental Figure 3A).

The MIP group contains genes encoding water channels (aquaporins) including plasma membrane intrinsic proteins (PIPs), tonoplast intrinsic proteins (TIPs) (Sakurai et al., 2005). Transcripts encoding three and two members of the PIP1 and PIP2 families respectively accumulated preferentially in the bundle sheath compared with mesophyll and veinal cells (Figure 2A). Consistent with the deep sequencing data, *in situ* RNA localization of *PIP1.1* and *PIP1.3* generated strong signal associated with the periphery of bundle sheath cells (Figure 2B). Transcripts encoding three tonoplast intrinsic proteins (*TIP1.1*, *TIP1.2* and *TIP2.2*) also accumulated preferentially in bundle sheath cells (Figure 2A), presumably allowing water storage in the large vacuole. Lastly, in addition to these transporters, specific P-type and V-type ATPases that could regulate leaf hydraulic conductance (Grunwald et al., 2021) and establish a proton gradient across plasma and vacuole membranes to power secondary transport were strongly expressed in bundle sheath cells (Supplemental Figure 3B). Taken together, this preferential patterning of multiple aquaporins to the rice bundle sheath suggests an important role for these cells associated with water transport and storage.

The rice bundle sheath preferentially accumulates transcripts associated with sulphur and nitrogen assimilation

Sulphur is an essential element required for both central metabolism and responding to biotic and abiotic stress. Although primarily taken up as sulphate by roots, reduction mainly takes place in leaves (Figure 2C). Prior to activation by ATP sulfurylase (ATPS) to adenosine 5'-phosphosulfate (APS), transport into the cell is mediated by SULTR transporters. APS is reduced into sulfite and sulfide by APS reductase (APR) and sulfite reductase (SIR) respectively, and then incorporated into O-acetylserine via O-acetylserine (thiol)lyase

(OASTL) to generate the amino acid cysteine. Notably, transcripts derived from two highly expressed *SULTR* transporters (*SULTR2;1*, *SULTR3;2*), both *ATPS* genes, *APR*, as well as *SIR* and *OASTL1* were more abundant in bundle sheath cells compared with the mesophyll (Figure 2C). With the exception of transcripts encoding *APR* and *OASTL1* that were most abundant in bundle sheath cells, most were even more highly expressed in veins (Figure 2C). RNA *in situ* hybridization for transcripts encoding *ATPS* and *SIR* showed stronger signals in bundle sheath and veins compared with mesophyll cells (Figure 2D). These results strongly imply that gene expression of the rice bundle sheath cells is conditioned to allow sulphur assimilation.

The data also indicate that gene expression in the bundle sheath is conditioned for synthesis of glutathione, a major sulphur-containing metabolite that plays critical roles in redox homeostasis and heavy metal(loid) detoxification. Biosynthesis of glutathione is catalyzed by γ -glutamylcysteine synthetase (ECS) to generate γ -glutamylcysteine (γ -EC) from glutamate and cysteine, followed by ligation of glycine and γ -EC by glutathione synthetase (Foyer and Noctor, 2011; Hernández et al., 2015). The intermediate γ -EC has to be exported from the plastid by the CRT-like transporter (CLT) (Maughan et al., 2010; Hernández et al., 2015; Yang et al., 2016) to sustain GSH biosynthesis in the cytosol (Pasternak et al., 2007). Interestingly, we found that transcripts encoding *ECS1* were preferentially expressed in bundle sheath and veinal cells (Figure 2C), and that γ -glutamylcysteine transporter *CLT1* accumulated preferentially in bundle sheath cells (Figure 2C). Rice absorbs both arsenate and arsenite by different transporters, but arsenate needs to be reduced into arsenite before it can be detoxified by phytochelatin. *CLT1* has been reported to be critical for rice tolerance to arsenic because it determines phytochelatin biosynthesis and arsenate reduction (Yang et al., 2016). Notably, the arsenate reductase *HAC1;1* also showed preferential expression in the bundle sheath, suggesting that this cell type may play an important role in arsenate reduction and detoxification (Figure 2E).

As with sulphur assimilation, transcripts encoding some of the pathway allowing nitrate reduction were more highly expressed in the bundle sheath and veinal cells compared with mesophyll cells. Interestingly, this included the nitrate transporters *NRT1.4*, *NRT1.1A*, *NRT1.2*, and *NRT2.3*, both nitrate reductases (*NIA1* and *NIA2*), nitrite reductase (*NIR*) and glutamine synthetase (*GS1.1*). Transcripts encoding *NRT2.3*, *NIA1* and *NIA2* were most

abundant in bundle sheath cells, while the rest were also highly expressed in veins. In contrast, transcripts encoding glutamine synthetase (GS2) and glutamate synthase (*Fd-GOGAT*) that allow ammonia assimilation in the chloroplast were preferentially expressed in the bundle sheath and mesophyll relative to veinal cells (Figure 2F, Supplemental Figure 4). These results indicate that gene expression in the rice bundle sheath is also tuned to specialise in nitrate assimilation and amino acid biosynthesis.

The Calvin Benson Bassham cycle, photorespiration and C₄ cycle gene expression

Compared with the mesophyll, bundle sheath cells contain few chloroplasts, and veins only contain rudimentary plastids (Figure 3A). Consistent with this, transcripts encoding most components of photosynthetic electron transport chain were abundant in mesophyll cells, and whilst barely detectable in veins they were clearly expressed in the bundle sheath (Supplemental Figure 5A). Exceptions included one ferredoxin that showed highest transcript abundance in bundle sheath cells, three ferredoxin genes for which transcripts were more abundant in bundle sheath and vein cells compared with the mesophyll, and three homologs of photosystem I subunit PsaA, photosystem II subunit PsbK, and (Supplemental Figure 5A, 5B). In contrast to the primary ferredoxin Fd1 (LOC_Os08g01380) involved in photosynthetic electron transport (He et al., 2020) and primarily expressed in mesophyll cells, four other ferredoxins including the nitrate-inducible Fd (LOC_Os05g37140) (Doyama et al., 1998) were preferentially expressed in bundle sheath and veins indicating that rice bundle sheath cells have optimised reducing power for nitrate reduction, as well as sulphur assimilation. A gradient from high expression in mesophyll to low in veins was observed for most enzymes of the Calvin Benson Bassham cycle. Exceptions included *RUBISCO ACTIVASE* (*RCA*), which was poorly expressed in both bundle sheath and veinal cells, and *RBCS4* which had similar transcript abundance in bundle sheath and mesophyll cells (Figure 3B, Supplemental Figure 5C). The ratio of *RCA* to *RbcS* transcripts was twofold higher in mesophyll compared with bundle sheath cells. With the exception of ADP-glucose pyrophosphorylase subunits (*APL1*, *APS2*), starch synthase (*SSI*) and granule-bound starch synthase (*GBSSII*) which were more abundant in mesophyll cells, transcripts encoding enzymes of starch synthesis were similar in mesophyll and bundle sheath cells (Supplemental Figure 5D). Although transcripts associated the glucose 6-

phosphate/phosphate translocator (GPT1, GPT2-1) were more abundant in bundle sheath than mesophyll cells, their abundance was low compared with the triose phosphate/phosphate translocator (TPT1) (Supplemental Figure 5D). Together, these data indicate that gene expression in the bundle sheath is set up to favour starch synthesis using products of the Calvin Benson Bassham cycle.

Most transcripts encoding proteins of photorespiration showed strong expression in the mesophyll, weaker expression in the bundle sheath and very poor expression in veins (Figure 3C, Supplemental Figure 5E). The only exceptions were genes with very low absolute levels of expression that were preferential to veins, and which included one *L-GLYCINE DECARBOXYLASE* subunit (*GDCL2*), one *SERINE HYDROXYMETHYLTRANSFERASES* (*SHMT4*), one *PHOSPHOGLYCOLATE PHOSPHATASE* (*PGP2*) and two *GLYCINE DECARBOXYLASE COMPLEX H* subunits (*GDCH2&3*) (Figure 3C, Supplemental Figure 5E).

It has previously been reported that vascular tissue of C_3 plants possesses high activities of some enzymes associated with C_4 photosynthesis, including all three C_4 acid decarboxylases and pyruvate, orthophosphate dikinase (Hibberd and Quick, 2002; Brown et al., 2010; Shen et al., 2016). However, to our knowledge it has not been possible to delimit these activities to the specific cells associated with the vascular tissue. To investigate this, we assessed abundance of transcripts encoding core enzymes of the C_4 cycle in veins, bundle sheath and mesophyll cells of rice (Figure 3D). Transcripts of *CARBONIC ANHYDRASE* (*CA*) were more abundant in the mesophyll compared with bundle sheath and vein cells, a pattern consistent with that required for C_4 photosynthesis. Transcripts encoding PEP carboxylase (*PEPC*), NADP-MALIC DEHYDROGENASE (*NADP-MDH*), ASPARTATE AMINOTRANSFERASE (*AspAT*), NAD-MALIC ENZYME (*NAD-ME*), ALANINE AMINOTRANSFERASE (*AlaAT*), PYRUVATE, ORTHOPHOSPHATE DIKINASE (*PPDK*) and the *PPDK REGULATORY PROTEIN* (*PPDK-RP*) showed no significant difference in abundance between mesophyll and bundle sheath cells. However, transcripts encoding AMP KINASE (*AMK*) and PYROPHOSPHORYLASE (*PPASE*), which allow the *PPDK* reaction and APS synthesis by ATP sulfurylase to proceed, showed higher abundance in the bundle sheath than both veins and mesophyll cells (Figure 3D), suggesting that the activity of *PPDK* might be higher in this cell type. Notably, two C_4 acid decarboxylases *NADP-*

369 *DEPENDENT MALIC ENZYME (NADP-ME)* and *PHOSPHOENOLPYRUVATE*
 370 *CARBOXYKINASE (PCK)* showed a gradient in transcript abundance from veins, to bundle
 371 sheath to mesophyll cells. These data indicate that high activity of these C₄ decarboxylases
 372 in vascular bundles of rice (Shen et al., 2016) is likely caused by their expression in the vein
 373 rather than bundle sheath cells. Overall, C₄ genes could be partitioned into three main
 374 groups, either showing a strong negative gradient (*CA* and *PEPC*) from mesophyll to vein
 375 cells, a strong positive gradient (*NADP-ME*, *NAD-ME*, *PCK* and *AlaAT*) or a tendency to be
 376 most strongly expressed in the bundle sheath (*NADP-MDH*, *AspAT*, *PPDK*, *PPDK-RP*, *AMP*
 377 and *PPase*) (Figure 3D). This finding is consistent with them being part of multiple gene
 378 regulatory networks in the ancestral C₃ state that pattern their expression across these cell
 379 types.

380

Transcription factors and their cognate *cis*-elements associated with the rice mesophyll, bundle sheath and vein

383 To gain insight into the regulatory architecture associated with the three cell types, we
 384 identified transcription factors in each of the six clusters and designated these as TF_M, TF_{BS},
 385 TF_V, TF_{BS&M}, TF_{BS&V} and TF_{M&V}. From a total of 201 differentially expressed transcription
 386 factors, over 30% of them (66) showed vein-specific expression (TF_V, Figure 4A), including
 387 families such as bZIP, bHLH, G2-like, MYB-related and Dof (Supplemental Figure 6A). ERF,
 388 HD-ZIP and MYB-related were the most abundant transcription factor families in mesophyll-
 389 specific cluster TF_M (Figure 4A, Supplemental Figure 6A). Transcription factors known to
 390 regulate chloroplast biogenesis and photosynthesis were most abundant in mesophyll cells,
 391 but also detected in the bundle sheath. This included GNC (LOC_Os06g37450) in TF_M, and
 392 GLK1 (LOC_Os06g24070), GLK2 (LOC_Os01g13740) and CGA1 (LOC_Os02g12790) in
 393 TF_{BS&M}. Thus, consistent with the chloroplast complement of each cell type (Figure 3A),
 394 these transcription factors found in TF_{BS&M} were less abundant in bundle sheath compared
 395 with mesophyll cells (Supplemental Figure 6C).

396 The bundle sheath specific cluster contained only ten genes that derived from families
 397 such as the ERFs, bZIPs and bHLHs (Figure 4A, Supplemental Figure 6A). Transcription
 398 factors abundant in both bundle sheath and vein belonged to the MYB, ERF, bZIP and
 399 WRKY families, whilst CO-like, C₂H₂, and G2-like families were abundant in the mesophyll

400 and bundle sheath cluster (Figure 4A, Supplemental Figure 6A). The ZF-HD, G2-like, DBB,
401 MIKC-MADS, and Dof families were significantly enriched in the vein-specific cluster, CO-
402 like transcription factors were over-represented in $TF_{BS\&M}$, and Dof, MYB, HRT-like and VOZ
403 transcription factors were over-represented in $TF_{BS\&V}$ (Supplemental Figure 6B).

404 To determine whether DNA motifs known to be bound by transcription factors were
405 associated with genes preferential to each cell type, we performed motif enrichment
406 analyses using the FIMO and AME tools (Bailey et al., 2015). Although the two approaches
407 differ in their statistical testing they returned broadly similar estimates for the number of
408 enriched motifs. AME found the highest number of motifs in C_{BS} and $C_{BS\&M}$ ($n=44$ and 91
409 respectively), whilst FIMO identified the most enriched motifs in C_V and $C_{BS\&M}$ ($n=86$ and
410 100 respectively) (Supplemental table 4 and 5). Both found either few, or no enriched motifs
411 in the $C_{BS\&V}$ (AME=11; FIMO=5), C_M (AME=0; FIMO=3) or $C_{M\&V}$ (both 0) clusters
412 (Supplemental table 4) suggesting that genes making up these three clusters are
413 heterogeneously regulated.

