# 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 Hua[1], Sean R. Stevenson[1], Ivan Reyna-Llorens[1], Haiyan Xiong[1], Stanislav Kopriva[2] and Julian M. Hibberd[1]

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

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

Email addresses:

LH - lh556@cam.ac.uk

SRS - srs62@cam.ac.uk

IRL - suallorens@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 $C_4$ photosynthesis into species such as $C_3$ rice to increase yields (von Caemmerer et al., 2012; Hibberd et al., 2008; Langdale, 2011). As $C_4$ 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 $C_4$ pathway is thought to have evolved over 60 times independently in monocotyledons and dicotyledons in response to reduced water availability and $CO_2$ 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 $C_4$ 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 $C_4$ species, RuBisCO activity is replaced with phospho*enol*pyruvate carboxylase in mesophyll cells to allow bicarbonate to be fixed into oxaloacetate. High concentrations of $C_4$ acids derived from oxaloacetate build up in mesophyll cells and drive diffusion to the bundle sheath where $C_4$ acid decarboxylases resupply $CO_2$ 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 $C_4$ 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 (Bezrutczyk et al., 2021). The importance of bundle sheath cells in the $C_4$ 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 $C_4$ leaf bundle sheath cells are well characterized and carry out a variety of specialized roles.

Although the vast majority of plants use the ancestral $C_3$ pathway, our understanding of gene expression in bundle sheath cells specifically, and other major cell types more generally of $C_3$ 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 $C_3$ 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 $C_4$ plants we have a relatively poor understanding of the role of bundle sheath cells in $C_3$ 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).

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

Here we used an optimised laser capture microdissection protocol for RNA isolation from leaves (Hua and Hibberd 2019) to study the patterns of gene expression in bundle sheath, veinal (including phloem and xylem parenchyna, xylem, as well as sieve elements and companion cells), and mesophyll cells of rice. We had three main hypotheses. First, that as the rice bundle sheath contains a small chloroplast compartment, gene expression would be poorly set up for photosynthesis. Second, as in other species, gene expression in the rice bundle sheath would favour water transport. Third, in contrast to Arabidopsis, as rice does not make glucosinolates there has been no selection pressure to restrict sulphur assimilation to the bundle sheath. Mesophyll and bundle sheath cells were distinguished by over-representation of terms including photosynthesis and co-enzyme metabolism in the former, and solute transport and protein synthesis in the latter. Transcripts encoding the majority of aquaporins were more abundant in bundle sheath cells, and this was also the case for transcripts encoding proteins associated with sulphur assimilation and nitrate reduction. Transcription factors that were preferential to each cell type were identified, but in most cases a gradient from veins to bundle sheath to mesophyll cells, or *vice versa*, was detected. Direct comparison to publicly available data from Arabidopsis identified groups of genes encoding aquaporins, proteins allowing sulphur assimilation, jasmonic acid synthesis and also a small number of transcription factor families that showed preferential expression in the bundle sheath of both species. Whilst these findings could be due to evolutionary convergence, the most parsimonious explanation is that the bundle sheath of the last common ancestor of monocotyledons and dicotyledons was specialised in water transport, sulphur assimilation as well as jasmonic acid synthesis, and that members of the Basic Leucine Zipper, $C_2H_2$-type Zinc Finger, DNA-binding with One Finger, Ethylene Responsive Factor, Hairy-Related Transcription-Factor, GRAS, MYB, Nuclear Factor-YB and Vascular Plant One-Zinc Finger Protein transcription factor families play ancient and conserved roles in this cell type.

## Results

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

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

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

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

To quantify the extent to which transcript abundance differed between bundle sheath, mesophyll and veinal cells, we performed differential gene expression analysis using DESeq2 and edgeR (Love et al., 2014; Robinson et al., 2010). This identified 1,919 differentially expressed genes between bundle sheath and mesophyll cells, the majority of which (1,173) were more abundant in bundle sheath cells (FDR and adjusted *P* <5%) (Supplemental Table 2). Functional enrichment analysis identified three categories over-represented in mesophyll cells containing transcripts linked to photosynthesis, coenzyme metabolism and redox homeostasis (Figure 1L). Eight categories were associated with bundle sheath cells, including solute transport, protein biosynthesis, phytohormone action, nutrient uptake, polyamine metabolism and cellular respiration (Figure 1L). To provide an overview of metabolic specialization in mesophyll and bundle sheath cells, Mapman (Thimm et al., 2004; Schwacke et al., 2019) was used. Consistent with mesophyll cells being specialized for photosynthesis, this indicated that transcripts encoding components of the light reactions, Calvin Benson Bassham cycle and tetrapyrolle biosynthesis were upregulated in mesophyll cells, whilst those associated with many other metabolic processes including cell wall, minor carbohydrate, fatty acid, amino acid and secondary metabolism were more abundant in the bundle sheath (Figure 1M). Quantitative comparison of bundle sheath and veinal cells showed 1,258 and 660 genes were significantly up and down regulated in bundle sheath cells respectively (Supplemental Table 2) and indicated categories associated with the bundle sheath included photosynthesis, carbohydrate metabolism, redox homeostasis, secondary metabolism and solute transport (Supplemental Figure 2A). Mapman outputs confirmed transcripts encoding components of the light-dependent reactions of photosynthesis, as well as the Calvin-Benson-Bassham cycle and photorespiratory pathway were more abundant in bundle sheath than veinal cells (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 $C_4$ 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 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-*

*DEPENDENT MALIC ENZYME* (*NADP-ME*) and *PHOSPHOENOLPYRUVATE CARBOXYKINASE* (*PCK*) showed a gradient in transcript abundance from veins, to bundle sheath to mesophyll cells. These data indicate that high activity of these $C_4$ decarboxylases in vascular bundles of rice (Shen et al., 2016) is likely caused by their expression in the vein rather than bundle sheath cells. Overall, $C_4$ genes could be partitioned into three main groups, either showing a strong negative gradient (*CA* and *PEPC*) from mesophyll to vein cells, a strong positive gradient (*NADP-ME, NAD-ME, PCK* and *AlaAT*) or a tendency to be most strongly expressed in the bundle sheath (*NADP-MDH, AspAT, PPDK, PPDK-RP, AMP* and *PPase*) (Figure 3D). This finding is consistent with them being part of multiple gene regulatory networks in the ancestral $C_3$ state that pattern their expression across these cell types.

