**Title:** *Sorghum bicolor* cultivars have divergent and dynamic gene regulatory networks that control the temporal expression of genes in stem tissue

**Authors:** Madara Hetti Arachchilage[1], John E. Mullet[2], and Amy Marshall-Colon*[1,3]

[1]Center for Advanced Bioenergy and Bioproducts Innovation BRC, University of Illinois Urbana-Champaign, Urbana, IL, USA 61801

[2]Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX, 77843 USA

[3]Department of Plant Biology, University of Illinois Urbana-Champaign, 265 Morrill Hall, MC-116, 505 South Goodwin Ave., Urbana, IL 61801, USA

*For correspondence (email: amymc@illinois.edu)

**Running title:** Temporal and spatial gene networks of sorghum stems

**Keywords:** Stem-specific genes, *BTx623*, *Sweet Della*, *Sorghum bicolor*, Tissue-specific gene regulatory network, Spatial-temporal gene regulation, Cis-regulatory elements, Single nucleotide variants, Structural variants

**Abstract:** The genetic engineering of value-added traits such as the accumulation of bioproducts in high biomass C4 grass stems is one promising strategy to make plant-derived biofuels more economical for industrial use. A first step toward achieving this goal is to identify stem-specific promoters that can drive the expression of genes of interest with good temporal and spatial specificity. However, a comprehensive characterization of the spatial-temporal regulatory elements of stem tissue-specific promoters for C4 grasses has not been reported. Therefore, we performed an *in-silico* analysis on *Sorghum bicolor cv BTx623* transcriptomes from multiple tissues over development to identify stem-expressed genes. The analysis identified 10 genes that are "Always-On-Stem-Specific," 59 genes that are "Temporally-Stem-Specific during early development," and 21 genes that are "Temporally-Stem-Specific during late development." Promoter analysis revealed common and/or unique cis-regulatory elements in promoters of genes

within each of the three categories. Subsequent gene regulatory network (GRN) analysis revealed that different transcriptional regulatory programs are responsible for the temporal activation of the stem-expressed genes. The analysis of temporal stem GRNs between sweet (*cv Della*) and grain (*cv BTx623*) sorghum varieties revealed genetic variation that could influence the regulatory landscape. This study provides new insights about sorghum stem biology, and information for future genetic engineering efforts to fine-tune the spatial-temporal expression of transgenes in C4 grass stems.

## Introduction

High biomass grasses, such as sorghum (*Sorghum bicolor*), miscanthus (*Miscanthus x giganteus*), switchgrass (*Panicum virgatum*), and sugarcane (*Saccharum sp.*), offer an abundant and renewable source of biomass for biofuels production (McLaughlin and Kszos, 2005; Rooney *et al.*, 2007; Karp and Shield, 2008; Murray *et al.*, 2008; Waclawovsky *et al.*, 2010a; Byrt, Grof and Furbank, 2011; Sage and Zhu, 2011; Feltus and Vandenbrink, 2012; Olson *et al.*, 2012; Mullet *et al.*, 2014; Mathur *et al.*, 2017) providing a sustainable alternative to conventional fossil fuels (Demirbas, 2007; Demirbas and Demirbas, 2010; Kumar, Long and Singh, 2018; Kumar, 2020). Grasses with C4 photosynthesis can produce high biomass in marginal lands with low water and nitrogen inputs, leading to energy conservation (McLaughlin *et al.*, 2002; Heaton, Dohlemnn and Long, 2008; Varvel *et al.*, 2008). However, the cost associated with production of fuels and other bioproducts from plant biomass is relatively high, which has been a major constraint on the widespread adoption of grasses as sources of feedstock to produce biofuels or bioproducts (Hill *et al.*, 2006; Kumar, Long and Singh, 2018). Previous work showed that C4 grasses can be genetically engineered for increased fatty acid production, resulting in biomass with high oil content (Huang, Long and Singh, 2015; Huang *et al.*, 2016; Wang, 2016; Zale *et al.*, 2016; Beechey-Gradwell *et al.*, 2020; Mitchell *et al.*, 2020; Parajuli *et al.*, 2020), which increases their value as biofuels. Genetically engineering high-biomass grasses to accumulate valuable bioproducts that require minimal processing post-harvest is a desirable way forward, but has many challenges.