414 As many DNA motifs share considerable sequence similarity, and closely related
415 transcription factors can bind to similar motifs, we collapsed known DNA binding motifs from
416 closely related transcription factors into forty-one groups (Figure 4B). Twenty-three of these
417 groups were enriched or depleted in a specific cluster associated with transcript abundance
418 in the three cell types. Although many were found in multiple clusters, there were no
419 examples where a particular motif was statistically enriched or depleted in all six clusters
420 (Figure 4B). C_M contained the fewest enriched motifs ($n=4$), only one of which (ARF2) was
421 uniquely enriched in this cluster (Figure 4B). HD-Zip I and LHY motifs were unique to C_{BS} ,
422 but all other enriched motifs in the bundle sheath were also found in C_V suggesting overlap
423 in the regulation of gene expression between bundle sheath and vein cells (Figure 4B). C_V
424 contained the most enriched motifs ($n=29$) of which almost half of these were unique to this
425 cluster and included BIRD, Dof, GATA, HD-Zip II, PLT, FAR1 and FHY3 motifs (Figure 4B).

426 We next investigated whether transcription factors and their cognate DNA binding sites
427 were associated with the six clusters that had been defined by patterns of gene expression
428 across mesophyll, bundle sheath and veinal cells. As relatively few transcription factors from
429 rice have had their DNA binding characteristics determined, we first used protein homology
430 (BLASTP bit-score > 100) to link them with transcription factors for which DNA binding data

are available. Of the mesophyll specific transcription factors, although five are likely to bind motifs enriched in the cistrome of bundle sheath and mesophyll cells (Figure 4C), none were associated with motifs enriched only in the mesophyll cistrome. In contrast, three bundle sheath specific transcription factors coincided with motifs enriched in the BS-specific cistrome. Vein-specific transcription factors showed the greatest convergence with enrichment in their cognate DNA binding sites with fifteen mapping to the vein specific cistrome. This included transcription factors predicted to bind BIRD, Dof, AP2ERF, MYB, MADS and bHLH motifs (Figure 4C). Overall, this approach identifies families of transcription factors preferentially expressed in cell types in which their cognate motifs were over-represented in the cistrome of that cell type. We regard these transcription factors as strong candidates for patterning gene expression across these cell-types.

Conserved patterning of gene expression in bundle sheath cells from rice and Arabidopsis

Bundle sheath cells are present in both monocotyledons and dicotyledons. Previous analysis has compared transcripts loaded onto ribosomes in the bundle sheath of Arabidopsis with those from whole leaves. In so doing, it was concluded that the bundle sheath in Arabidopsis is important for sulphur and glucosinolate metabolism as well as trehalose synthesis (Aubry et al 2014b). Functional enrichment of bundle sheath and mesophyll preferential genes from rice and Arabidopsis were compared. Consistent with the mesophyll representing the major cell type conducting photosynthesis, gene sets common to the mesophyll from the two species included transcripts important for the photosynthetic electron transport chain and photophosphorylation, the Calvin Benson Bassham cycle, photorespiration, tetrapyrrole biosynthesis, chloroplast redox homeostasis, organelle protein biosynthesis and terpenoid biosynthesis (Figure 5A). In contrast, only four categories of genes were enriched in the bundle sheath from both species, namely carrier-mediated transport, sulphur assimilation, amino acid biosynthesis and jasmonic acid action (Figure 5A).

We next identified genes from both species likely to be descended from a common ancestor and placed these into orthogroups (Emms and Kelly, 2019). Genes were assigned to 10,665 orthogroups and the extent to which each was preferentially expressed in

mesophyll or bundle sheath cells determined. Genes from 380 orthogroups were preferentially expressed in mesophyll cells of rice and whole leaves of Arabidopsis, whilst genes from 293 orthogroups were shared by bundle sheath of both species (Figure 5B), in both cases a greater overlap than would be expected by chance (Fisher's exact test, $p < 2.2e^{-16}$ for both bundle sheath and mesophyll). However, the odds ratio that is also generated from the Fisher's exact test indicated that there was a greater degree of overlap between the mesophyll of rice and whole leaves of Arabidopsis than between their bundle sheath cells (odds ratios were 5.32 and 2.06 for mesophyll and bundle sheath respectively). We thus conclude that the expression of orthologs has diverged more in the bundle sheath than the mesophyll of these species. Specific orthogroups associated with the mesophyll in both species were mainly associated with photosynthesis but also included eleven orthogroups of transcription factors, one of which included the transcription factor *GLK1* (OG0002393) (Supplemental table 5) that is known to activate photosynthesis related gene expression (Waters et al., 2008; Waters et al., 2009). Although fewer orthogroups were common to the bundle sheath in rice and Arabidopsis, overlap included three groups encoding aquaporins (PIP1 - OG0000721, PIP2 -OG0000170 and TIP1 - OG0002706), as well as orthologs involved in sulphate transport (OG0005139 and OG0013047), sulphate assimilation (OG0001744, OG0003507 and OG0009859) and jasmonic acid biosynthesis (OG0002794, OG0005175, OG0010544) (Figure 5C-F, Supplemental Figure 7, Supplemental Table 5). We compared bundle sheath preferentially expressed genes of rice and Arabidopsis with the bundle sheath marker genes identified through single cell sequencing of Arabidopsis (Kim et al., 2021). As sequencing depth from single celled sequencing is not as great, fewer transcripts were detected (Supplemental Figure 8A). However, 41 orthogroups were identified containing at least one gene preferential to the bundle sheath in all three studies, including two sulphate transporters (OG0005139 and OG0013047), ATP sulfurylase (OG0001744), and allene oxide cyclase (OG0005175). To investigate whether these genes are associated with bundle sheath cells in a broader range of species, we assessed transcript abundance of consensus orthogroups associated with water transport, sulphur assimilation, jasmonic acid biosynthesis as well as nitrate reduction in publicly available data from *Panicum virgatum* (Rao et al., 2016), *Sorghum bicolor* (Emms et al., 2016), *Setaria italica* (John et al., 2014) and *Zea mays* (Chang et al., 2012) as well as the C₄ dicotyledon

493 *Gynandropsis gynandra* (Aubry et al., 2014a). Except for nitrate reduction genes that are
494 more highly expressed in mesophyll cells, this indicated that in the majority of these species,
495 genes in each of orthogroup are preferentially expressed in bundle sheath cells
496 (Supplemental Figure 9). Taken together, the data strongly suggest that sulphur assimilation,
497 jasmonic acid biosynthesis, and water transport represent ancestral functions associated
498 with the bundle sheath derived from the last common ancestor of monocotyledons and
499 dicotyledons.

500 We next wished to identify if any transcription factors were preferentially expressed in
501 bundle sheath cells of both species. The rice ortholog of the key regulator of the sulphur
502 starvation response *SULFUR LIMITATION1* (Maruyama-Nakashita et al., 2006), *OsEIL3*
503 (LOC_Os09g31400) showed low expression, but consistent with *Arabidopsis*, was more
504 strongly expressed in bundle sheath compared with mesophyll cells (Supplemental Figure
505 7). Fourteen transcription factor orthogroups were identified containing at least one ortholog
506 in both species that was preferential to the bundle sheath (Figure 5G). These included three
507 Basic Leucine Zipper (OG0000499, OG0002563, OG0005764), C₂H₂-type Zinc Finger
508 (OG0000197), DNA-binding with One Finger (OG0002717), Ethylene Responsive Factor
509 (OG0000957), two GRAS (OG0000296, OG0005593), Hairy-Related Transcription-Factor
510 (OG0005688), three MYB (OG0000100, OG0001367 and OG0001681), Nuclear Factor-YB
511 (OG0000404), and Vascular Plant One-Zinc Finger Protein (OG0002457) transcription
512 factor families (Figure 5G, Supplemental Figure 7, Supplemental table 6). These data imply
513 purifying selection has acted to maintain expression of these transcription factors in the
514 bundle sheath since these species diverged from their last common ancestor prior to the
515 divergence of the dicotyledons and monocotyledons.

516 Discussion

517 Patterning of photosynthesis gene expression between cell-types in the rice leaf

518 Separation of protoplasts followed by cell sorting has provided significant insight into gene
 519 expression in specific cells of roots (Birnbaum et al., 2003; Brady et al., 2007). In leaves,
 520 the ability to separate mesophyll and bundle sheath cells from C₄ plants (Kanai and Edwards,
 521 1973; Edwards and Black, 1971; Moore et al., 1984) has allowed similar levels of insight into
 522 these cells, but leaf cells from C₃ species have been more challenging to isolate. As a
 523 consequence, although major cell-types of leaves such as the mesophyll, phloem, xylem
 524 and guard cells have well defined roles, others such as the bundle sheath are relatively
 525 poorly understood (Leegood, 2008). In Arabidopsis the bundle sheath is thought to play
 526 important roles in regulating hydraulic conductance, substrate transport and storage (Shatil-
 527 Cohen et al., 2011; Sade et al., 2014; Griffiths et al., 2013). Analysis of mRNAs resident on
 528 ribosomes showed that patterns of gene expression in the Arabidopsis bundle sheath are
 529 conditioned to facilitate sulfur metabolism and glucosinolate biosynthesis (Aubry et al.,
 530 2014b). Although, suppression of chlorophyll synthase in veinal tissue including bundle
 531 sheath cells of Arabidopsis reduced photosynthesis, growth and fitness (Janacek et al.,
 532 2009), the importance of the bundle sheath itself was not defined. Moreover, in C₃
 533 monocotyledons, the group containing many of our most important crops, little is known
 534 about the function of bundle sheath cells. To address this, we defined gene expression in
 535 the rice bundle sheath as well as mesophyll and vein cells using laser capture
 536 microdissection coupled with mRNA sequencing. Contrary to our expectations, this indicated
 537 that although bundle sheath cells of rice contain few chloroplasts (Wang et al., 2017; Sage
 538 and Sage, 2009), transcripts encoding components of the photosynthetic electron transport
 539 chain and the Calvin Benson Bassham cycle were clearly expressed in rice bundle sheath
 540 cells. The consequence of this finding is that expression of photosynthesis genes represents
 541 a continuum from low to medium to high in veins, bundle sheath and mesophyll cells
 542 respectively. Whether this is also the case in Arabidopsis remains to be determined. There
 543 is significant interest in activating photosynthesis the bundle sheath of rice such that
 544 mesophyll and bundle sheath cells could be engineered to carry out C₄ photosynthesis
 545 (Wang et al., 2017; Sage, 2004). The finding that photosynthesis gene expression in the
 546 bundle sheath resembles the mesophyll more than veins indicates that it needs to be re-

tuned rather than completely re-programmed to achieve this demanding aim.

Expression of C₄ genes in the C₃ rice leaf

Vascular bundles of C₃ plants are known to carry out C₄-like metabolism via high activities of C₄ acid decarboxylases and PPDK to make use of C₄ acids present in the transpiration stream (Hibberd and Quick, 2002; Brown et al., 2010; Shen et al., 2016). To date, it has not been possible to define whether high activity of these C₄ enzymes in C₃ plants is associated with their expression in vein or bundle sheath cells. The analysis of rice we present here shows that transcripts encoding two C₄ acid decarboxylases, NADP-ME and PCK, were more abundant in bundle sheath compared with mesophyll cells, but in both cases, transcripts were even more strongly expressed in veins. These findings suggest that the high activity of NADP-ME and PCK in vascular tissue is primarily caused by expression in veins rather than the bundle sheath. Thus, the gradient in expression of genes encoding these C₄ acid decarboxylases from high in veins to low in mesophyll is the opposite of photosynthesis genes in these cell types. Whether this is also the case in C₃ dicotyledons such as Arabidopsis remains to be determined.

Although there was no significant difference in *PPDK* transcript abundance between bundle sheath and mesophyll cells, we detected greater abundance of transcripts encoding AMK and PPase which carry out ancillary reactions allowing the PPDK reaction to proceed. As activity of PPDK is higher in vascular strands of rice than in mesophyll cells (Shen et al., 2016), it is therefore possible that this is caused by greater activity of AMP and PPase. PPDK is important for nitrogen recycling in Arabidopsis and tobacco (Taylor et al., 2010) and so these data also suggest that this is likely the case in rice. Although *CA* and *PEPC* transcripts in rice were most abundant in mesophyll cells, the particular isoforms these transcripts encode are predicted to generate proteins localized to the chloroplast (Masumoto et al., 2010; Chen et al., 2017). In C₄ leaves, both of these proteins are cytosolic (Hatch, 1987; Hatch and Burnell, 1990). Overall, these results suggest that at least three modifications to expression of these genes would be required to build a C₄ cycle into rice – amplifying the existing expression of C₄ acid decarboxylases in bundle sheath cells, repositioning *CA* and *PEPC* proteins such that they reside in the cytosol, and expressing AMK and PPase in the mesophyll for effective PPDK activity.

578

579 **The rice bundle sheath is conditioned to allow water transport, sulphate and nitrate** 580 **metabolism**

581 Our analysis of the rice bundle sheath indicates that gene expression is poised to allow
582 photosynthesis in this cell type. This finding is consistent with the fact that although during
583 early leaf development plastids of the rice bundle sheath contain significant amount of starch,
584 as the leaf matures these plastids develop into chloroplasts (Miyake, 2016). Transcripts
585 patterns in the bundle sheath were consistent with starch being synthesized from carbon
586 skeletons derived from the Calvin Benson Bassham cycle with strong expression of *PGI*,
587 *PGM*, *UGP1*, *SSIIB* and *SSIII-1*, and low expression of hexose phosphate transporters
588 (*GPT1*, *GPT2-1*) compared with the *TPT1*. Our analysis also indicates that the rice bundle
589 sheath is patterned to facilitate water transport and storage, sulphate and nitrate assimilation.
590 It was notable that most of highly expressed aquaporins were preferentially expressed in
591 bundle sheath cells, and included members of the PIP1, PIP2, TIP1 and TIP2 subfamilies.
592 It has been reported that water transport activity and plasma membrane localization of PIP1
593 requires interaction with PIP2 (Fetter et al., 2004; Zelazny et al., 2007) and so their co-
594 expression in the bundle sheath is compatible with efficient water transport and storage in
595 this cell type. Notably, analysis of publicly available data indicates that strong expression of
596 *PIP1* and *PIP2* in the bundle sheath is also found in the C₄ grasses *Panicum virgatum*,
597 *Sorghum bicolor*, *Setaria italica* and *Zea mays* as well as the C₄ dicotyledon *Gynandropsis*
598 *gynandra* (Rao et al., 2016; Emms et al., 2016; John et al., 2014, Chang et al., 2012; Aubry
599 et al., 2014a), implying that bundle sheath plays important role in water transport in the
600 common ancestor of monocotyledons and dicotyledons (Supplemental Figure 9).