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

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

The bundle sheath specific cluster contained only ten genes that derived from families such as the ERFs, bZIPs and bHLHs (Figure 4A, Supplemental Figure 6A). Transcription factors abundant in both bundle sheath and vein belonged to the MYB, ERF, bZIP and WRKY families, whilst CO-like, $C_2H_2$, and G2-like families were abundant in the mesophyll

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

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

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

We next investigated whether transcription factors and their cognate DNA binding sites were associated with the six clusters that had been defined by patterns of gene expression across mesophyll, bundle sheath and veinal cells. As relatively few transcription factors from rice have had their DNA binding characteristics determined, we first used protein homology (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, tetrapyrrolle 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

462 mesophyll or bundle sheath cells determined. Genes from 380 orthogroups were
463 preferentially expressed in mesophyll cells of rice and whole leaves of Arabidopsis, whilst
464 genes from 293 orthogroups were shared by bundle sheath of both species (Figure 5B), in
465 both cases a greater overlap than would be expected by chance (Fisher's exact test, $p<$
466 $2.2e^{-16}$ for both bundle sheath and mesophyll). However, the odds ratio that is also generated
467 from the Fisher's exact test indicated that there was a greater degree of overlap between
468 the mesophyll of rice and whole leaves of Arabidopsis than between their bundle sheath
469 cells (odds ratios were 5.32 and 2.06 for mesophyll and bundle sheath respectively). We
470 thus conclude that the expression of orthologs has diverged more in the bundle sheath than
471 the mesophyll of these species. Specific orthogroups associated with the mesophyll in both
472 species were mainly associated with photosynthesis but also included eleven orthogroups
473 of transcription factors, one of which included the transcription factor *GLK1* (OG0002393)
474 (Supplemental table 5) that is known to activate photosynthesis related gene expression
475 (Waters et al., 2008; Waters et al., 2009). Although fewer orthogroups were common to the
476 bundle sheath in rice and Arabidopsis, overlap included three groups encoding aquaporins
477 (PIP1 - OG0000721, PIP2 -OG0000170 and TIP1 - OG0002706), as well as orthologs
478 involved in sulphate transport (OG0005139 and OG0013047), sulphate assimilation
479 (OG0001744, OG0003507 and OG0009859) and jasmonic acid biosynthesis (OG0002794,
480 OG0005175, OG0010544) (Figure 5C-F, Supplemental Figure 7, Supplemental Table 5).
481 We compared bundle sheath preferentially expressed genes of rice and Arabidopsis with
482 the bundle sheath marker genes identified through single cell sequencing of Arabidopsis
483 (Kim et al., 2021). As sequencing depth from single celled sequencing is not as great, fewer
484 transcripts were detected (Supplemental Figure 8A). However, 41 orthogroups were
485 identified containing at least one gene preferential to the bundle sheath in all three studies,
486 including two sulphate transporters (OG0005139 and OG0013047), ATP sulfurylase
487 (OG0001744), and allene oxide cyclase (OG0005175). To investigate whether these genes
488 are associated with bundle sheath cells in a broader range of species, we assessed
489 transcript abundance of consensus orthogroups associated with water transport, sulphur
490 assimilation, jasmonic acid biosynthesis as well as nitrate reduction in publicly available data
491 from *Panicum virgatum* (Rao et al., 2016), *Sorghum bicolor* (Emms et al., 2016), *Setaria*
492 *italica* (John et al., 2014) and *Zea mays* (Chang et al., 2012) as well as the $C_4$ 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), $C_2H_2$-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.

## Discussion

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

Separation of protoplasts followed by cell sorting has provided significant insight into gene expression in specific cells of roots (Birnbaum et al., 2003; Brady et al., 2007). In leaves, the ability to separate mesophyll and bundle sheath cells from $C_4$ plants (Kanai and Edwards, 1973; Edwards and Black, 1971; Moore et al., 1984) has allowed similar levels of insight into these cells, but leaf cells from $C_3$ species have been more challenging to isolate. As a consequence, although major cell-types of leaves such as the mesophyll, phloem, xylem and guard cells have well defined roles, others such as the bundle sheath are relatively poorly understood (Leegood, 2008). In Arabidopsis the bundle sheath is thought to play important roles in regulating hydraulic conductance, substrate transport and storage (Shatil-Cohen et al., 2011; Sade et al., 2014; Griffiths et al., 2013). Analysis of mRNAs resident on ribosomes showed that patterns of gene expression in the Arabidopsis bundle sheath are conditioned to facilitate sulfur metabolism and glucosinolate biosynthesis (Aubry et al., 2014b). Although, suppression of chlorophyll synthase in veinal tissue including bundle sheath cells of Arabidopsis reduced photosynthesis, growth and fitness (Janacek et al., 2009), the importance of the bundle sheath itself was not defined. Moreover, in $C_3$ monocotyledons, the group containing many of our most important crops, little is known about the function of bundle sheath cells. To address this, we defined gene expression in the rice bundle sheath as well as mesophyll and vein cells using laser capture microdissection coupled with mRNA sequencing. Contrary to our expectations, this indicated that although bundle sheath cells of rice contain few chloroplasts (Wang et al., 2017; Sage and Sage, 2009), transcripts encoding components of the photosynthetic electron transport chain and the Calvin Benson Bassham cycle were clearly expressed in rice bundle sheath cells. The consequence of this finding is that expression of photosynthesis genes represents a continuum from low to medium to high in veins, bundle sheath and mesophyll cells respectively. Whether this is also the case in Arabidopsis remains to be determined. There is significant interest in activating photosynthesis the bundle sheath of rice such that mesophyll and bundle sheath cells could be engineered to carry out $C_4$ photosynthesis (Wang et al., 2017; Sage, 2004). The finding that photosynthesis gene expression in the 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 $C_4$ genes in the $C_3$ rice leaf

Vascular bundles of $C_3$ plants are known to carry out $C_4$-like metabolism via high activities of $C_4$ acid decarboxylases and PPDK to make use of $C_4$ 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 $C_4$ enzymes in $C_3$ plants is associated with their expression in vein or bundle sheath cells. The analysis of rice we present here shows that transcripts encoding two $C_4$ 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 $C_4$ 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 $C_3$ 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 $C_4$ 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 $C_4$ cycle into rice – amplifying the existing expression of $C_4$ 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 $C_4$ grasses *Panicum virgatum,*
597 *Sorghum bicolor, Setaria italica* and *Zea mays* as well as the $C_4$ 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 $C_4$
602 grasses and $C_3$ 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 $C_4$ 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

localized to bundle sheath cells of rice, which is consistent with expression patterns in Arabidopsis (Aubry et al., 2014b). This was also true for genes indirectly associated with sulphur assimilation such as the PPase that increases the rate of the ATP sulphurylase reaction. Several evolutionary drivers for localization of sulphur assimilation in the bundle sheath of $C_4$ grasses have been discussed including co-localisation with photorespiration as a source of serine for cysteine synthesis and protection of the reaction intermediates from oxidation by Photosystem II derived oxygen (Kopriva and Koprivova, 2005).