Seeds are the primary site of lipid synthesis and storage in plants (Harwood *et al.*, 1971; Dyer and Mullen, 2005; Lung and Weselake, 2006; Graham, 2008; Koçar and Civaş, 2013; Li *et al.*, 2016; Baud, 2018) in the form of triacylglycerols (TAGs) which have twice the energy density of carbohydrates and can be easily converted to other biofuel products. While oil-accumulating

seeds have been a source of lipids for the biofuel industry, use of oil seeds or fruits for biofuels could negatively impact food supplies and prices. Therefore, genetic engineering of vegetative tissues to accumulate higher triacylglycerol (TAG) levels has drawn attention (Durrett, Benning and Ohlrogge, 2008; Qazi, Paranjpe and Bhargava, 2012; Napier *et al.*, 2014a; Kumar, 2020). Since C4 grasses can be grown on marginal lands not suitable for most food crops, the high biomass producing vegetative tissues of sorghum, miscanthus, and sugarcane are desirable targets for engineering efforts aimed at producing and/or storing lipids (Carlsson *et al.*, 2011; Sanjaya *et al.*, 2011; Weselake, 2016a; Lima *et al.*, 2017). For example, stem tissues engineered to divert energy from nonstructural carbohydrates into the lipid biosynthesis pathway through strategies such as increasing the supply of FAs, increasing TAG assembly activities, and blocking TAG breakdown pathways have resulted in higher amounts of TAGs accumulating in vegetative tissues (Papini-Terzi *et al.*, 2009; Waclawovsky *et al.*, 2010b; Sanjaya *et al.*, 2011; Qazi *et al.*, 2014; Sekhon *et al.*, 2016). Other proof-of-concept studies for increasing TAGs in vegetative tissues have been performed in *Arabidopsis thaliana, Brachypodium distachyon, Nicotiana benthamiana, Nicotiana tabacum,* sugarcane (*Saccharum* spp.)*,* and *Sorghum bicolor* (Thelen and Ohlrogge, 2002; Fan, Yan and Xu, 2013; Vanhercke *et al.*, 2013, 2019; Yang *et al.*, 2015; Zale *et al.*, 2016; Mitchell *et al.*, 2020; Parajuli *et al.*, 2020; Xu *et al.*, 2020). For example, in engineered sugarcane, TAGs accumulated to an average of 8.0% of the dry weight of leaves and 4.3% of the dry weight of stems (Huang, Long and Singh, 2015; Zale *et al.*, 2016; Parajuli *et al.*, 2020). Another study in tobacco achieved TAGs accumulation up to 19% of the dry weight of the total biomass production by over-expressing the genes encoding WRINKLED1, DGAT, and oleosins (Vanhercke *et al.*, 2014; Zale *et al.*, 2016). However, many of these engineering efforts to increase TAG accumulation in immature vegetative tissues have resulted in negative impacts on plant growth as observed in sorghum, potato, tobacco, and Arabidopsis (Slocombe *et al.*, 2009; Feltus and Vandenbrink, 2012; Kelly *et al.*, 2013; Vanhercke *et al.*, 2014, 2019; Hofvander *et al.*, 2016; Liu *et al.*, 2017; Ramšak *et al.*, 2018; Xiaoyu Xu *et al.*, 2019; Xu *et al.*, 2020; Mitchell *et al.*, 2020). One hypothesis is that driving lipid accumulation under the control of tissue- and/or developmental-stage specific promoters, specifically those active during late development (Moyle and Birch, 2013; Mudge *et al.*, 2013), will have less of an impact on photosynthetic efficiencies and plant growth than constitutive overexpression of genes of interest.

Sink–source dynamics within the plant direct how much, where, and when carbohydrates are allocated, and determine final biomass of vegetative plant organs and seed. Knowledge of the spatial and temporal heterogeneity of carbon sinks among tissues and cell types in vegetative tissues during development could enable more informed genetic engineering (Schnyder, 1993; Hoffmann-Thoma *et al.*, 1996; Wang *et al.*, 2007; Sanjaya *et al.*, 2011; Slewinski, 2012; da Costa *et al.*, 2014; Sekhon *et al.*, 2016; Hennet *et al.*, 2020). For example, previous studies reported that sorghum, sugarcane, and wheat store excess carbohydrates in the form of soluble sugars in stem pith parenchyma, in close proximity to vascular bundles located within internode, and these stem carbohydrate reserves can be remobilized and partitioned to the developing grain during the reproductive phase (Hoffmann-Thoma *et al.*, 1996; Douglas *et al.*, 2002; Walsh, Sky and Brown, 2005; Ruuska *et al.*, 2006; Tarpley and Vietor, 2007; Rae *et al.*, 2009; Qazi, Paranjpe and Bhargava, 2012; Sekhon *et al.*, 2016; Kebrom, McKinley and Mullet, 2017; McKinley *et al.*, 2018). In such cases, tissue and developmental-stage specific promoters could be leveraged to divert the nonstructural carbohydrates to tissues engineered for elevated TAG production and accumulation, potentially avoiding the previously observed inhibition of photosynthesis and growth at early developmental stages. Many tissue-specific promoters have been successfully characterized and applied to targeted genetic engineering of maize, rice, soybean, tomato, potato, and tobacco (Deikman, Kline and Fischer, 1992; Trindade *et al.*, 2003; Cao *et al.*, 2007; Li *et al.*, 2013, 2019; Molla *et al.*, 2013; Dutt *et al.*, 2014; Napier *et al.*, 2014b, 2014a; Chen *et al.*, 2015; Wang *et al.*, 2017). Thus, increasing oil accumulation in stems of biofuel crops without impacting plant growth will require manipulation of developmental-stage specific and stem-specific promoters. Potential strategies include: a) converting sucrose to lipid in specific cells/tissues at stages when non-structural carbohydrates are highly abundant, and b) increasing stem and other vegetative tissue carbon sink strength, and sink storage capacity at late developmental stages (Slocombe *et al.*, 2009; Zhang *et al.*, 2010; P. Joshi and Nookaraju, 2012; Kim *et al.*, 2015; Weselake, 2016b; Xu and Shanklin, 2016; Shih, 2018).