601 Enzymes of sulphur assimilation preferentially accumulate in the bundle sheath of C₄
602 grasses and C₃ dicotyledon *Arabidopsis* (Gerwick et al., 1980; Passera and Ghisi, 1982;
603 Burnell, 1984; Schmutz and Brunold, 1984; Burgener et al., 1998; Aubry et al., 2014b;
604 Chang et al., 2012; John et al., 2014; Emms et al. 2016; Rao et al., 2016), but this is not the
605 case in wheat and the C₄ dicotyledons *Flaveria* and *Gynandropsis gynandra* where ATPS
606 and APR showed similar activity and/or transcript abundance in bundle sheath and
607 mesophyll cells (Schmutz and Brunold, 1984; Kopriva et al., 2001; Aubry et al., 2014a). We
608 found that transcripts associated with sulphur assimilation were restricted or preferentially

609 localized to bundle sheath cells of rice, which is consistent with expression patterns in
610 Arabidopsis (Aubry et al., 2014b). This was also true for genes indirectly associated with
611 sulphur assimilation such as the PPase that increases the rate of the ATP sulphurylase
612 reaction. Several evolutionary drivers for localization of sulphur assimilation in the bundle
613 sheath of C₄ grasses have been discussed including co-localisation with photorespiration as
614 a source of serine for cysteine synthesis and protection of the reaction intermediates from
615 oxidation by Photosystem II derived oxygen (Kopriva and Koprivova, 2005).

616 In Arabidopsis glucosinolate synthesis is controlled by a MYC-MYB transcription factor
617 module that has recently been shown to pattern gene expression to the bundle sheath
618 (Dickinson et al 2020). It seemed likely that localisation of glucosinolates to the bundle
619 sheath of Arabidopsis led to upregulation of ATPS and APR in these cells. However, the
620 preferential expression of genes associated with sulphur assimilation in the rice bundle
621 sheath that does not synthesise glucosinolates indicates a much more ancient role for the
622 bundle sheath in sulphur assimilation. It also removes the proposed connection between C₄
623 photosynthesis and localisation of sulphur assimilation in the bundle sheath. More broadly,
624 it appears that the bundle sheath is conditioned for sulphur assimilation in all species
625 analysed to date. In plants such as wheat, *Flaveria* and *Gynandropsis*, mesophyll cells are
626 also used (Schmutz and Brunold, 1984; Koprivova et al., 2001; Aubry et al., 2014a), but in
627 others the pathway appears to be restricted to the bundle sheath (Supplemental Figure 9).
628 It is not clear whether the ancestral state was for gene expression allowing sulphur
629 assimilation to be associated with bundle sheath and mesophyll cells or whether in some
630 species expression in mesophyll cells has been gained. Moreover, a full complement of
631 transporters and enzymes for nitrate reduction were found to preferentially be expressed in
632 rice bundle sheath cells, other enzymes that serve to provide carbon skeleton for amino acid
633 metabolism such as PEPC (Masumoto et al., 2010) and PPDK (Taylor et al., 2010) also
634 showed high expression in these cells. The compartmentation of nitrate and ammonia
635 assimilation gene expression between bundle sheath and mesophyll cells of rice contrasts
636 with their spatial separation of C₄ grasses and *Gynandropsis gynandra*, where nitrate
637 reduction predominately takes place in the mesophyll, and ammonia assimilation occurs in
638 the bundle sheath (Supplemental Figure 9). The strong expression of genes associated with
639 nitrate reduction to our knowledge has not previously been reported in any system.

Furthermore, it has implications for our understanding of transitions associated with the evolution of the C₂ and C₄ pathways, since modelling revealed that early changes to C₂ metabolism should induce an imbalance in nitrogen metabolism between bundle sheath and mesophyll cells (Mallmann et al., 2014). It has been proposed that this imbalance could be counteracted by upregulating genes associated with C₄ photosynthesis (Mallmann et al., 2014). Indeed, transcripts for key genes of nitrate assimilation are consistently more strongly expressed in mesophyll cells of C₄ plants (Supplementary Figure 9, Rao et al., 2016; Emms et al., 2016; John et al., 2014, Chang et al., 2012; Aubry et al., 2014a). The consequences and the drivers of gene expression associated with nitrate reduction being focused on the bundle sheath in rice are unknown. However, this co-localisation of sulphur and nitrate assimilation in the rice bundle sheath was associated with preferential expression of specific ferredoxins that have previously only been implicated in differences in electron transfer reactions allowing CO₂ or mineral nutrient reduction in shoots and roots (Yonekura-Sakakibara et al., 2000).

The ancestors of rice and Arabidopsis diverged ~140 million years ago (Chaw et al., 2004). Despite this timescale significant overlap in cell-specific gene expression in bundle sheath and mesophyll cells from these two C₃ species was detected. We conclude that the two cell types have retained specific roles over this extended time. Less overlap was found in bundle sheath cells suggesting that the role of the bundle sheath has evolved faster than the mesophyll. However, despite these apparent changes to gene expression in the bundle sheath, some genes associated with water transport and jasmonic acid biosynthesis were preferential to bundle sheath cells in majority of species. In contrast to sulphur assimilation, we therefore propose that these processes represent ancestral functions derived from the last common ancestor of monocotyledons and dicotyledons. Moreover, a small number of transcription factors were strongly bundle sheath preferential in both species, and so we propose that these regulators underpin ancestral and conserved functions of bundle sheath cells in dicotyledons and monocotyledons.

667 **Materials and methods**

668 **Plant growth condition and sample preparation**

669 The temperate rice (*Oryza sativa* ssp. *japonica*) Kitaake was germinated and grown in a
 670 mixture of 1:1 topsoil and sand for 2 weeks in a controlled environment growth room.
 671 Temperature was set to 28°C day, 25°C night, and photoperiod at 12 hr light and 12 hr dark.
 672 Relative humidity was 60% and photon flux density 300 $\mu\text{mol m}^{-2} \text{s}^{-1}$. 1cm sections from the
 673 middle of the fourth fully expanded leaves were sampled 4 hours after dawn. Leaf tissue
 674 was fixed and embedded into Steedman's wax as described previously (Hua and Hibberd,
 675 2019) with minor modifications. For example, rice leaves were fixed in 100% (v/v) acetone
 676 on ice for 4 hours and before embedding tissue infiltrated with 100% Steedman's wax at
 677 37 °C overnight. For laser capture microdissection (LCM) and RNA extraction, paradermal
 678 sections of 7 microns were prepared with a microtome and mounted on PEN membrane
 679 slides. Prior to LCM Steedman's wax was removed by incubating slides in 100% (v/v)
 680 acetone for 1 min. LCM was performed on an Arcturus Laser Capture Microdissection
 681 platform, with isolated cells being collected on CapSure Macro Caps and RNA extracted
 682 using the PicoPure RNA Isolation Kit with on-column DNaseI treatment. RNA quality and
 683 concentration were analyzed using an Agilent Bioanalyser RNA 6000 Pico assay.

684 For RNA *in situ* hybridization, middle sections from the fourth fully expanded leaf were
 685 sampled and fixed overnight using FAA fixative (50% (v/v) ethanol, 5% (v/v) acetic acid and
 686 3.7% (v/v) formaldehyde. They were then dehydrated through an ethanol series of 50%,
 687 70%, 85%, 95% and 100% (v/v) and embedded into Steedman's wax as described
 688 previously (Hua and Hibberd, 2019). 8 micron thick sections were obtained on a microtome
 689 and mounted onto Superfrost plus slides. Tissue pretreatment, hybridization and colour
 690 development were performed as previously described (Jackson, 1992), except that
 691 Steedman's wax was removed by incubating slides in 100% ethanol for 5 mins twice.

692 **Library preparation, RNA sequencing and data processing**

694 20 ng of bundle sheath, mesophyll and veinal RNA was used as input for the Quantseq
 695 3' mRNA-seq library preparation kit (Lexogen, Moll et al., 2014) according to the
 696 manufacturer's instructions. Libraries were sequenced using Nextseq 500 sequencer to
 697 produce single-ended 150-bp reads for each sample. The leading and tailing 10-bp of

Quantseq reads were trimmed and reads with a quality score less than 20 and shorter than 50-bp were removed using BBDuk (Bushnell, 2015). Transcript abundance was determined after the remaining reads were quantified using Salmon version 0.8.2 (Patro et al., 2017) against the rice cDNA reference (MSU version7), '--noLengthCorrection' flag was used to disable length correction (Corley et al., 2019). Gene level abundance (TPM, transcript per million) and counts were summarized using tximport version 1.10.1 (Soneson et al., 2016) and gene level counts were used for downstream differential gene expression analysis using DESeq2 version 1.22.2 (Love et al., 2014) and edgeR version 3.24.3 (Robinson et al., 2010). Poorly expressed genes with row sum of TPM < 1 in three samples were excluded, resulting in 15168 genes; and a Benjamini-Hochberg corrected *P*-value (DESeq2) and False discovery rate (edgeR) of <0.05 were used to define differentially expressed genes.

To quantify the abundance of transcript associated with ribosomes of Arabidopsis BS and total leaf samples, reads were obtained at Sequence Read Archive (SRA, <https://www.ncbi.nlm.nih.gov/sra/>) Bioproject accession PRJEB5030 and quantified using Salmon version 0.8.2 against the Arabidopsis TAIR10 reference (Berardini et al., 2015) using default parameters, gene level abundance (TPM, transcript per million) was summarized using tximport version 1.10.1 (Soneson et al., 2016). Differentially expressed genes were defined according to Aubry et al., 2014b.

716

717 **Gene clustering, over-representation analysis and gene expression visualisation**

Consensus differentially expressed genes identified using DESeq2 and edgeR in each pairwise comparison were used for gene expression clustering and functional enrichment analysis. Gene expression clustering were performed with K-Means method (Hartigan and Wong, 1979) using log₂ transformed quantile normalized TPM (transcript per million). For functional enrichment, rice proteome (MSU version7) were firstly annotated by Mapman categories using Mercator 4 (Schwacke et al., 2019), over-representation analysis of Mapman categories were performed using Fisher's exact test and using the expressed 15168 genes as background with false discovery rate (FDR) < 0.1. Rice transcription factor families were annotated according to PlantTFDB v5.0 (Jin et al., 2017), using all expressed transcription factor as background, over-representation test were performed using Fisher's exact test with a cutoff FDR < 0.1. Significant Mapman categories and transcription factor families were

plotted using ggplot2 (Wickham, 2016). Heatmap of gene expression clusters and pathways were plotted using ComplexHeatmap package (Gu et al., 2016). Transcript abundance were presented in boxplot of TPM (transcript per million) using default settings of hinge and whisker in ggplot2.

Motif enrichment analysis

Leaf DNase I hypersensitive sites (DHS) (Zhang et al., 2012) within 2000-bp of the genes were extracted and used as input to the AME tool (<http://meme-suite.org/tools/ame>, Bailey et al., 2015; McLeay and Bailey, 2010) using default settings and a custom background of all DHS regions within 2000bp of gene loci as well as for FIMO command line tool (Grant et al., 2011) using default settings. The meme format Jaspar Plant Non-Redundant Motif Database (Fornes et al., 2020, <http://jaspar2018.genereg.net/downloads/>) was used for both methods. FIMO scanning identified significant matches to known motifs and the frequencies of these sites were statistically tested for enrichment or depletion using permutation testing with the regioneR R package (Gel et al., 2016). For permutation testing, all DHS regions within 2000bp of genes were used as a background for random subsampling and observed frequencies were statistically tested against the random subsampling distributions. As the Jaspar motif database contains many motifs with high similarity predicted to be bound by closely related transcription factors, we grouped similar motifs together that showed a higher than 70% overlap in predicted target sites within the DHS background sequences representing motifs that were deemed highly redundant. The most strongly enriched individual motif from each such group was used as a representative value for the group and plotted using ggplot2 (Wickham and Sievert, 2016). In order to predict which *O. sativa* transcription factors might bind predicted motifs, we mapped rice transcription factors to their best match motif through protein homology with the Jaspar Non-Redundant Motif database proteins (Fornes et al., 2020). The best scoring BLASTP match was used as long as bit-score was greater than 100 to avoid spurious assignments to motifs from other TF families. For all transcription factors found in the 6 clusters, those that were mapped to an enriched motif from any of the 6 clusters was plotted using the single motif enrichment scores, rice transcription factors were annotated according to funRiceGenes database (Yao et al., 2018).

760

761 **Defining orthogroups in rice and Arabidopsis**

762 Protein sequences of primary transcripts of *Oryza sativa* (v7.0), *Arabidopsis thaliana*
763 (TAIR10), *Gynandropsis gynandra* (Aubry et al., 2014a), *Brachypodium distachyon* (v3.2),
764 *Panicum virgatum* (v1.1), *Sorghum bicolor* (v1.4), *Setaria italica* (v2.1), *Zea mays* (Ensembl-
765 18) and *Marchantia polymorpha* (v3.1) obtained from Phytozome v12.1 (Goodstein et al.,
766 2012) were clustered into orthogroups using Orthofinder version 2.4.1 (Emms and Kelly,
767 2019) with the default parameters, orthogroups that contain both rice and Arabidopsis genes
768 were used for comparative analysis in the two species. Resolved gene trees of orthogroups
769 were visualized using R package ggtree version 2.41 (Yu, 2020).