In Arabidopsis glucosinolate synthesis is controlled by a MYC-MYB transcription factor module that has recently been shown to pattern gene expression to the bundle sheath (Dickinson et al 2020). It seemed likely that localisation of glucosinolates to the bundle sheath of Arabidopsis led to upregulation of ATPS and APR in these cells. However, the preferential expression of genes associated with sulphur assimilation in the rice bundle sheath that does not synthesise glucosinolates indicates a much more ancient role for the bundle sheath in sulphur assimilation. It also removes the proposed connection between $C_4$ photosynthesis and localisation of sulphur assimilation in the bundle sheath. More broadly, it appears that the bundle sheath is conditioned for sulphur assimilation in all species analysed to date. In plants such as wheat, *Flaveria* and *Gynandropsis*, mesophyll cells are also used (Schmutz and Brunold, 1984; Koprivova et al., 2001; Aubry et al., 2014a), but in others the pathway appears to be restricted to the bundle sheath (Supplemental Figure 9). It is not clear whether the ancestral state was for gene expression allowing sulphur assimilation to be associated with bundle sheath and mesophyll cells or whether in some species expression in mesophyll cells has been gained. Moreover, a full complement of transporters and enzymes for nitrate reduction were found to preferentially be expressed in rice bundle sheath cells, other enzymes that serve to provide carbon skeleton for amino acid metabolism such as PEPC (Masumoto et al., 2010) and PPDK (Taylor et al., 2010) also showed high expression in these cells. The compartmentation of nitrate and ammonia assimilation gene expression between bundle sheath and mesophyll cells of rice contrasts with their spatial separation of $C_4$ grasses and *Gynandropsis gynandra*, where nitrate reduction predominately takes place in the mesophyll, and ammonia assimilation occurs in the bundle sheath (Supplemental Figure 9). The strong expression of genes associated with 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 $C_2$ and $C_4$ pathways, since modelling revealed that early changes to $C_2$ 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 $C_4$ photosynthesis (Mallmann et al., 2014). Indeed, transcripts for key genes of nitrate assimilation are consistently more strongly expressed in mesophyll cells of $C_4$ 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 $CO_2$ 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 $C_3$ 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.

## Materials and methods

### Plant growth condition and sample preparation

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

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

### Library preparation, RNA sequencing and data processing

20 ng of bundle sheath, mesophyll and veinal RNA was used as input for the Quantseq 3' mRNA-seq library preparation kit (Lexogen, Moll et al., 2014) according to the manufacturer's instructions. Libraries were sequenced using Nextseq 500 sequencer to 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.

**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 $\log_2$ transformed quantile normalized TPM (transcript per million). For functional enrichment, rice proteome (MSU version7) were firstly annotated by Mapman cateories 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

729 plotted using ggplot2 (Wickham, 2016). Heatmap of gene expression clusters and pathways
730 were plotted using ComplexHeatmap package (Gu et al., 2016). Transcript abundance were
731 presented in boxplot of TPM (transcript per million) using default settings of hinge and
732 whisker in ggplot2.
733
734 **Motif enrichment analysis**
735 Leaf DNase I hypersensitive sites (DHS) (Zhang et al., 2012) within 2000-bp of the
736 genes were extracted and used as input to the AME tool (http://meme-suite.org/tools/ame,
737 Bailey et al., 2015; McLeay and Bailey, 2010) using default settings and a custom
738 background of all DHS regions within 2000bp of gene loci as well as for FIMO command
739 line tool (Grant et al., 2011) using default settings. The meme format Jaspar Plant Non-
740 Redundant Motif Database (Fornes et al., 2020, http://jaspar2018.genereg.net/downloads/)
741 was used for both methods. FIMO scanning identified significant matches to known motifs
742 and the frequencies of these sites were statistically tested for enrichment or depletion using
743 permutation testing with the regioneR R package (Gel et al., 2016). For permutation testing,
744 all DHS regions within 2000bp of genes were used as a background for random subsampling
745 and observed frequencies were statistically tested against the random subsampling
746 distributions. As the Jaspar motif database contains many motifs with high similarity
747 predicted to be bound by closely related transcription factors, we grouped similar motifs
748 together that showed a higher than 70% overlap in predicted target sites within the DHS
749 background sequences representing motifs that were deemed highly redundant. The most
750 strongly enriched individual motif from each such group was used as a representative value
751 for the group and plotted using ggplot2 (Wickham and Sievert, 2016). In order to predict
752 which *O. sativa* transcription factors might bind predicted motifs, we mapped rice
753 transcription factors to their best match motif through protein homology with the Jaspar Non-
754 Redundant Motif database proteins (Fornes et al., 2020). The best scoring BLASTP match
755 was used as long as bit-score was greater than 100 to avoid spurious assignments to motifs
756 from other TF families. For all transcription factors found in the 6 clusters, those that were
757 mapped to an enriched motif from any of the 6 clusters was plotted using the single motif
758 enrichment scores, rice transcription factors were annotated according to funRiceGenes
759 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-
774 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 $C_4$ 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.

## Figure legends

**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

820 negative $log_{10}$ transformed FDR, size of dots (GeneRatio) represents the ratio of matched
821 genes in each category relative to total number of genes in each cluster.
822
823 **Figure 2. Preferential accumulation of transcripts associated with water transport,**
824 **sulphur and nitrogen assimilation in the rice bundle sheath.** (A) Relative transcript
825 abundance for aquaporins in mesophyll, bundle sheath and veinal cells. $Log_2$ transformed
826 quantile normalized TPM were scaled and genes with similar expression pattern were
827 clustered using hierarchical method. (B) Representative image after *in situ* hybridization
828 localization for *PIP1.1* and *PIP1.3* mRNAs, scale bars represent 20 microns, red arrows
829 indicate specific signal on bundle sheath cell periphery. (C) Schematic illustrating sulphur
830 assimilation and relative transcript abundance in mesophyll (M), bundle sheath (BS) and
831 veinal (V) cells depicted as Z-score from $log_2$ transformed quantile normalized TPM. (D)
832 Representative image after *in situ* hybridization for *ATPSb* and *SIR* mRNAs, scale bars
833 represent 20 microns, red arrows indicate specific signal on bundle sheath cell periphery.
834 (E) Transcript abundance of *CLT1* and *HAC1;1*. (F) Relative transcript abundance for genes
835 involved in nitrogen assimilation.
836
837 **Figure 3. Relative transcript abundance for genes involved in Calvin Benson**
838 **Bassham cycle (B) and photorespiration (C) and $C_4$ pathway (D) in mesophyll (M),**
839 **bundle sheath (BS) and veinal (V) cells of rice.** (A) Paradermal section of rice leaf stained
840 with toluidine blue shows bundle sheath cells are less occupied by chloroplasts compared
841 with mesophyll cells, scale bar represents 50 microns. (B&C) Transcripts associated with
842 Calvin Benson Bassham cycle (B) and photorespiration (C) were preferentially expressed in
843 mesophyll cells, $log_2$ transformed quantile normalized TPM were scaled and genes with
844 similar expression pattern were clustered using hierarchical method**.** (D) Transcripts
845 encoding the $C_4$ acid decarboxylases NADP-DEPENDENT MALIC ENZYME (NADP-ME)
846 and PHOSPHOENOLPYRUVATE CARBOXYKINASE (PCK) accumulated preferentially in
847 bundle sheath and veinal cells whilst ancillary enzymes AMP KINASE (AMK) and
848 PYROPHOSPHORYLASE (PPASE) for PYRUVATE,ORTHOPHOSPHATE DIKINASE
849 (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).