Cis -regulatory elements in promoters upstream of a gene provide specific binding sites for their corresponding transcription factors (TFs), which regulate the transcription of gene expression. Studies in Arabidopsis and a few crop species have uncovered cis-regulatory elements that govern stem vascular tissue specific expression patterns, including phloem-specific, xylem-specific, and companion cell-specific promoter elements (***Table S1***) (Dutt *et al.*, 2014). However,

a comprehensive characterization of the spatial-temporal regulatory elements of stem specific promoters for C4 grasses has not been reported but is necessary to understand the regulation of stem-expressed genes. Therefore, we performed an *in-silico* genome-wide analysis to identify stem-specific genes in Sorghum and the regulatory network that drives their spatial-temporal expression dynamics. We also examined the conservation/divergence of TF regulators of stem-specific genes between the two sorghum cultivars, sweet sorghum (*Sorghum bicolor cv* Della) and grain sorghum (*Sorghum bicolor cv* BTx623).

## Results and Discussion

We analyzed a gene expression atlas of sorghum tissues (McCormick *et al.*, 2018a) to identify genes specifically expressed in stem tissue over development. For this purpose, we used 38 expression profiles of five different *Sorghum bicolor* (*cv* BTx623) organs (stem, leaves, roots, peduncle, panicle, seeds) across five developmental stages (juvenile, vegetative, floral initiation, anthesis, grain maturity) to perform a tissue specificity analysis. We explored the gene regulatory networks (GRNs) for stem-expressed genes in an attempt to identify regulatory promoter elements that may contribute to the tissue-specific expression of these genes. The identification of such elements may help to uncover a stem-specific promoter that could be used as a molecular tool for the temporal and spatial expression of genes of interest that are not normally expressed in stem tissue. We performed further investigation to identify genetic variations in stem-preferred genes and their regulators, which cause large changes in the stem-specific GRNs between sweet sorghum (*Sorghum bicolor cv* Della) and grain sorghum (*Sorghum bicolor cv* BTx623).

### Tau index identifies stem-expressed genes over development

Several methods have been developed to determine tissue-specific gene expression (Huminiecki, Lloyd and Wolfe, 2003; Schug *et al.*, 2005; Yanai *et al.*, 2005; Smith *et al.*, 2006; Yu *et al.*, 2006; Xiao *et al.*, 2010; Julien *et al.*, 2012; Uhlen *et al.*, 2015; O'Hagan *et al.*, 2018). A recent benchmark study compared nine algorithms to determine tissue specificity and identified that the Tau ($\tau$) index was the most robust analysis (Kryuchkova-Mostacci and Robinson-Rechavi, 2016). According to the tissue-specificity benchmark study, $\tau$ value analysis identified the most known tissue-specific

genes in the human genome; including genes not detected by other algorithms, and avoided bias toward classifying most genes as "housekeeping."

In this study, we used τ index to estimate the tissue-specificity of expressed genes at a given developmental stage. A summary of the number of tissue-specific genes found for each tissue type at each development stage is listed in ***Table S2***. A small portion of genes exhibited tissue-specific expression at each of the developmental stages (τ = 1) in stem, root, leaf, peduncle, panicle and seed tissues. There were fewer tissue-specific genes in stem tissue compared to leaf, root, and reproductive tissues. We observed a higher number of stem-specific genes during early stages compared to late stages (157, 133, 21, 21 and 44 stem-specific genes at juvenile, vegetative, floral, anthesis, and grain stages, respectively). Of the stem-specific genes identified at each developmental stage, some genes were always expressed in the stem throughout development, while some genes were transiently stem-specific over development. In this study, we classified three groups of stem-expressed genes: AOSS (Always-On-Stem-Specific), TSS-Early (Temporally-Stem-Specific during early development), and TSS-Late (Temporally-Stem-Specific during late development), in which there were 10 AOSS, 59 TSS-Early, and 21 TSS-Late genes (***Table S3***).