770

771 **Code and data availability**

772 Code associated with this manuscript and the underlying data required to generate plots
773 are available in the Github repository: [https://github.com/hibberd-](https://github.com/hibberd-lab/Hua_et_al_Kitaake_LCM)
774 [lab/Hua_et_al_Kitaake_LCM](https://github.com/hibberd-lab/Hua_et_al_Kitaake_LCM). All other data are available on request.

775

776 **Accession numbers**

777 Raw sequencing data are deposited in the National Center for Biotechnology Information
778 under BioProject ID PRJNA702624, BioSample ID SAMN17976370 - SAMN17976383.

779

780 **Acknowledgements**

781 The work was funded by C₄ Rice project grant from The Bill and Melinda Gates
782 Foundation to the University of Oxford (2015–2019), European Research Council Grant
783 694733 Revolution, BBSRC Grants BBP0031171 and BBL014130 to JMH. Research in SK's
784 lab is funded by the Deutsche Forschungsgemeinschaft (DFG) under Germany's Excellence
785 Strategy – EXC 2048/1 – project 390686111.

786

787 **Competing interests**

788 The authors declare that they have no competing interests.

789 **Figure legends**

790 **Figure 1. RNA was isolated from mesophyll, bundle sheath and veinal cells of rice**
 791 **using laser capture microdissection.** Each cell-type was identified and then sequentially
 792 removed from paradermal sections prior to RNA quality being assessed. (A&B)
 793 Representative image of mesophyll cells outlined with red dash line (A) that are cut with a
 794 UV laser and then captured with an infrared laser and placed on a cap (B). (C&D) Bundle
 795 sheath cells (C) were then cut and captured (D). (E&F) Lastly, veinal cells (E) were cut and
 796 captured (F). Scale bars represents 50 microns. (G-I) Representative RNA profiles from
 797 microdissected mesophyll (G), bundle sheath (H) and veinal cells (I). Peaks from the
 798 cytosolic 25S, 18S and 5S ribosomal RNAs were detected in all cell types. Chloroplastic
 799 ribosomal RNAs (CP rRNA) were clearly detectable in mesophyll and bundle sheath cells.
 800 (J) Spearman ranked correlations of $\log_2$ transformed TPM (transcripts per million) indicates
 801 little variation between biological replicates from each cell type, and distinct patterns of gene
 802 expression in each cell type. (K) Principal component analysis of normalised counts after
 803 variance-stabilizing transformation showing that cell type accounted for 46.1% of the
 804 variance detected in the data. (L) Primary Mapman categories associated with differentially
 805 expressed genes in bundle sheath and mesophyll cells. Terms were defined using Fisher's
 806 exact test (False discovery rate, $FDR < 0.1$), colour scale represents negative $\log_{10}$
 807 transformed FDR, gene ratio represents the ratio of matched genes in categories relative to
 808 total number of differentially expressed genes in each cell type. (M) Metabolic overviews of
 809 differentially expressed genes between bundle sheath and mesophyll. Colour scale presents
 810 the $\log_2$ fold change. (N) k-mean clustering of 4155 differentially expressed transcripts was
 811 performed using $\log_2$ transformed quantile-normalized TPM (transcripts per million) and
 812 visualized in a heatmap. Clusters were named as C_M , C_{BS} , C_V , $C_{BS\&M}$, $C_{BS\&V}$, and $C_{M\&V}$. C_M
 813 contained 613 genes that were strongest in mesophyll cells, C_{BS} contained 285 genes with
 814 preferential expression in bundle sheath cells, C_V contained 972 genes that were strongest
 815 in veins, $C_{BS\&M}$ 1136 genes mostly highly expressed in mesophyll and bundle sheath cells,
 816 $C_{BS\&V}$ 1015 genes that were preferential to bundle sheath and veinal cells, and $C_{M\&V}$ 134
 817 genes strongly expressed in both mesophyll and veinal cells. Colour scale represents Z-
 818 score. (O) Schematic illustrating the enriched primary categories derived from Mapman
 819 using Fisher's exact test ($FDR < 0.1$) for each of the six clusters, colour scale indicates

negative log₁₀ transformed FDR, size of dots (GeneRatio) represents the ratio of matched genes in each category relative to total number of genes in each cluster.

822

Figure 2. Preferential accumulation of transcripts associated with water transport, sulphur and nitrogen assimilation in the rice bundle sheath. (A) Relative transcript abundance for aquaporins in mesophyll, bundle sheath and veinal cells. Log₂ transformed quantile normalized TPM were scaled and genes with similar expression pattern were clustered using hierarchical method. (B) Representative image after *in situ* hybridization localization for *PIP1.1* and *PIP1.3* mRNAs, scale bars represent 20 microns, red arrows indicate specific signal on bundle sheath cell periphery. (C) Schematic illustrating sulphur assimilation and relative transcript abundance in mesophyll (M), bundle sheath (BS) and veinal (V) cells depicted as Z-score from log₂ transformed quantile normalized TPM. (D) Representative image after *in situ* hybridization for *ATPSb* and *SIR* mRNAs, scale bars represent 20 microns, red arrows indicate specific signal on bundle sheath cell periphery. (E) Transcript abundance of *CLT1* and *HAC1;1*. (F) Relative transcript abundance for genes involved in nitrogen assimilation.

836

Figure 3. Relative transcript abundance for genes involved in Calvin Benson Bassham cycle (B) and photorespiration (C) and C₄ pathway (D) in mesophyll (M), bundle sheath (BS) and veinal (V) cells of rice. (A) Paradermal section of rice leaf stained with toluidine blue shows bundle sheath cells are less occupied by chloroplasts compared with mesophyll cells, scale bar represents 50 microns. (B&C) Transcripts associated with Calvin Benson Bassham cycle (B) and photorespiration (C) were preferentially expressed in mesophyll cells, log₂ transformed quantile normalized TPM were scaled and genes with similar expression pattern were clustered using hierarchical method. (D) Transcripts encoding the C₄ acid decarboxylases NADP-DEPENDENT MALIC ENZYME (NADP-ME) and PHOSPHOENOLPYRUVATE CARBOXYKINASE (PCK) accumulated preferentially in bundle sheath and veinal cells whilst ancillary enzymes AMP KINASE (AMK) and PYROPHOSPHORYLASE (PPASE) for PYRUVATE, ORTHOPHOSPHATE DIKINASE (PPDK) accumulated preferentially in bundle sheath cells, data are presented as TPM

(transcript per million), asterisks indicate statistically significant difference (FDR and adjust $P < 0.05$ using edgeR and DESeq2 analysis).

852

Figure 4 Patterning of transcription factors between mesophyll, bundle sheath and veinal cells of rice. (A) Transcription factors from cluster C_M , C_{BS} , C_V , $C_{BS\&M}$, $C_{BS\&V}$, and $C_{M\&V}$ were designated as TF_M , TF_{BS} , TF_V , $TF_{BS\&M}$, $TF_{BS\&V}$, and $TF_{M\&V}$ respectively, relative abundance of differentially expressed transcription factors were presented as heatmap and line plot of Z-score which is calculated from $\log_2$ transformed quantile normalized TPM, red lines in line plot represent mean of Z-score. (B) Significantly enriched or depleted motifs were identified in each of the cistromes from the 6 gene expression clusters, enrichment was calculated using the regioneR permutation testing package (Gel et al., 2016) following motif scanning using FIMO to identify motifs from the plant Jaspar non-redundant database (Fornes et al., 2020). The Z-scores are shown with a colour scale to show the magnitude of enrichment (dark blue) or depletion (yellow) for motifs that were significant after multiple testing correction. Motifs derived from closely related TFs were grouped together for visualisation based on their degree of overlap to predicted target sites (e.g. AP2ERFs). The cistrome from cluster C_V shows the greatest number of enriched motifs, including 13 uniquely enriched, while the C_M and C_{BS} cistromes have far fewer. (C) Cluster specific TFs (left of panel) were mapped to motifs (right of panel) they would be most likely to bind based on high protein sequence similarity with the proteins in the Jaspar plant motif database. The TFs that mapped to any enriched motifs are shown with the motif enrichment data. This allows visualisation of the intersection between TF transcript abundance with potential activation activity. Gene symbols of rice transcription factors were retrieved from funRiceGene database (Yao et al., 2018) but for the symbols not found in the database, symbols of best hit Arabidopsis transcription factors were used and presented in blue. The matching motifs show first the best match and then the motif group if part of a group as shown in 4B.

877

Figure 5. Conserved patterning of gene expression in the Arabidopsis and rice bundle sheath. Orthologs from Arabidopsis and rice associated with aquaporins, sulphate transport and assimilation as well as jasmonic acid biosynthesis are strongly expressed in bundle

sheath cells. (A) The enriched Mapman categories (secondary level) of BS and M preferential genes in rice and Arabidopsis were defined using Fisher's exact test (FDR<0.1). (B) Venn diagram illustrating the extent to which genes in the same orthogroup are preferentially expressed in mesophyll or bundle sheath cells of both rice and Arabidopsis. (C-G) Transcript abundance of Arabidopsis and rice genes belonging to the same orthogroups of aquaporins (C), sulphate transport (D), sulphate assimilation (E), jasmonic acid biosynthesis (F), and transcription factors (G). Data are presented as TPM and statistically significant differences annotated with an asterisk (FDR and adjust $P < 0.05$ using edgeR and DESeq2 analysis in this study, PPDE>0.95 in Aubry et al., 2014b), red and green bars represent bundle sheath and mesophyll respectively.

Supplemental Data

Supplemental Figure S1. Transcript abundance of genes previously reported to be associated with mesophyll, bundle sheath or veins.

Supplemental Figure S2. Mapman categories and metabolic overview of pairwise comparisons between bundle sheath (BS) and vein (V) and between mesophyll (M) and vein (V).

Supplemental Figure S3. Enriched transporter families in the six gene expression clusters.

Supplemental Figure S4. Transcript abundance of genes associated with nitrogen assimilation.

Supplemental Figure S5. Transcript abundance of genes associated with the photosynthetic electron transport chain (A, B), Calvin Benson Bassham cycle (C), starch biosynthesis (D) and photorespiration (E).

Supplemental Figure S6. Transcription factors associated with different clusters.

Supplemental Figure S7. Transcript abundance of orthogroups shared by bundle sheath cells in rice and Arabidopsis associated with aquaporins, sulphate transport and assimilation, jasmonic acid biosynthesis, and transcription factors.

Supplemental Figure S8. Comparison of bundle sheath preferentially expressed genes among different studies.

Supplemental Figure S9. Transcript abundance of Aquaporins (*PIP1* and *PIP2*), *ATP sulfurylase* (*ATPS*), *APS reductase* (*APR*), *sulfite reductase* (*SIR*), *13-lipoxygenase* (*LOX*),

912 *allene oxidase cyclase (AOC)* and *oxophytodienoate reductase (OPR)*, *nitrate reductase*
 913 *(NIA)* and *nitrite reductase (NIR)* in bundle sheath and mesophyll cells of *A. thaliana* (Aubry
 914 et al., 2014b), *G. gynandra* (Aubry et al., 2014a), *O. sativa* (this study), *P. virgatum* (Rao et
 915 al., 2016), *Z. mays* (Chang et al., 2012), *S. italica* (John et al., 2014) and *S. bicolor* (Emms
 916 et al., 2016).

917 **Supplemental Table S1.** RNA sequencing statistics.

918 **Supplemental Table S2.** Summary of differential gene expression analysis using DESeq2
 919 and edgeR.

920 **Supplemental Table S3.** Pairwise comparison and gene expression clusters.

921 **Supplemental Table S4.** Motif enrichment analysis of the six gene expression clusters.

922 **Supplemental Table S5.** Bundle sheath and mesophyll differentially expressed orthologous
 923 genes in rice and Arabidopsis.

References

- Attia, Z., Dalal, A., and Moshelion, M.** (2020). Vascular bundle sheath and mesophyll cells modulate leaf water balance in response to chitin. *Plant J.* **101**: 1368–1377.
- Aubry, S., Kelly, S., Kümpers, B.M.C.C., Smith-Unna, R.D., and Hibberd, J.M.** (2014a). Deep Evolutionary Comparison of Gene Expression Identifies Parallel Recruitment of Trans-Factors in Two Independent Origins of C₄ Photosynthesis. *PLoS Genet.* **10**: e1004365.
- Aubry, S., Smith-Unna, R.D., Bournsnel, C.M., Kopriva, S., and Hibberd, J.M.** (2014b). Transcript residency on ribosomes reveals a key role for the *Arabidopsis thaliana* bundle sheath in sulfur and glucosinolate metabolism. *Plant J.* **78**: 659–673.
- Bailey, T.L., Johnson, J., Grant, C.E., and Noble, W.S.** (2015). The MEME Suite. *Nucleic Acids Res.* **43**: 39–49.
- Berardini, T.Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., and Huala, E.** (2015). The *Arabidopsis* information resource: Making and mining the “gold standard” annotated reference plant genome. *Genesis* **53**: 474–485.
- Berry, J.A.** (2012). There Ought to Be an Equation for That. *Annu. Rev. Plant Biol.* **63**: 1–17.
- Bezruczyk, M., Zöllner, N.R., Kruse, C.P.S., Hartwig, T., Lautwein, T., Köhrer, K., Frommer, W.B., and Kim, J.-Y.** (2021). Evidence for Phloem Loading via the Abaxial Bundle Sheath Cells in Maize Leaves. *Plant Cell*: doi.org/10.1093/plcell/koaa055.
- Birnbaum, K., Shasha, D.E., Wang, J.Y., Jung, J.W., Lambert, G.M., Galbraith, D.W., and Benfey, P.N.** (2003). A gene expression map of the *Arabidopsis* root. *Science* **302**: 1956–60.
- Brady, S.M., Orlando, D.A., Lee, J.Y., Wang, J.Y., Koch, J., Dinneny, J.R., Mace, D., Ohler, U., and Benfey, P.N.** (2007). A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science* **318**: 801–806.
- Brown, N.J. et al.** (2010). C₄ acid decarboxylases required for C₄ photosynthesis are active in the mid-vein of the C₃ species *Arabidopsis thaliana*, and are important in sugar and amino acid metabolism. *Plant J.* **61**: 122–133.
- Burgener, M., Suter, M., Jones, S., and Brunold, C.** (1998). Cyst(e)ine Is the Transport Metabolite of Assimilated Sulfur from Bundle-Sheath to Mesophyll Cells in Maize

955 Leaves. *Plant Physiol.* **116**: 1315–1322.

956 **Burnell, J.N.** (1984). Sulfate Assimilation in C₄ Plants. *Plant Physiol.* **75**: 873–875.