**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

881 sheath cells. (A) The enriched Mapman categories (secondary level) of BS and M
882 preferential genes in rice and Arabidopsis were defined using Fisher's exact test (FDR<0.1).
883 (B) Venn diagram illustrating the extent to which genes in the same orthogroup are
884 preferentially expressed in mesophyll or bundle sheath cells of both rice and
885 Arabidopsis. (C-G) Transcript abundance of Arabidopsis and rice genes belonging to the
886 same orthogroups of aqaporins (C), sulphate transport (D), sulphate assimilation (E) ,
887 jasmonic acid biosynthesis (F), and transcription factors (G). Data are presented as TPM
888 and statistically significant differences annotated with an asterisk (FDR and adjust $P < 0.05$
889 using edgeR and DESeq2 analysis in this study, PPDE>0.95 in Aubry et al., 2014b), red and
890 green bars represent bundle sheath and mesophyll respectively.
891
892 **Supplemental Data**
893 **Supplemental Figure S1.** Transcript abundance of genes previously reported to be
894 associated with mesophyll, bundle sheath or veins.
895 **Supplemental Figure S2.** Mapman categories and metabolic overview of pairwise
896 comparisons between bundle sheath (BS) and vein (V) and between mesophyll (M) and vein
897 (V).
898 **Supplemental Figure S3.** Enriched transporter families in the six gene expression clusters.
899 **Supplemental Figure S4.** Transcript abundance of genes associated with nitrogen
900 assimilation.
901 **Supplemental Figure S5.** Transcript abundance of genes associated with the
902 photosynthetic election transport chain (A, B), Calvin Benson Bassham cycle (C), starch
903 biosynthesis (D) and photorespiration (E).
904 **Supplemental Figure S6.** Transcription factors associated with different clusters.
905 **Supplemental Figure S7.** Transcript abundance of orthogroups shared by bundle sheath
906 cells in rice and Arabidopsis associated with aquaporins, sulphate transport and assimilation,
907 jasmonic acid biosynthesis, and transcription factors.
908 **Supplemental Figure S8.** Comparison of bundle sheath preferentially expressed genes
909 among different studies.
910 **Supplemental Figure S9.** Transcript abundance of Aquaporins (*PIP1* and *PIP2*), *ATP*
911 *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. virgartum* (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.

924 **References**

925 **Attia, Z., Dalal, A., and Moshelion, M.** (2020). Vascular bundle sheath and mesophyll cells
926 modulate leaf water balance in response to chitin. Plant J. **101**: 1368–1377.

927 **Aubry, S., Kelly, S., Kümpers, B.M.C.C., Smith-Unna, R.D., and Hibberd, J.M.** (2014a).
928 Deep Evolutionary Comparison of Gene Expression Identifies Parallel Recruitment of
929 Trans-Factors in Two Independent Origins of $C_4$ Photosynthesis. PLoS Genet. **10**:
930 e1004365.

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

934 **Bailey, T.L., Johnson, J., Grant, C.E., and Noble, W.S.** (2015). The MEME Suite. Nucleic
935 Acids Res. **43**: 39–49.

936 **Berardini, T.Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., and Huala, E.**
937 (2015). The *Arabidopsis* information resource: Making and mining the "gold standard"
938 annotated reference plant genome. Genesis **53**: 474–485.

939 **Berry, J.A.** (2012). There Ought to Be an Equation for That. Annu. Rev. Plant Biol. **63**: 1–
940 17.

941 **Bezrutczyk, M., Zöllner, N.R., Kruse, C.P.S., Hartwig, T., Lautwein, T., Köhrer, K.,**
942 **Frommer, W.B., and Kim, J.-Y.** (2021). Evidence for Phloem Loading via the Abaxial
943 Bundle Sheath Cells in Maize Leaves. Plant Cell: doi.org/10.1093/plcell/koaa055.

944 **Birnbaum, K., Shasha, D.E., Wang, J.Y., Jung, J.W., Lambert, G.M., Galbraith, D.W.,**
945 **and Benfey, P.N.** (2003). A gene expression map of the *Arabidopsis* root. Science **302**:
946 1956–60.

947 **Brady, S.M., Orlando, D.A., Lee, J.Y., Wang, J.Y., Koch, J., Dinneny, J.R., Mace, D.,**
948 **Ohler, U., and Benfey, P.N.** (2007). A high-resolution root spatiotemporal map reveals
949 dominant expression patterns. Science **318**: 801–806.

950 **Brown, N.J. et al.** (2010). $C_4$ acid decarboxylases required for $C_4$ photosynthesis are active
951 in the mid-vein of the $C_3$ species *Arabidopsis thaliana*, and are important in sugar and
952 amino acid metabolism. Plant J. **61**: 122–133.

953 **Burgener, M., Suter, M., Jones, S., and Brunold, C.** (1998). Cyst(e)ine Is the Transport
954 Metabolite of Assimilated Sulfur from Bundle-Sheath to Mesophyll Cells in Maize

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

**Burnell, J.N.** (1984). Sulfate Assimilation in $C_4$ Plants. Plant Physiol. **75**: 873–875.

**von Caemmerer, S., Edwards, G.E., Koteyeva, N., and Cousins, A.B.** (2014). Single cell $C_4$ photosynthesis in aquatic and terrestrial plants: A gas exchange perspective. Aquat. Bot. **118**: 71–80.

**von Caemmerer, S., Quick, W.P., and Furbank, R.T.** (2012). The development of $C_4$ rice: Current progress and future challenges. Science **336**: 1671–1672.

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

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

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

**Chen, T., Wu, H., Wu, J., Fan, X., Li, X., and Lin, Y.** (2017). Absence of *OsβCA1* causes a $CO_2$ deficit and affects leaf photosynthesis and the stomatal response to $CO_2$ in rice. Plant J. **90**: 344–357.

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

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

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

**Edwards, G.E. and Black, C.C.** (1971). Isolation of Mesophyll Cells and Bundle Sheath 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 $C_4$ Photosynthesis Provides New Insight into the Mechanism of Phloem Loading in $C_4$ 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 $C_4$ 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 $C_4$ 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 $C_3$ and $C_4$ plants. Plant. Cell Environ. **36**: 249–261.

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

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

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

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

**Hatch, M.D.** (1987). $C_4$ photosynthesis: a unique elend of modified biochemistry, anatomy and ultrastructure. BBA Rev. Bioenerg. **895**: 81–106.

**Hatch, M.D. and Burnell, J.N.** (1990). Carbonic anhydrase activity in leaves and its role in the first step of $C_4$ photosynthesis. Plant Physiol. **93**: 825–828.

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

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

**Hibberd, J.M. and Quick, W.P.** (2002). Characteristics of $C_4$ photosynthesis in stems and petioles of $C_3$ flowering plants. Nature **415**: 451–454.

**Hibberd, J.M., Sheehy, J.E., and Langdale, J.A.** (2008). Using $C_4$ photosynthesis to increase the yield of rice—rationale and feasibility. Curr. Opin. Plant Biol. **11**: 228–231.

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

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

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

**Janacek, S.H., Trenkamp, S., Palmer, B., Brown, N.J., Parsley, K., Stanley, S., Astley, H.M., Rolfe, S.A., Paul Quick, W., Fernie, A.R., and Hibberd, J.M.** (2009). Photosynthesis in cells around veins of the $C_3$ plant *Arabidopsis thaliana* is important for both the shikimate pathway and leaf senescence as well as contributing to plant fitness. Plant J. **59**: 329–343.