**Expression patterns of AOSS, TSS-Early, TSS-Late genes and their functional characterization**

There were 10 genes in the AOSS group that had high stem-specificity (no significant expression at threshold of 5 TPM in non-stem tissues). Specifically, Sobic.006G024800 (an Andropogoneae family gene), Sobic.006G026300 (a Poaceae family gene) and Sobic.003G272000 (a Poales family gene) had high tissue specificity (τ = 1) and high expression (expression > 100 TPM) across the five developmental time points (***Figure 2***, ***Figure S1***). The high stem-specificity value for these genes in *Sorghum cv BTx623* make them interesting candidates for further investigation. Other genes in the AOSS group had high stem-specificity at early developmental stages; however, during the later stages these genes were expressed at similar levels among vegetative stems, the developing peduncle, and panicles which also contain stem tissue (rachis, pedicles) (***Figure 2***).These genes include Sobic.001G106200 (KNAT1; TALE TF family gene with KNOX DNA binding domain), Sobic.001G106000 (KNAT1; TALE TF), Sobic.010G194600 (HSP90-like ATPase family protein), Sobic.006G147500 (light sensitive

hypocotyls 6, LSH6), Sobic.002G419500 (HAD superfamily, subfamily IIIB acid phosphatase), Sobic.003G007200 (TIP4; tonoplast intrinsic protein), Sobic.007G008700 (ARM repeat superfamily protein, Poaceae specific family). KNAT1 is involved in meristem maintenance promoting cell proliferation and inhibiting differentiation and was previously shown to move between adjacent cells via intercellular pores formed by plasmodesmata in Arabidopsis (Kehr & Kragler, 2018). In Arabidopsis, *knat1* mutant plants had reduced formation of xylem fibers (Helizon *et al.*, 2014a).

There were 59 TSSE genes that are "stem-preferred" rather than stem-specific, which are highly expressed in the stem during early development, but have no, low, or moderate expression in non-stem tissues (leaf, root, seed) during late development. The TSSE genes that have high stem specificity ($\tau = 1$) during vegetative and juvenile stages and are not expressed during later stages (***Table S3***), including Sobic.001G239000, Sobic.005G018900 (PLC2; Phospholipase C2), Sobic.005G051600 (DUF247), Sobic.005G188500 (DUF94), Sobic.006G017500 (Disease resistance protein family/CC-NBS-LRR class), and Sobic.007G099966 (Andropogoneae specific family). Other TSSE genes also have high stem specificity ($\tau = 1$) during vegetative and juvenile stages but have low to moderate expression in non-stem tissues during late development, thus their tissue specificity is more temporal. Many of the genes in this latter group have known stem-related functions related to early regulation of vascular tissue development and maintenance as listed in ***Table 2***. Two of these genes, YAB1 and CLAVATA3, are involved in CLAVATA (CLV)-WUSCHEL (WUS) pathway to regulate stem cell maintenance in shoot and floral meristems (Caño-Delgado, Lee and Demura, 2010; Ruonala, Ko and Helariutta, 2017). CLVA3 is expressed in the central domain of SAM and influences the relative positioning of the organ primordia by acting together with CLV1 and WUS (Ottoline Leyser and Furner, 1992), while YAB1 defines the primordia domain of the SAM periphery and promotes robust partitioning of the SAM by non-autonomously communicating with the meristem to regulate the expression of CLV3 and WUS. This coordination is essential for organized growth of the SAM (Goldshmidt *et al.*, 2008a; Stahle *et al.*, 2009a).