957 **von Caemmerer, S., Edwards, G.E., Koteyeva, N., and Cousins, A.B.** (2014). Single cell
958 C₄ photosynthesis in aquatic and terrestrial plants: A gas exchange perspective. *Aquat.*
959 *Bot.* **118**: 71–80.

960 **von Caemmerer, S., Quick, W.P., and Furbank, R.T.** (2012). The development of C₄ rice:
961 Current progress and future challenges. *Science* **336**: 1671–1672.

962 **Chang, Y.M., Liu, W.Y., Shih, A.C.C., Shen, M.N., Lu, C.H., Lu, M.Y.J., Yang, H.W., Wang,**
963 **T.Y., Chen, S.C.C., Chen, S.M., Li, W.H., and Ku, M.S.B.** (2012). Characterizing
964 regulatory and functional differentiation between maize mesophyll and bundle sheath
965 cells by transcriptomic analysis. *Plant Physiol.* **160**: 165–77.

966 **Chávez-Bárcenas, A.T., Valdez-Alarcón, J.J., Martínez-Trujillo, M., Chen, L.,**
967 **Xoconostle-Cázares, B., Lucas, W.J., and Herrera-Estrella, L.** (2000). Tissue-
968 specific and developmental pattern of expression of the rice *sps1* gene. *Plant Physiol.*
969 **124**: 641–54.

970 **Chaw, S.M., Chang, C.C., Chen, H.L., and Li, W.H.** (2004). Dating the monocot-dicot
971 divergence and the origin of core eudicots using whole chloroplast genomes. *J. Mol.*
972 *Evol.* **58**: 424–441.

973 **Chen, T., Wu, H., Wu, J., Fan, X., Li, X., and Lin, Y.** (2017). Absence of *OsβCA1* causes a
974 CO₂ deficit and affects leaf photosynthesis and the stomatal response to CO₂ in rice.
975 *Plant J.* **90**: 344–357.

976 **Cho, J. II et al.** (2010). Expression analysis and functional characterization of the
977 monosaccharide transporters, *OsTMTs*, involving vacuolar sugar transport in rice. *New*
978 *Phytol.* **186**: 657–668.

979 **Corley, S.M., Troy, N.M., Bosco, A., and Wilkins, M.R.** (2019). QuantSeq 3' Sequencing
980 combined with Salmon provides a fast, reliable approach for high throughput RNA
981 expression analysis. *Sci. Rep.* **9**: 1–15.

982 **Doyama, N., Kajiwara, H., and Ida, S.** (1998). Cloning and expression of a ferredoxin gene
983 in rice roots in response to nitrate and ammonium. *Plant Sci.* **137**: 53–62.

984 **Edwards, G.E. and Black, C.C.** (1971). Isolation of Mesophyll Cells and Bundle Sheath
985 Cells from *Digitaria sanguinalis* (L.) Scop. Leaves and a Scanning Microscopy Study of

the Internal Leaf Cell Morphology. *Plant Physiol.* **47**: 149–56.

Emms, D.M., Covshoff, S., Hibberd, J.M., and Kelly, S. (2016). Independent and Parallel Evolution of New Genes by Gene Duplication in Two Origins of C₄ Photosynthesis Provides New Insight into the Mechanism of Phloem Loading in C₄ Species. *Mol. Biol. Evol.* **33**: 1796–1806.

Emms, D.M. and Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**: 238.

Fetter, K., Van Wilder, V., Moshelion, M., and Chaumont, F. (2004). Interactions between Plasma Membrane Aquaporins Modulate Their Water Channel Activity. *Plant Cell* **16**: 215–228.

Fornes, O. et al. (2020). JASPAR 2020: Update of the open-Access database of transcription factor binding profiles. *Nucleic Acids Res.* **48**: D87–D92.

Foyer, C.H. and Noctor, G. (2011). Ascorbate and glutathione: The heart of the redox hub. *Plant Physiol.* **155**: 2–18.

Fryer, M.J., Oxborough, K., Mullineaux, P.M., and Baker, N.R. (2002). Imaging of photo-oxidative stress responses in leaves. *J. Exp. Bot.* **53**: 1249–1254.

Furbank, R.T. (2016). Walking the C₄ pathway: past, present, and future. *J. Exp. Bot.* **67**: 4057–4066.

Gel, B., Díez-Villanueva, A., Serra, E., Buschbeck, M., Peinado, M.A., and Malinverni, R. (2016). RegioneR: An R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics* **32**: 289–291.

Gerwick, B.C., Ku, S.B., and Black, C.C. (1980). Initiation of sulfate activation: A variation in C₄ photosynthesis plants. *Science* **209**: 513–515.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., and Rokhsar, D.S. (2012). Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res.* **40**: D1178.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: Scanning for occurrences of a given motif. *Bioinformatics* **27**: 1017–1018.

Griffiths, H., Weller, G., Toy, L.F.M., and Dennis, R.J. (2013). You're so vein: bundle sheath physiology, phylogeny and evolution in C₃ and C₄ plants. *Plant. Cell Environ.* **36**: 249–261.

1017 **Grunwald, Y., Wigoda, N., Sade, N., Yaaran, A., Torne, T., Chaka Gosa, S., Moran, N.,**
1018 **and Moshelion, M.** (2021). Arabidopsis leaf hydraulic conductance is regulated by
1019 xylem-sap pH, controlled, in turn, by a P-type H⁺-ATPase of vascular bundle sheath
1020 cells. *Plant J.*: tpj.15235.

1021 **Gu, Z., Eils, R., and Schlesner, M.** (2016). Complex heatmaps reveal patterns and
1022 correlations in multidimensional genomic data. *Bioinformatics* **32**: 2847–2849.

1023 **Haberlandt, G.** (1884). *Physiologische Pflanzenanatomie*.

1024 **Hartigan, J.A. and Wong, M.A.** (1979). A K-Means Clustering Algorithm. *Appl. Stat.* **28**:
1025 100.

1026 **Hatch, M.D.** (1987). C₄ photosynthesis: a unique elend of modified biochemistry, anatomy
1027 and ultrastructure. *BBA Rev. Bioenerg.* **895**: 81–106.

1028 **Hatch, M.D. and Burnell, J.N.** (1990). Carbonic anhydrase activity in leaves and its role in
1029 the first step of C₄ photosynthesis. *Plant Physiol.* **93**: 825–828.

1030 **He, L. et al.** (2020). Primary leaf-type ferredoxin 1 participates in photosynthetic electron
1031 transport and carbon assimilation in rice. *Plant J.* **104**: 44–58.

1032 **Hernández, L.E., Sobrino-Plata, J., Belén Montero-Palmero, M., Carrasco-Gil, S.,**
1033 **Flores-Cáceres, L., Ortega-Villasante, C., and Escobar, C.** (2015). Contribution of
1034 glutathione to the control of cellular redox homeostasis under toxic metal and metalloid
1035 stress. *J. Exp. Bot.* **66**: 2901–2911.

1036 **Hibberd, J.M. and Quick, W.P.** (2002). Characteristics of C₄ photosynthesis in stems and
1037 petioles of C₃ flowering plants. *Nature* **415**: 451–454.

1038 **Hibberd, J.M., Sheehy, J.E., and Langdale, J.A.** (2008). Using C₄ photosynthesis to
1039 increase the yield of rice—rationale and feasibility. *Curr. Opin. Plant Biol.* **11**: 228–231.

1040 **Hua, L. and Hibberd, J.M.** (2019). An optimized protocol for isolation of high-quality RNA
1041 through laser capture microdissection of leaf material. *Plant Direct* **3**: e00156.

1042 **Ibraheem, O., J Botha, C.E., Bradley, G., Dealtry, G., Roux, S., E J Botha, C.C., and**
1043 **Thiel, E.G.** (2013). Rice sucrose transporter1 (OsSUT1) up-regulation in xylem
1044 parenchyma is caused by aphid feeding on rice leaf blade vascular bundles. *Plant Biol.*
1045 **16**: 783–791.

1046 **Jackson** (1992). In situ hybridization in plants. In *Molecular plant pathology : a practical*
1047 *approach* (Oxford University Press), pp. 163–174.

1048 **Janacek, S.H., Trenkamp, S., Palmer, B., Brown, N.J., Parsley, K., Stanley, S., Astley,**
1049 **H.M., Rolfe, S.A., Paul Quick, W., Fernie, A.R., and Hibberd, J.M. (2009).**
1050 Photosynthesis in cells around veins of the C₃ plant *Arabidopsis thaliana* is important
1051 for both the shikimate pathway and leaf senescence as well as contributing to plant
1052 fitness. *Plant J.* **59**: 329–343.

1053 **Jiao, Y. et al. (2009).** A transcriptome atlas of rice cell types uncovers cellular, functional
1054 and developmental hierarchies. *Nat. Genet.* **41**: 258–263.

1055 **Jin, J., Tian, F., Yang, D.C., Meng, Y.Q., Kong, L., Luo, J., and Gao, G. (2017).** PlantTFDB
1056 4.0: Toward a central hub for transcription factors and regulatory interactions in plants.
1057 *Nucleic Acids Res.* **45**: D1040–D1045.

1058 **John, C.R., Smith-Unna, R.D., Woodfield, H., Covshoff, S., and Hibberd, J.M. (2014).**
1059 Evolutionary Convergence of Cell-Specific Gene Expression in Independent Lineages
1060 of C₄ Grasses. *Plant Physiol.* **165**: 62–75.

1061 **Jurić, I., González-Pérez, V., Hibberd, J.M., Edwards, G., and Burroughs, N.J. (2016).**
1062 Size matters for single-cell C₄ photosynthesis in *Bienertia*. *J. Exp. Bot.* **68**: 255–267.

1063 **Kanai, R. and Edwards, G.E. (1973).** Separation of mesophyll protoplasts and bundle
1064 sheath cells from maize leaves for photosynthetic studies. *Plant Physiol.* **51**: 1133–7.

1065 **Kim, J.-Y. et al. (2021).** Distinct identities of leaf phloem cells revealed by single cell
1066 transcriptomics. *Plant Cell*: doi.org/10.1093/plcell/koaa060.

1067 **Kinsman, E.A. and Pyke, K.A. (1998).** Bundle sheath cells and cell-specific plastid
1068 development in *Arabidopsis* leaves. *Development* **125**: 1815–1822.

1069 **Kopriva, S. and Koprivova, A. (2005).** Sulfate assimilation and glutathione synthesis in C₄
1070 plants. *Photosynth. Res.* **86**: 363–372.

1071 **Koprivova, A., Melzer, M., Von Ballmoos, P., Mandel, T., Brunold, C., and Kopriva, S.**
1072 **(2001).** Assimilatory sulfate reduction in C₃, C₃-C₄, and C₄ species of *Flaveria*. *Plant*
1073 *Physiol.* **127**: 543–550.

1074 **Koroleva, O.A., Gibson, T.M., Cramer, R., and Stain, C. (2010).** Glucosinolate-
1075 accumulating S-cells in *Arabidopsis* leaves and flower stalks undergo programmed cell
1076 death at early stages of differentiation. *Plant J.* **64**: 456–469.

1077 **Langdale, J.A. (2011).** C₄ Cycles: Past, Present, and Future Research on C₄
1078 Photosynthesis. *Plant Cell* **23**: 3879–3892.