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

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

**John, C.R., Smith-Unna, R.D., Woodfield, H., Covshoff, S., and Hibberd, J.M.** (2014). Evolutionary Convergence of Cell-Specific Gene Expression in Independent Lineages of $C_4$ Grasses. Plant Physiol. **165**: 62–75.

**Jurić, I., González-Pérez, V., Hibberd, J.M., Edwards, G., and Burroughs, N.J.** (2016). Size matters for single-cell $C_4$ photosynthesis in *Bienertia*. J. Exp. Bot. **68**: 255–267.

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

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

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

**Kopriva, S. and Koprivova, A.** (2005). Sulfate assimilation and glutathione synthesis in $C_4$ plants. Photosynth. Res. **86**: 363–372.

**Koprivova, A., Melzer, M., Von Ballmoos, P., Mandel, T., Brunold, C., and Kopriva, S.** (2001). Assimilatory sulfate reduction in $C_3$, $C_3$-$C_4$, and $C_4$ species of *Flaveria*. Plant Physiol. **127**: 543–550.

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

**Langdale, J.A.** (2011). $C_4$ Cycles: Past, Present, and Future Research on $C_4$ Photosynthesis. Plant Cell **23**: 3879–3892.

**Leegood, R.C.** (2008). Roles of the bundle sheath cells in leaves of $C_3$ plants. J. Exp. Bot. **59**: 1663–1673.

**Li, G.W., Zhang, M.H., Cai, W.M., Sun, W.N., and Su, W.A.** (2008). Characterization of OsPIP2;7, a Water Channel Protein in Rice. Plant Cell Physiol. **49**: 1851–1858.

**Love, M.I., Huber, W., and Anders, S.** (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. **15**: 550.

**Lunn, J.E. and Furbank, R.T.** (1997). Localisation of sucrose-phosphate synthase and starch in leaves of $C_4$ plants. Planta **202**: 106–111.

**Mallmann, J., Heckmann, D., Bräutigam, A., Lercher, M.J., Weber, A.P.M., Westhoff, P., and Gowik, U.** (2014). The role of photorespiration during the evolution of $C_4$ photosynthesis in the genus *Flaveria*. Elife **3**: e02478.

**Maruyama-Nakashita, A., Nakamura, Y., Tohge, T., Saito, K., and Takahashi, H.** (2006). *Arabidopsis* SLIM1 is a central transcriptional regulator of plant sulfur response and metabolism. Plant Cell **18**: 3235–3251.

**Masumoto, C., Miyazawa, S.-I.I., Ohkawa, H., Fukuda, T., Taniguchi, Y., Murayama, S., Kusano, M., Saito, K., Fukayama, H., and Miyao, M.** (2010). Phospho*enol*pyruvate carboxylase intrinsically located in the chloroplast of rice plays a crucial role in ammonium assimilation. Proc. Natl. Acad. Sci. U. S. A. **107**: 5226–31.

**Maughan, S.C. et al.** (2010). Plant homologs of the *Plasmodium falciparum* chloroquine-resistance transporter, *Pf*CRT, are required for glutathione homeostasis and stress responses. Proc. Natl. Acad. Sci. U. S. A. **107**: 2331–2336.

**McLeay, R.C. and Bailey, T.L.** (2010). Motif Enrichment Analysis: A unified framework and an evaluation on ChIP data. BMC Bioinformatics **11**: 165.

**Miyake, H.** (2016). Starch Accumulation in the Bundle Sheaths of $C_3$ Plants: A Possible Pre-Condition for $C_4$ Photosynthesis. Plant Cell Physiol. **57**: 890–896.

**Moll, P., Ante, M., Seitz, A., and Reda, T.** (2014). QuantSeq 3' mRNA sequencing for RNA quantification. Nat. Methods **11**: i–iii.

**Moore, B. d., Ku, M.S.B., and Edwards, G.E.** (1984). Isolation of leaf bundle sheath protoplasts from $C_4$ dicot species and intracellular localization of selected enzymes. Plant Sci. Lett. **35**: 127–138.

**Mustroph, A., Zanetti, M.E., Jang, C.J.H.H., Holtan, H.E., Repetti, P.P., Galbraith, D.W.,**

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

**Nomura, M., Higuchi, T., Ishida, Y., Ohta, S., Komari, T., Imaizumi, N., Miyao-Tokutomi, M., Matsuoka, M., and Tajima, S.** (2005). Differential Expression Pattern of $C_4$ Bundle Sheath Expression Genes in Rice, a $C_3$ Plant. Plant Cell Physiol **46**: 754–761.

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

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

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

**Rao, X., Lu, N., Li, G., Nakashima, J., Tang, Y., and Dixon, R.A.** (2016). Comparative cell-specific transcriptomics reveals differentiation of $C_4$ photosynthesis pathways in switchgrass and other $C_4$ lineages. J. Exp. Bot. **67**: 1649–1662.

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

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

**Sage, R.F.** (2001). Environmental and Evolutionary Preconditions for the Origin and Diversification of the $C_4$ Photosynthetic Syndrome. Plant Biol. **3**: 202–213.

**Sage, R.F.** (2004). The evolution of $C_4$ photosynthesis. New Phytol. **161**: 341–370.

**Sage, R.F., Sage, T.L., and Kocacinar, F.** (2012). Photorespiration and the Evolution of $C_4$ Photosynthesis. Annu. Rev. Plant Biol. **63**: 19–47.

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

Refixation of Photorespiratory $CO_2$ and Efforts to Engineer $C_4$ Photosynthesis into Rice. Plant Cell Physiol. **50**: 756–772.

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

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

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

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

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

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

**Shen, W., Ye, L., Ma, J., Yuan, Z., Zheng, B., Chuangen, L. V., Zhu, Z., Chen, X., Gao, Z., and Chen, G.** (2016). The existence of $C_4$-bundle-sheath-like photosynthesis in the mid-vein of $C_3$ rice. Rice **9**: 20.

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

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

**Thimm, O., Bläsing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., Selbig, J., Müller, 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 $C_4$ 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 $C_4$ 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 $C_3$ to $C_4$ 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.

Bioinforma. **69**.

**Zelazny, E., Borst, J.W., Muylaert, M., Batoko, H., Hemminga, M.A., and Chaumont, F.** (2007). FRET imaging in living maize cells reveals that plasma membrane aquaporins interact to regulate their subcellular localization. Proc. Natl. Acad. Sci. U. S. A. **104**: 12359–12364.

**Zhang, W., Wu, Y., Schnable, J.C., Zeng, Z., Freeling, M., Crawford, G.E., and Jiang, J.** (2012). High-resolution mapping of open chromatin in the rice genome. Genome Res. **22**: 151–162.