- 1079 **Leegood, R.C.** (2008). Roles of the bundle sheath cells in leaves of C₃ plants. J. Exp. Bot.
1080 **59**: 1663–1673.
- 1081 **Li, G.W., Zhang, M.H., Cai, W.M., Sun, W.N., and Su, W.A.** (2008). Characterization of
1082 OsPIP2;7, a Water Channel Protein in Rice. Plant Cell Physiol. **49**: 1851–1858.
- 1083 **Love, M.I., Huber, W., and Anders, S.** (2014). Moderated estimation of fold change and
1084 dispersion for RNA-seq data with DESeq2. Genome Biol. **15**: 550.
- 1085 **Lunn, J.E. and Furbank, R.T.** (1997). Localisation of sucrose-phosphate synthase and
1086 starch in leaves of C₄ plants. Planta **202**: 106–111.
- 1087 **Mallmann, J., Heckmann, D., Bräutigam, A., Lercher, M.J., Weber, A.P.M., Westhoff, P.,
1088 and Gowik, U.** (2014). The role of photorespiration during the evolution of C₄
1089 photosynthesis in the genus *Flaveria*. Elife **3**: e02478.
- 1090 **Maruyama-Nakashita, A., Nakamura, Y., Tohge, T., Saito, K., and Takahashi, H.** (2006).
1091 *Arabidopsis* SLIM1 is a central transcriptional regulator of plant sulfur response and
1092 metabolism. Plant Cell **18**: 3235–3251.
- 1093 **Masumoto, C., Miyazawa, S.-I.I., Ohkawa, H., Fukuda, T., Taniguchi, Y., Murayama, S.,
1094 Kusano, M., Saito, K., Fukayama, H., and Miyao, M.** (2010). Phosphoenolpyruvate
1095 carboxylase intrinsically located in the chloroplast of rice plays a crucial role in
1096 ammonium assimilation. Proc. Natl. Acad. Sci. U. S. A. **107**: 5226–31.
- 1097 **Maughan, S.C. et al.** (2010). Plant homologs of the *Plasmodium falciparum* chloroquine-
1098 resistance transporter, *PfCRT*, are required for glutathione homeostasis and stress
1099 responses. Proc. Natl. Acad. Sci. U. S. A. **107**: 2331–2336.
- 1100 **McLeay, R.C. and Bailey, T.L.** (2010). Motif Enrichment Analysis: A unified framework and
1101 an evaluation on ChIP data. BMC Bioinformatics **11**: 165.
- 1102 **Miyake, H.** (2016). Starch Accumulation in the Bundle Sheaths of C₃ Plants: A Possible Pre-
1103 Condition for C₄ Photosynthesis. Plant Cell Physiol. **57**: 890–896.
- 1104 **Moll, P., Ante, M., Seitz, A., and Reda, T.** (2014). QuantSeq 3' mRNA sequencing for RNA
1105 quantification. Nat. Methods **11**: i–iii.
- 1106 **Moore, B. d., Ku, M.S.B., and Edwards, G.E.** (1984). Isolation of leaf bundle sheath
1107 protoplasts from C₄ dicot species and intracellular localization of selected enzymes.
1108 Plant Sci. Lett. **35**: 127–138.
- 1109 **Mustroph, A., Zanetti, M.E., Jang, C.J.H.H., Holtan, H.E., Repetti, P.P., Galbraith, D.W.,**

1110 **Girke, T., and Bailey-Serres, J.** (2009). Profiling translomes of discrete cell
1111 populations resolves altered cellular priorities during hypoxia in *Arabidopsis*. *Proc. Natl.*
1112 *Acad. Sci.* **106**: 18843–18848.

1113 **Nomura, M., Higuchi, T., Ishida, Y., Ohta, S., Komari, T., Imaizumi, N., Miyao-Tokutomi,**
1114 **M., Matsuoka, M., and Tajima, S.** (2005). Differential Expression Pattern of C₄ Bundle
1115 Sheath Expression Genes in Rice, a C₃ Plant. *Plant Cell Physiol* **46**: 754–761.

1116 **Passera, C. and Ghisi, R.** (1982). ATP Sulphurylase and O-Acetylserine Sulphydrylase in
1117 Isolated Mesophyll Protoplasts and Bundle Sheath Strands of S-deprived Maize Leaves.
1118 *J. Exp. Bot.* **33**: 432–438.

1119 **Pasternak, M., Lim, B., Wirtz, M., Hell, R., Cobbett, C.S., and Meyer, A.J.** (2007).
1120 Restricting glutathione biosynthesis to the cytosol is sufficient for normal plant
1121 development. *Plant J.* **53**: 999–1012.

1122 **Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C.** (2017). Salmon
1123 provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**:
1124 417–419.

1125 **Rao, X., Lu, N., Li, G., Nakashima, J., Tang, Y., and Dixon, R.A.** (2016). Comparative cell-
1126 specific transcriptomics reveals differentiation of C₄ photosynthesis pathways in
1127 switchgrass and other C₄ lineages. *J. Exp. Bot.* **67**: 1649–1662.

1128 **Robinson, M.D., McCarthy, D.J., and Smyth, G.K.** (2010). edgeR: a Bioconductor package
1129 for differential expression analysis of digital gene expression data. *Bioinforma. Appl.*
1130 *NOTE* **26**: 139–140.

1131 **Sade, N., Shatil-Cohen, A., Attia, Z., Maurel, C., Boursiac, Y., Kelly, G., Granot, D.,**
1132 **Yaaran, A., Lerner, S., and Moshelion, M.** (2014). The Role of Plasma Membrane
1133 Aquaporins in Regulating the Bundle Sheath-Mesophyll Continuum and Leaf Hydraulics.
1134 *Plant Physiol.* **166**: 1609–1620.

1135 **Sage, R.F.** (2001). Environmental and Evolutionary Preconditions for the Origin and
1136 Diversification of the C₄ Photosynthetic Syndrome. *Plant Biol.* **3**: 202–213.

1137 **Sage, R.F.** (2004). The evolution of C₄ photosynthesis. *New Phytol.* **161**: 341–370.

1138 **Sage, R.F., Sage, T.L., and Kocacinar, F.** (2012). Photorespiration and the Evolution of C₄
1139 Photosynthesis. *Annu. Rev. Plant Biol.* **63**: 19–47.

1140 **Sage, T.L. and Sage, R.F.** (2009). The Functional Anatomy of Rice Leaves: Implications for

1141 Refixation of Photorespiratory CO₂ and Efforts to Engineer C₄ Photosynthesis into Rice.
1142 Plant Cell Physiol. **50**: 756–772.

1143 **Sakurai, J., Ishikawa, F., Yamaguchi, T., Uemura, M., and Maeshima, M. (2005).**
1144 Identification of 33 rice aquaporin genes and analysis of their expression and function.
1145 Plant Cell Physiol. **46**: 1568–1577.

1146 **Sawers, R.J.H., Liu, P., Anufrikova, K., Gene, J.T.G., and Brutnell, T.P. (2007).** A multi-
1147 treatment experimental system to examine photosynthetic differentiation in the maize
1148 leaf. BMC Genomics **8**: 12.

1149 **Schmutz, D. and Brunold, C. (1984).** Intercellular Localization of Assimilatory Sulfate
1150 Reduction in Leaves of *Zea mays* and *Triticum aestivum*. Plant Physiol. **74**: 866–870.

1151 **Schwacke, R., Ponce-Soto, G.Y., Krause, K., Bolger, A.M., Arsova, B., Hallab, A.,**
1152 **Gruden, K., Stitt, M., Bolger, M.E., and Usadel, B. (2019).** MapMan4: A Refined
1153 Protein Classification and Annotation Framework Applicable to Multi-Omics Data
1154 Analysis. Mol. Plant **12**: 879–892.

1155 **Scofield, G.N., Hirose, T., Aoki, N., and Furbank, R.T. (2007).** Involvement of the sucrose
1156 transporter, OsSUT1, in the long-distance pathway for assimilate transport in rice. J.
1157 Exp. Bot. **58**: 3155–69.

1158 **Shatil-Cohen, A., Attia, Z., and Moshelion, M. (2011).** Bundle-sheath cell regulation of
1159 xylem-mesophyll water transport via aquaporins under drought stress: a target of xylem-
1160 borne ABA? Plant J. **67**: 72–80.

1161 **Shen, W., Ye, L., Ma, J., Yuan, Z., Zheng, B., Chuangen, L. V., Zhu, Z., Chen, X., Gao,**
1162 **Z., and Chen, G. (2016).** The existence of C₄-bundle-sheath-like photosynthesis in the
1163 mid-vein of C₃ rice. Rice **9**: 20.

1164 **Soneson, C., Love, M.I., and Robinson, M.D. (2016).** Differential analyses for RNA-seq:
1165 Transcript-level estimates improve gene-level inferences. F1000Research **4**: 1521.

1166 **Taylor, L., Nunes-Nesi, A., Parsley, K., Leiss, A., Leach, G., Coates, S., Wingler, A.,**
1167 **Fernie, A.R., and Hibberd, J.M. (2010).** Cytosolic pyruvate, orthophosphate dikinase
1168 functions in nitrogen remobilization during leaf senescence and limits individual seed
1169 growth and nitrogen content. Plant J. **62**: 641–652.

1170 **Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., Selbig, J., Müller,**
1171 **L.A., Rhee, S.Y., and Stitt, M. (2004).** MAPMAN: a user-driven tool to display genomics

1172 data sets onto diagrams of metabolic pathways and other biological processes. *Plant J.*
1173 **37**: 914–939.

1174 **Tipple, B.J. and Pagani, M.** (2007). The Early Origins of Terrestrial C₄ Photosynthesis.
1175 *Annu. Rev. Earth Planet. Sci.* **35**: 435–461.

1176 **Voznesenskaya, E. V., Franceschi, V.R., Kiirats, O., Freitag, H., and Edwards, G.E.**
1177 (2001). Kranz anatomy is not essential for terrestrial C₄ plant photosynthesis. *Nature*
1178 **414**: 543–546.

1179 **Wang, P., Khoshravesh, R., Karki, S., Tapia, R., Balahadia, C.P., Bandyopadhyay, A.,**
1180 **Quick, W.P., Furbank, R., Sage, T.L., and Langdale, J.A.** (2017). Re-creation of a Key
1181 Step in the Evolutionary Switch from C₃ to C₄ Leaf Anatomy. *Curr. Biol.* **27**: 3278-
1182 3287.e6.

1183 **Waters, M.T., Moylan, E.C., and Langdale, J.A.** (2008). GLK transcription factors regulate
1184 chloroplast development in a cell-autonomous manner. *Plant J.* **56**: 432–444.

1185 **Waters, M.T., Wang, P., Korkaric, M., Capper, R.G., Saunders, N.J., and Langdale, J.A.**
1186 (2009). GLK Transcription Factors Coordinate Expression of the Photosynthetic
1187 Apparatus in *Arabidopsis*. *Plant Cell* **21**: 1109–1128.

1188 **Wickham, H.** (2016). ggplot2: Elegant Graphics for Data Analysis.

1189 **Wigoda, N., Pasmanik-Chor, M., Yang, T., Yu, L., Moshelion, M., and Moran, N.** (2017).
1190 Differential gene expression and transport functionality in the bundle sheath versus
1191 mesophyll-a potential role in leaf mineral homeostasis. *J. Exp. Bot.* **68**: 3179–3190.

1192 **Williams, M.L., Farrar, J.F., and Pollock, C.J.** (1989). Cell specialization within the
1193 parenchymatous bundle sheath of barley. *Plant, Cell Environ.* **12**: 909–918.

1194 **Yang, J., Gao, M., Hu, H., Ding, X., Lin, H., Wang, L., Xu, J., Mao, C., Zhao, F., and Wu,**
1195 **Z.** (2016). OsCLT1, a CRT-like transporter 1, is required for glutathione homeostasis
1196 and arsenic tolerance in rice. *New Phytol.* **211**: 658–670.

1197 **Yao, W., Li, G., Yu, Y., and Ouyang, Y.** (2018). funRiceGenes dataset for comprehensive
1198 understanding and application of rice functional genes. *Gigascience* **7**: 1–9.

1199 **Yonekura-Sakakibara, K., Onda, Y., Ashikari, T., Tanaka, Y., Kusumi, T., and Hase, T.**
1200 (2000). Analysis of reductant supply systems for ferredoxin-dependent sulfite reductase
1201 in photosynthetic and nonphotosynthetic organs of maize. *Plant Physiol.* **122**: 887–894.

1202 **Yu, G.** (2020). Using ggtree to Visualize Data on Tree-Like Structures. *Curr. Protoc.*

1203 Bioinforma. **69**.

1204 **Zelazny, E., Borst, J.W., Muylaert, M., Batoko, H., Hemminga, M.A., and Chaumont, F.**

1205 (2007). FRET imaging in living maize cells reveals that plasma membrane aquaporins

1206 interact to regulate their subcellular localization. Proc. Natl. Acad. Sci. U. S. A. **104**:

1207 12359–12364.

1208 **Zhang, W., Wu, Y., Schnable, J.C., Zeng, Z., Freeling, M., Crawford, G.E., and Jiang, J.**

1209 (2012). High-resolution mapping of open chromatin in the rice genome. Genome Res.

1210 **22**: 151–162.