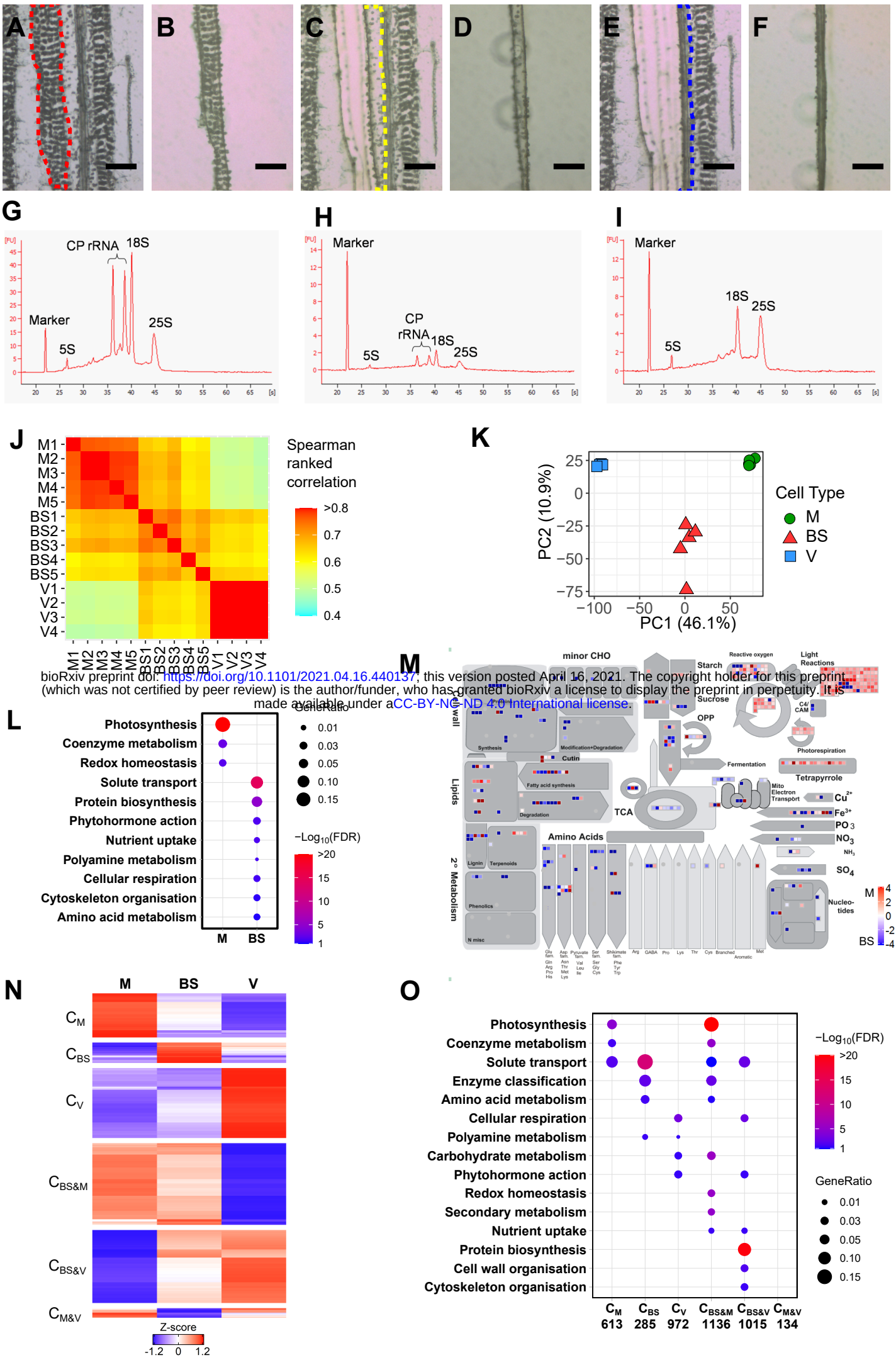


**Figure 1. RNA was isolated from mesophyll, bundle sheath and veinal cells of rice using laser capture microdissection.** Each cell-type was identified and then sequentially removed from paradermal sections prior to RNA quality being assessed. (A&B) Representative image of mesophyll cells outlined with 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.

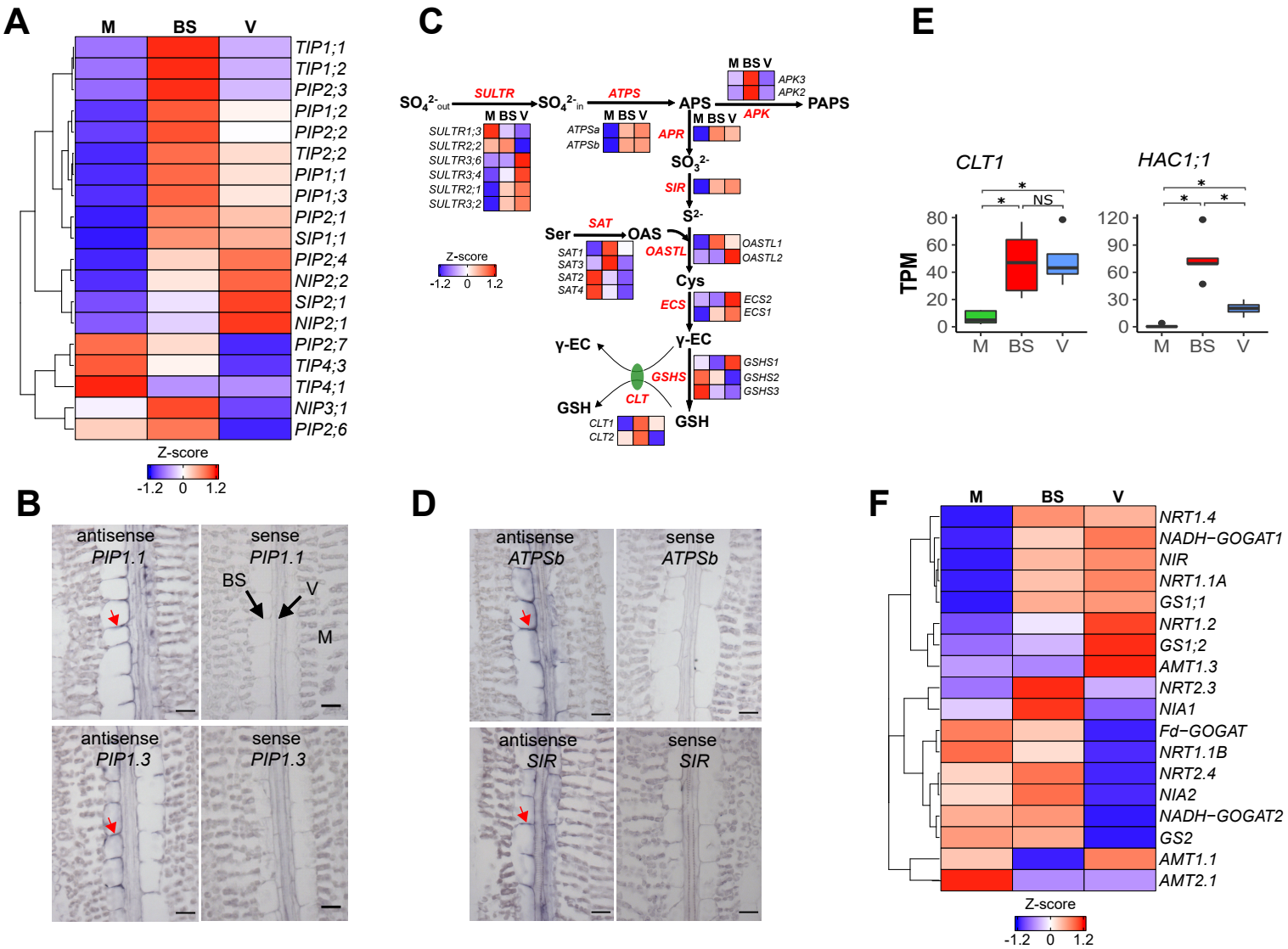


**Figure 2. Preferential accumulation of transcripts associated with water transport, sulphur and nitrogen assimilation in the rice bundle sheath.** (A) Relative transcript abundance for aquaporins in mesophyll, bundle sheath and veinal cells. $Log_2$ transformed quantile normalized TPM were scaled and 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 $log_2$ 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; GSHS, glutathione synthetase. Abbreviations for (F): NRT, nitrate transporter; AMT, ammonia transporter; NIA, nitrate reductase; NIR, nitrite reductase; GS, glutamine synthetase; NADH-GOGAT, NADH-dependent glutamate synthase; Fd-GOGAT, ferredoxin-dependent glutamate synthase.

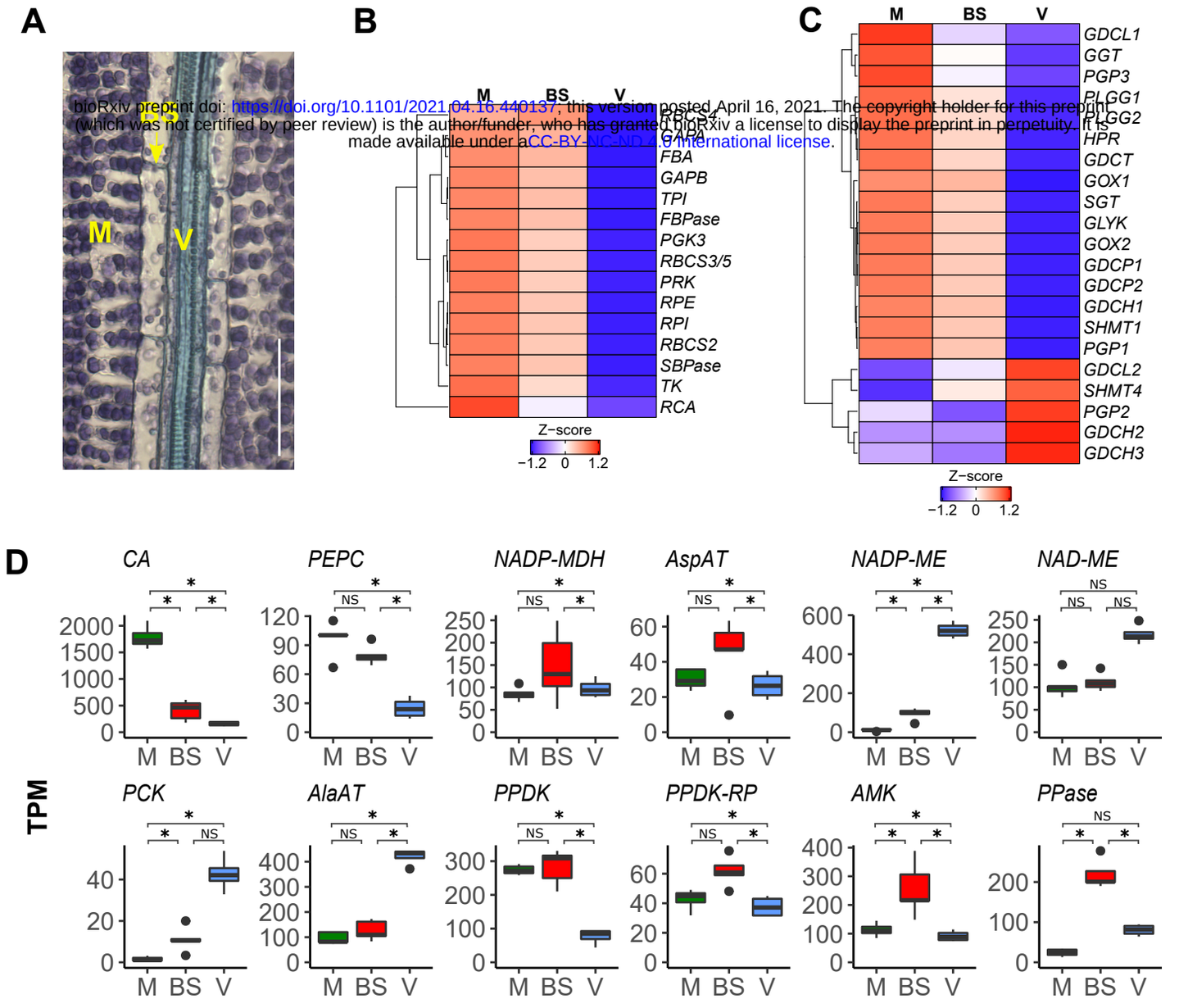


**Figure 3. Relative transcript abundance for genes involved in Calvin Benson Bassham cycle (B) and photorespiration (C) and $C_4$ pathway (D) in mesophyll (M), bundle sheath (BS) and veinal (V) cells of rice.** (A) Paradermal section of rice leaf stained with toluidine blue shows bundle sheath cells 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; GAPA/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, phospho*enol*pyruvate carboxylase; NADP-MDH, NADP-dependent malic dehydrogenase; AspAT, aspartate aminotransferase ; NADP-ME, NADP-dependent malic enzyme; NAD-ME, NAD-dependent malic enzyme; PCK, phospho*enol*pyruvate carboxykinase; AlaAT, alanine aminotransferase; PPDK, pyruvate,orthophosphate dikinase; PPDK-RP, regulatory protein of PPDK; AMK, AMP kinase; PPase, pyrophosphorylase.