1211

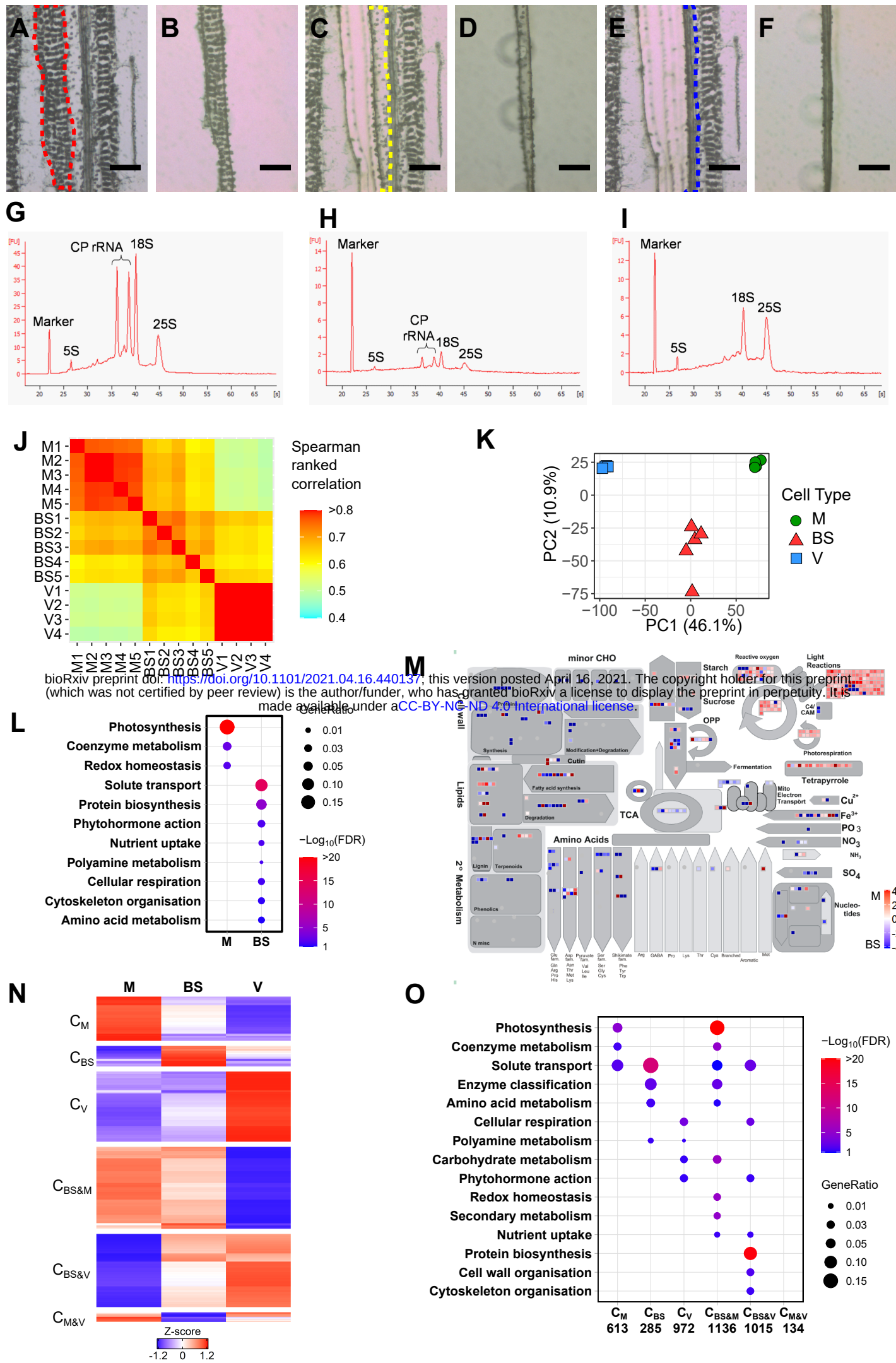


Figure 1. RNA was isolated from mesophyll, bundle sheath and veinal cells of rice using laser capture microdissection. Each cell-type was identified and then sequentially removed from paradermal sections prior to RNA quality being assessed. (A&B) Representative image of mesophyll cells outlined with red dash line (A) that are cut with a UV laser and then captured with an infrared laser and placed on a cap (B). (C&D) Bundle sheath cells (C) were then cut and captured (D). (E&F) Lastly, veinal cells (E) were cut and captured (F). Scale bars represents 50 microns. (G-I) Representative RNA profiles from microdissected mesophyll (G), bundle sheath (H) and veinal cells (I). Peaks from the cytosolic 25S, 18S and 5S ribosomal RNAs were detected in all cell types. Chloroplastic ribosomal RNAs (CP rRNA) were clearly detectable in mesophyll and bundle sheath cells. (J) Spearman ranked correlations of $\log_2$ transformed TPM (transcripts per million) indicates little variation between biological replicates from each cell type, and distinct patterns of gene expression in each cell type. (K) Principal component analysis of normalised counts after variance-stabilizing transformation showing that cell type accounted for 46.1% of the variance detected in the data. (L) Primary Mapman categories associated with differentially expressed genes in bundle sheath and mesophyll cells. Terms were defined using Fisher's exact test (False discovery rate, $FDR < 0.1$), colour scale represents negative $\log_{10}$ transformed FDR, gene ratio represents the ratio of matched genes in categories relative to total number of differentially expressed genes in each cell type. (M) Metabolic overviews of differentially expressed genes between bundle sheath and mesophyll. Colour scale presents the $\log_2$ fold change. (N) k-mean clustering of 4155 differentially expressed transcripts was performed using $\log_2$ transformed quantile-normalized TPM (transcripts per million) and visualized in a heatmap. Clusters were named as C_M , C_{BS} , C_V , $C_{BS\&M}$, $C_{BS\&V}$ and $C_{M\&V}$. C_M contained 613 genes that were strongest in mesophyll cells, C_{BS} contained 285 genes with preferential expression in bundle sheath cells, C_V contained 972 genes that were strongest in veins, $C_{BS\&M}$ 1136 genes mostly highly expressed in mesophyll and bundle sheath cells, $C_{BS\&V}$ 1015 genes that were preferential to bundle sheath and veinal cells, and $C_{M\&V}$ 134 genes strongly expressed in both mesophyll and veinal cells. Colour scale represents Z-score. (O) Schematic illustrating the enriched primary categories derived from Mapman using Fisher's exact test ($FDR < 0.1$) for each of the six clusters, colour scale indicates negative $\log_{10}$ transformed FDR, size of dots (GeneRatio) represents the ratio of matched genes in each category relative to total number of genes in each cluster.

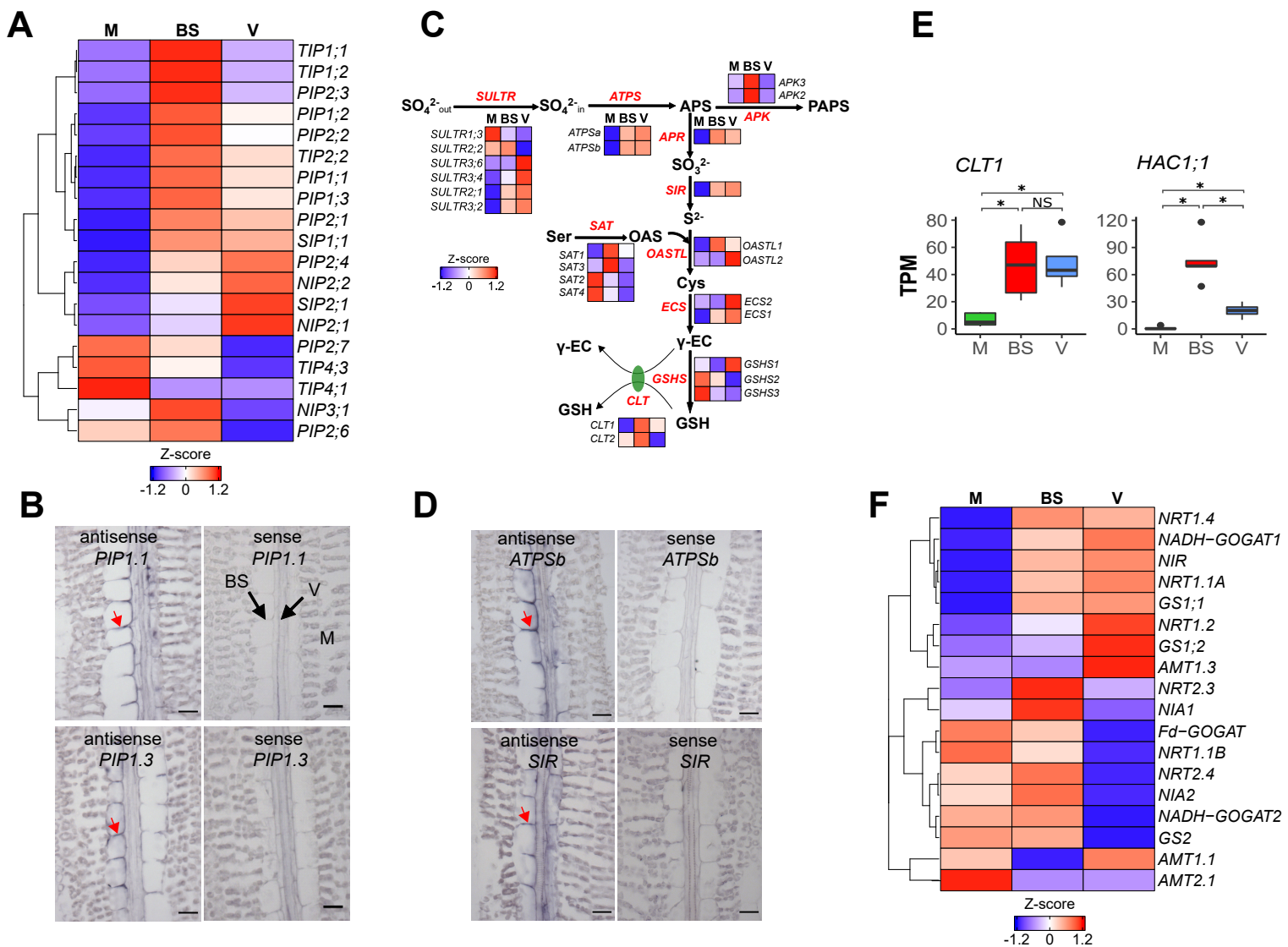


Figure 2. Preferential accumulation of transcripts associated with water transport, sulphur and nitrogen assimilation in the rice bundle sheath. (A) Relative transcript abundance for aquaporins in mesophyll, bundle sheath and veinal cells. Log₂ transformed quantile normalized TPM were scaled and genes with similar expression pattern were clustered using hierarchical method. (B) Representative image after *in situ* hybridization localization for *PIP1.1* and *PIP1.3* mRNAs, scale bars represent 20 microns, red arrows indicate specific signal on bundle sheath cell periphery. (C) Schematic illustrating sulphur assimilation and relative transcript abundance in mesophyll (M), bundle sheath (BS) and veinal (V) cells depicted as Z-score from log₂ transformed quantile normalized TPM. (D) Representative image after *in situ* hybridization for *ATPSb* and *SIR* mRNAs, scale bars represent 20 microns, red arrows indicate specific signal on bundle sheath cell periphery. (E) Transcript abundance of *CLT1* and *HAC1;1*. (F) Relative transcript abundance for genes involved in nitrogen assimilation.

Abbreviations for (A): PIP, plasma membrane intrinsic proteins; TIP, tonoplast intrinsic proteins; SIP, small basic intrinsic proteins; NIP, NOD26-like intrinsic proteins. Abbreviations for (C): SULTR, sulphate transporter; ATPS, ATP sulfurylase; APS, adenosine 5'-phosphosulfate; PAPS, 3'-phosphoadenosine 5'-phosphosulfate; APR, APS reductase; SIR, sulfite reductase; APK, APS kinase; SAT, serine acetyltransferase; OAS, O-acetylserine; OASTL, O-acetylserine (thiol)lyase; γ -ECS, glutamate-cysteine ligase; γ -EC, γ -glutamylcysteine; CLT, chloroquine-resistance transporter-like transporter; GS, glutamine synthetase; NADH-GOGAT, NADH-dependent glutamate synthase; Fd-GOGAT, ferredoxin-dependent glutamate synthase.

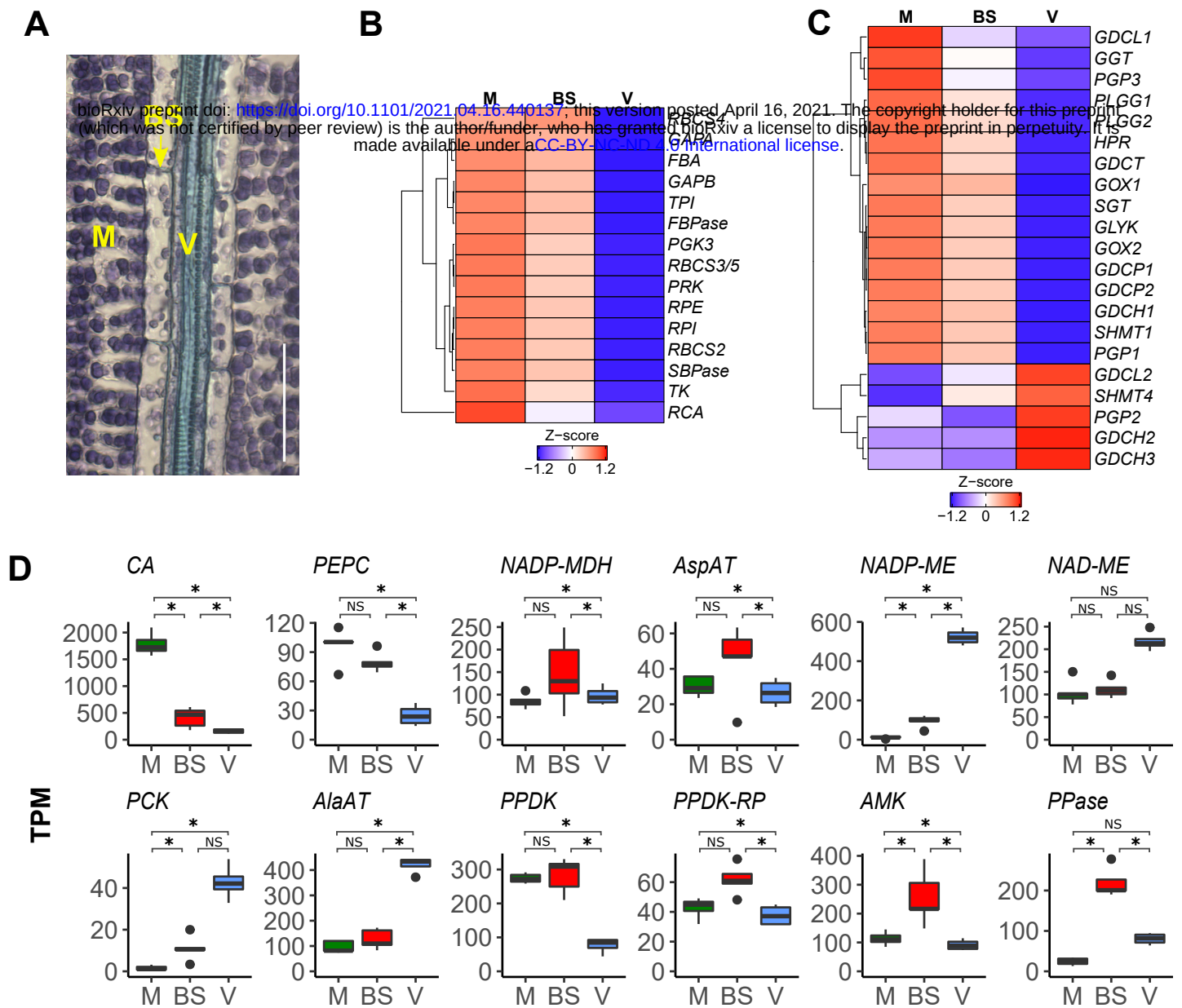


Figure 3. Relative transcript abundance for genes involved in Calvin Benson Bassham cycle (B) and photorespiration (C) and C_4 pathway (D) in mesophyll (M), bundle sheath (BS) and veinal (V) cells of rice. (A) Paradermal section of rice leaf stained with toluidine blue shows bundle sheath cells are less occupied by chloroplasts compared with mesophyll cells, scale bar represents 50 microns. (B&C) Transcripts associated with Calvin Benson Bassham cycle (B) and photorespiration (C) were preferentially expressed in mesophyll cells, $\log_2$ transformed quantile normalized TPM were scaled and genes with similar expression pattern were clustered using hierarchical method. (D) Transcripts encoding the C_4 acid decarboxylases NADP-DEPENDENT MALIC ENZYME (NADP-ME) and PHOSPHOENOLPYRUVATE CARBOXYKINASE (PCK) accumulated preferentially in bundle sheath and veinal cells whilst ancillary enzymes AMP KINASE (AMK) and PYROPHOSPHORYLASE (PPASE) for PYRUVATE, ORTHOPHOSPHATE DIKINASE (PPDK) accumulated preferentially in bundle sheath cells, data are presented as TPM (transcript per million), asterisks indicate statistically significant difference (FDR and adjust $P < 0.05$ using edgeR and DESeq2 analysis).