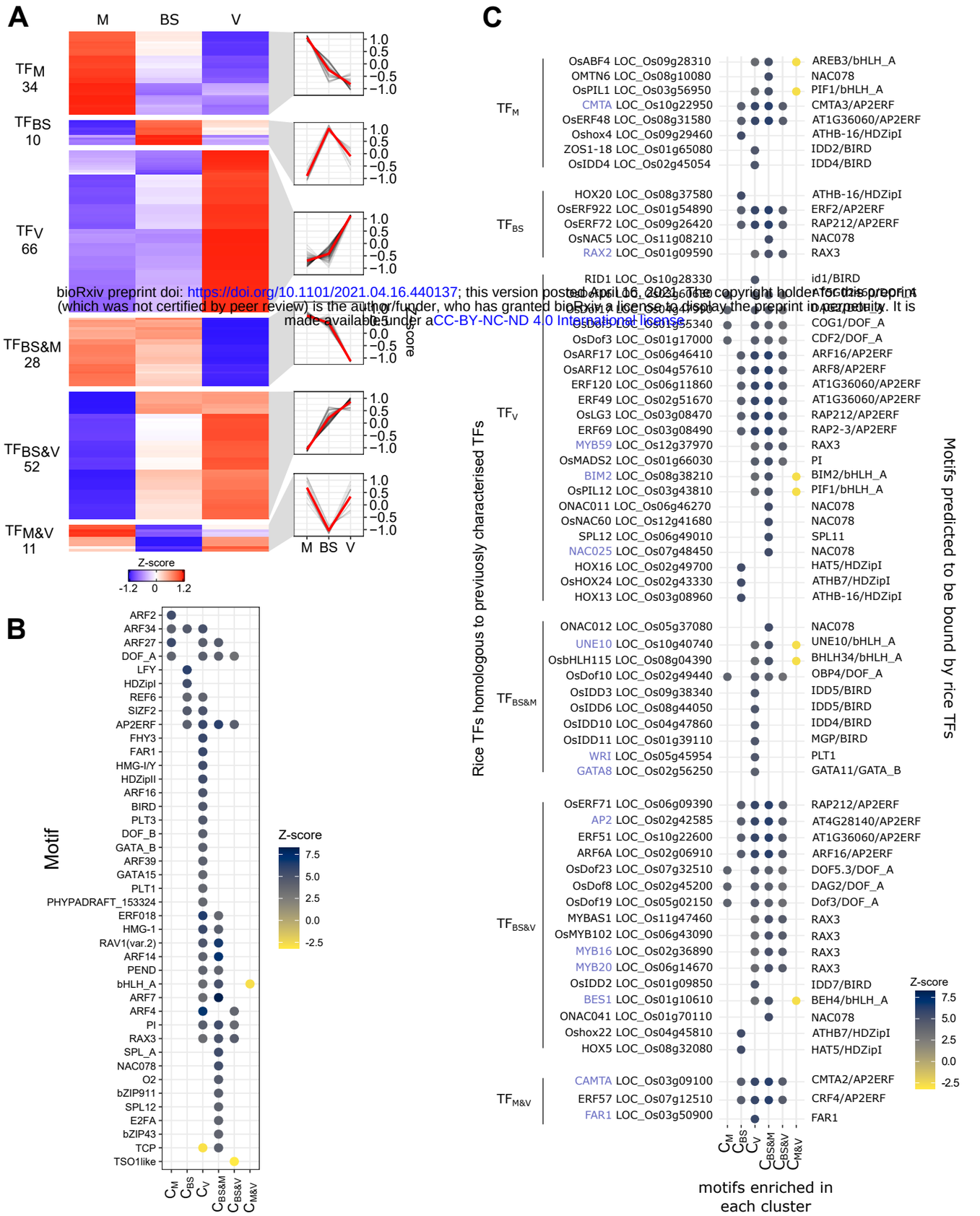


**Figure 4 Patterning of transcription factors between mesophyll, bundle sheath and veinal cells of rice. (A)** Transcription factors from cluster $C_M$, $C_{BS}$, $C_V$, $C_{BS\&M}$, $C_{BS\&V}$, and $C_{M\&V}$ were designated as $TF_M$, $TF_{BS}$, $TF_V$, $TF_{BS\&M}$, $TF_{BS\&V}$, and $TF_{M\&V}$ 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.

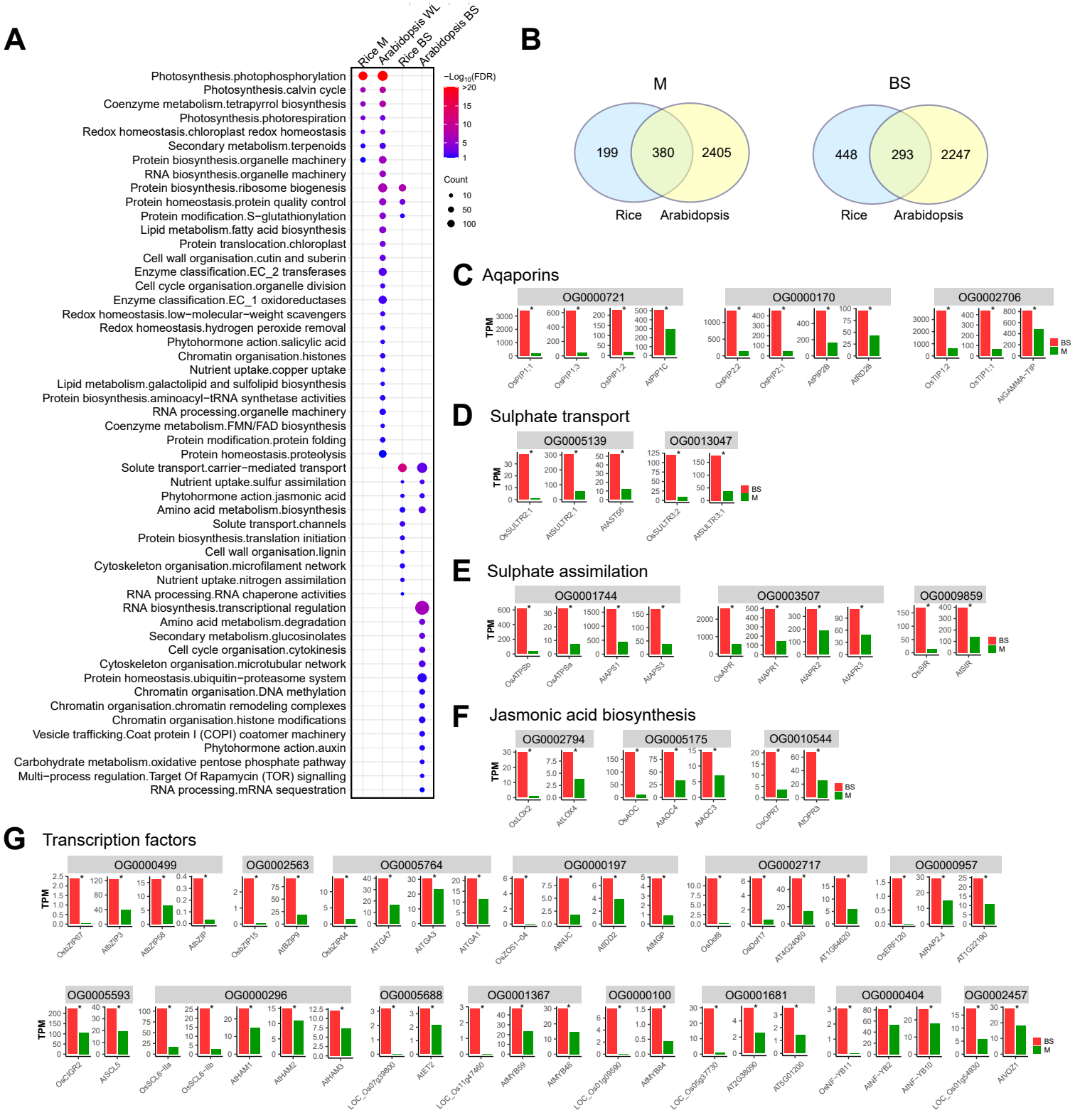


**Figure 5. Conserved patterning of gene expression in the Arabidopsis and rice bundle sheath**. Orthologs from Arabidopsis and rice associated with aquaporins, sulphate transport and assimilation as well as jasmonic acid biosynthesis are strongly expressed in bundle sheath cells. (A) The enriched Mapman categories (secondary level) of BS and M preferential genes in rice and Arabidopsis were defined using Fisher's exact test (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 aqaporins (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.