Abbreviations in (B): PGK, phosphoglycerate kinase; GAP/B, glyceraldehyde-3-phosphate dehydrogenase; TPI, triose-phosphate isomerase; FBA, fructose-1,6-bisphosphate aldolase, FBPAse, fructose-1,6-bisphosphatase; TK, transketolase; SBPAse, sedoheptulose 1,7-bisphosphatase; RPI, phosphopentose isomerase; RPE, phosphopentose epimerase; PRK, phosphoribulokinase; RBCS, RuBisCO small subunit; RCA, RuBisCO activase. Abbreviations in (C): PGP, phosphoglycolate phosphatase; GOX, glycolate oxidase; GGT, glutamate:glyoxylate aminotransferase; GDC, glycine decarboxylase; SHMT, serine hydroxymethyltransferase; SGT, serine:glyoxylate aminotransferase; HPR, hydroxypyruvate reductase; GLYK, glycerate kinase. Abbreviations in (D): CA, carbonic anhydrase; PEPC, phosphoenolpyruvate carboxylase; NADP-MDH, NADP-dependent malic dehydrogenase; AspAT, aspartate aminotransferase; NADP-ME, NADP-dependent malic enzyme; NAD-ME, NAD-dependent malic enzyme; PCK, phosphoenolpyruvate carboxykinase; AlaAT, alanine aminotransferase; PPDK, pyruvate, orthophosphate dikinase; PPDK-RP, regulatory protein of PPDK; AMK, AMP kinase; PPase, pyrophosphorylase.

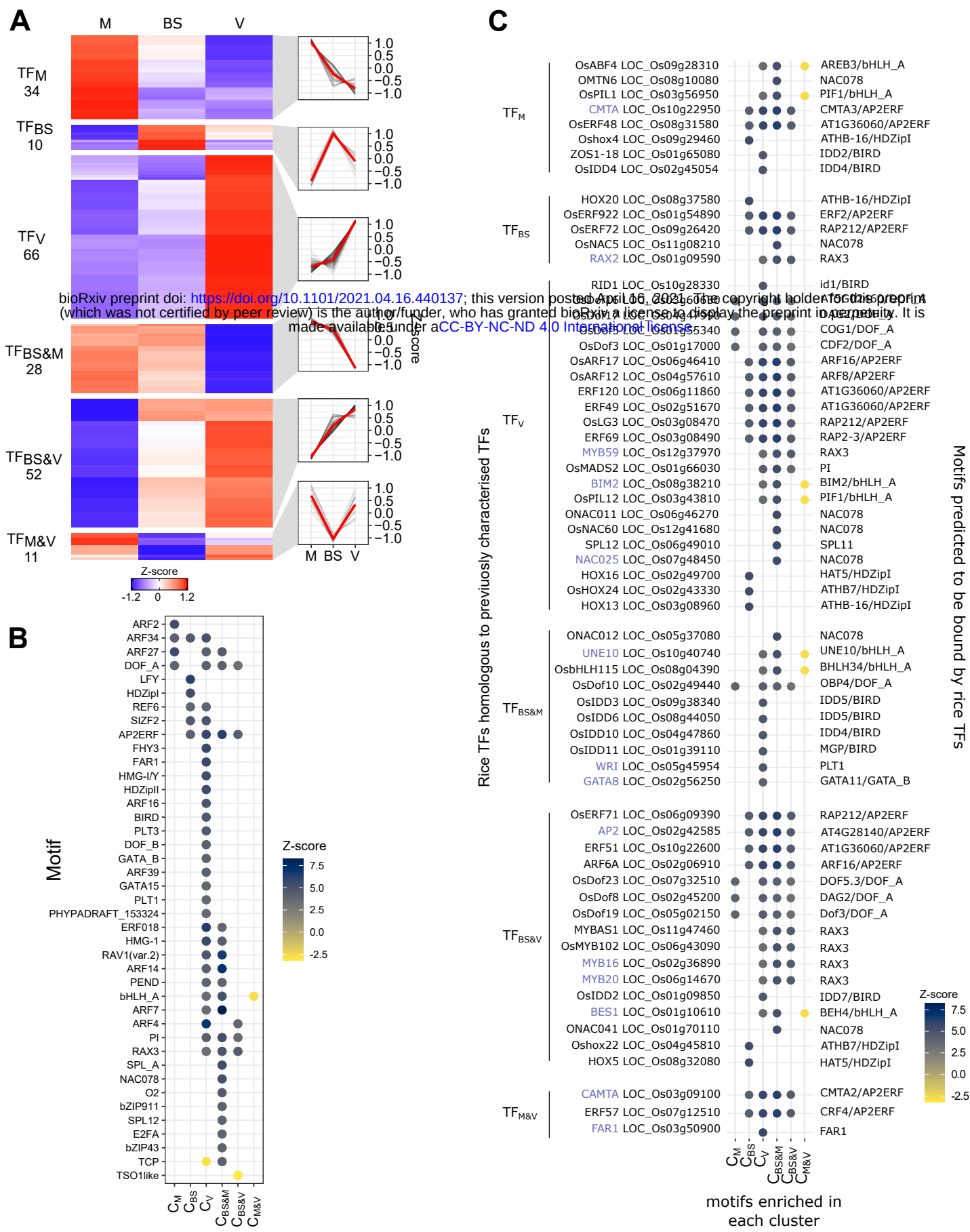


Figure 4 Patterning of transcription factors between mesophyll, bundle sheath and veinal cells of rice. (A) Transcription factors from cluster C_M , C_{BS} , C_V , $C_{BS\&M}$, $C_{BS\&V}$, and $C_{M\&V}$ were designated as TF_M , TF_{BS} , TF_V , $TF_{BS\&M}$, $TF_{BS\&V}$ and $TF_{M\&V}$ respectively, relative abundance of differentially expressed transcription factors were presented as heatmap and line plot of Z-score which is calculated from $\log_2$ transformed quantile normalized TPM, red lines in line plot represent mean of Z-score. (B) Significantly enriched or depleted motifs were identified in each of the cistromes from the 6 gene expression clusters, enrichment was calculated using the regioneR permutation testing package (Gel et al., 2016) following motif scanning using FIMO to identify motifs from the plant Jaspar non-redundant database (Fornes et al., 2020). The Z-scores are shown with a colour scale to show the magnitude of enrichment (dark blue) or depletion (yellow) for motifs that were significant after multiple testing correction. Motifs derived from closely related TFs were grouped together for visualisation based on their degree of overlap to predicted target sites (e.g. AP2ERFs). The cistrome from cluster C_V shows the greatest number of enriched motifs, including 13 uniquely enriched, while the C_M and C_{BS} cistromes have far fewer. (C) Cluster specific TFs (left of panel) were mapped to motifs (right of panel) they would be most likely to bind based on high protein sequence similarity with the proteins in the Jaspar plant motif database. The TFs that mapped to any enriched motifs are shown with the motif enrichment data. This allows visualisation of the intersection between TF transcript abundance with potential activation activity. Gene symbols of rice transcription factors were retrieved from funRiceGene database (Yao et al., 2018) but for the symbols not found in the database, symbols of best hit Arabidopsis transcription factors were used and presented in blue. The matching motifs show first the best match and then the motif group if part of a group as shown in 4B.

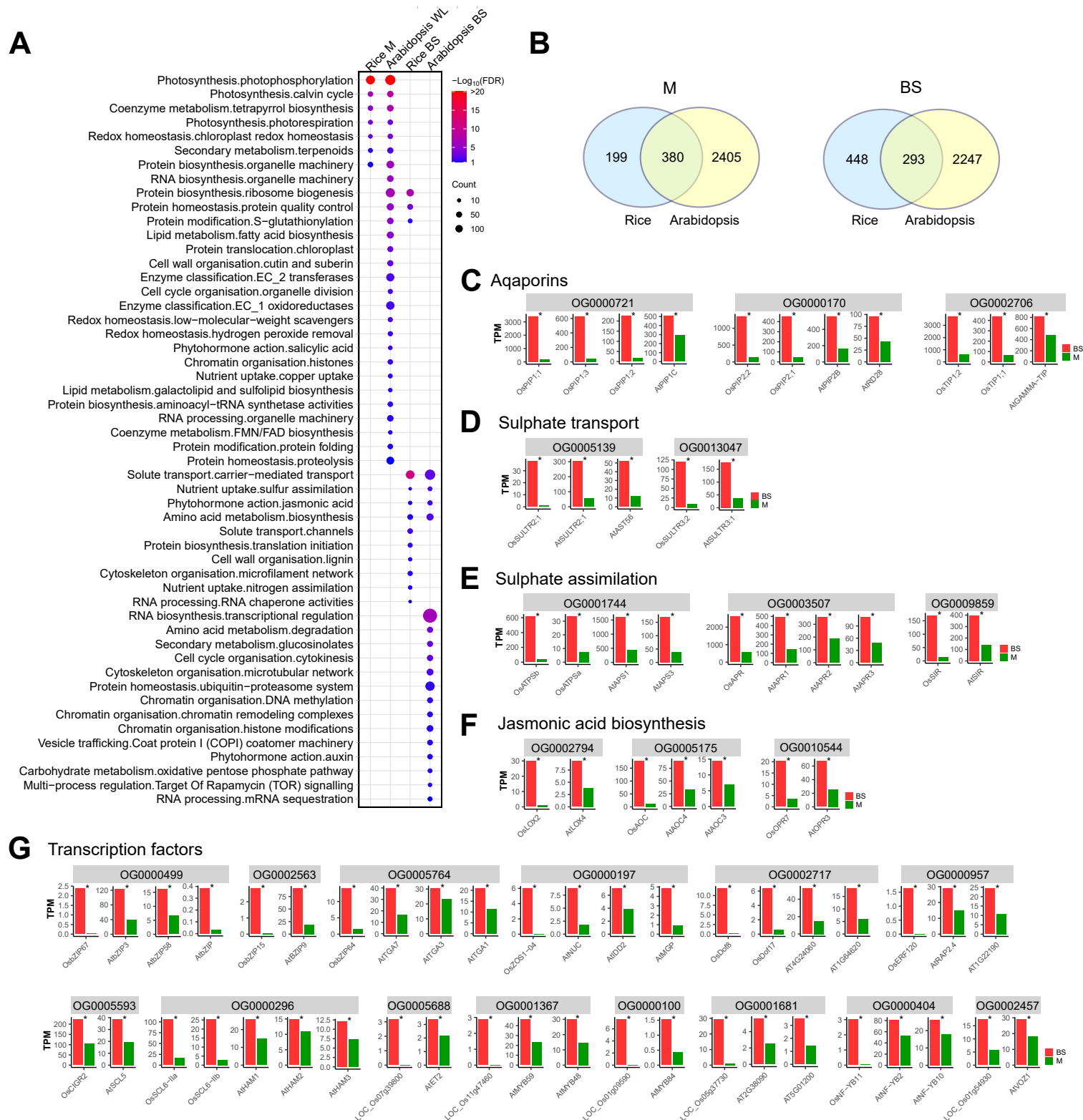


Figure 5. Conserved patterning of gene expression in the Arabidopsis and rice bundle sheath. Orthologs from Arabidopsis and rice associated with aquaporins, sulphate transport and assimilation as well as jasmonic acid biosynthesis are strongly expressed in bundle sheath cells. (A) The enriched Mapman categories (secondary level) of BS and M preferential genes in rice and Arabidopsis were defined using Fisher's exact test ($\text{FDR} < 0.1$). (B) Venn diagram illustrating the extent to which genes in the same orthogroup are preferentially expressed in mesophyll or bundle sheath cells of both rice and Arabidopsis. (C-G) Transcript abundance of Arabidopsis and rice genes belonging to the same orthogroups of aquaporins (C), sulphate transport (D), sulphate assimilation (E), jasmonic acid biosynthesis (F), and transcription factors (G). Data are presented as TPM and statistically significant differences annotated with an asterisk (FDR and adjusted $P < 0.05$ using edgeR and DESeq2 analysis in this study, PPDE > 0.95 in Aubry et al., 2014b), red and green bars represent bundle sheath and mesophyll respectively.