There were 21 TSSL genes that, like the TSSE genes, are stem-preferred temporally during late development but have no, low, or moderate expression in non-stem tissues (leaf, root) during early development (***Table S3***). Two of these TSSL genes that have high stem specificity during floral,

anthesis, and grain maturity stages, but are expressed in non-stem tissues during early development, include Sobic.002G299200 (AGO1) and Sobic.006G166300 (WRKY12). AGO1 is required for the function of most miRNAs, including miR165/166, to repress the HD-ZIP TFs such as PHB, PHV, REV related to the developmental regulators required for SAM establishment and vascular development in *Arabidopsis* (Kidner and Martienssen, 2005; Mallory *et al.*, 2009; Zhang and Zhang, 2012). WRKY12 was previously shown to negatively regulate the expression of NST2, which acts as a master regulator in switching on and off secondary cell wall biosynthesis and represses cell wall biosynthetic genes in pith cells in *Arabidopsis*, *Medicago sativa L.* (Sanchez, Nehlin and Greb, 2012; Gallego-Giraldo *et al.*, 2016; Yang and Wang, 2016; Yang *et al.*, 2016). Genes involved in secondary cell wall formation were down-regulated in Della post anthesis, possibly mediates in part through the action of WRKY12 (McKinley et al., 2016).

To elucidate the enriched functional categories of the identified stem-preferred genes, GO term analysis was performed using PlantTFDB for AOSS, TSSE and TSSL gene sets separately (***Figure 3***). AOSS and TSSE genes were enriched in biological process-related GO terms for biological regulation and regulation of transcription, while TSSL genes were enriched for plasma membrane-related cellular component GO terms. Based on functional annotations, there were 10 stem-preferred TFs from TALE, HD-ZIP, B3, YABBY, WRKY, AP2, bHLH protein families (***Table S3***). Also, there are 15 Andropogoneae family-specific stem-preferred genes (homologs are only found in *Sorghum bicolor, Saccharum spontaneum, Miscanthus sinesis,* and *Zea mays*), in which 10 of them are "orphan" genes without detectable homologues in other lineages for any previously annotated gene among eukaryotes, and are unique to a species and/or phylogenetic clade. Previous reports have shown that orphan genes generally have lower expression levels but higher tissue-specific expression compared to phylogenetically conserved genes (Levine *et al.*, 2006; Donoghue *et al.*, 2011; Wu, Irwin and Zhang, 2011). Several hypotheses have been proposed to explain the regulation of these younger genes that lead to tissue-specific and/or stage-specific expression, including transposon insertion upstream of the transcription start-site, shared regulatory elements with older genes (possibly by gene overlap), association with a bidirectional promoter, or by being located within an intron (Maksakova and Mager, 2005; Toll-Riera *et al.*, 2008; Arendsee, Li and Wurtele, 2014; Sundaram and Wysocka, 2020). We explored the cis-regulatory elements in the putative promoter regions of these genes below.

## Cis-regulatory element enrichment of AOSS, TSS-Early, TSS-Late stem-preferred genes

The three groups of stem genes (AOSS, TSSE, TSSL) were further explored in an attempt to uncover common and unique CREs that regulate these groups. Many CREs were common among promoters of genes of interest (GOI), including the orphan genes, across the three categories of stem-expressed genes (AOSS, TSS-Early, TSS-Late) (***Figure 4***), indicating that these genes are co-regulated by TFs that bind to the enriched CREs. These common CREs included binding sites for TF known to play roles in stem-specific functions such as bZIP, AT-Hook, HD-ZIP, GATA, MYB, SBP, TCP, WOX, which are previously reported to be present in vascular-specific genes (***Table S1***) (YH and SR, 2003; Ruiz-Medrano *et al.*, 2011; Le Hir and Bellini, 2013; Zhang *et al.*, 2014). Interestingly, the AP2 CRE was uniquely enriched in the AOSS group of genes, while the CPP CRE was strongly enriched in the TSSL group of genes compared to the AOSS and TSSE groups (***Figure 4***). This result suggests that TFs binding to the AP2 and CPP CREs might have an important role in the specific regulation of genes in the AOSS and TSSL groups.

## Gene regulatory networks of AOSS, TSS-Early, TSS-Late stem-preferred genes

Differential gene expression over development is controlled by temporal changes in TF binding to promoter elements. It is not only the timing of a TF's expression, but also its temporal promoter occupancy that determines time evolved gene expression (Para *et al.*, 2014). However, extensive patterns of TF occupancy in developmental contexts are not known for sorghum (i.e. ChiP-Seq or DAP-seq). Therefore, elucidating the dynamic properties of cis-regulatory networks from available spatial-temporal datasets is useful in understanding how dynamic expression states arise. To explore which TFs regulate the expression of stem-specific genes, and examine whether stem-expressed genes are differentially regulated over time, gene regulatory networks (GRNs) were constructed for each group of stem-expressed genes at each developmental stage. GRNs were constructed by layering TF binding predictions onto gene co-expression networks. We generated 15-temporal stem-preferred gene regulatory networks (5 temporal GRNs for AOSS, TSS-Early, TSS-Late). See ***Data S1*** for list of stem-preferred genes and TFs and unique directed TF–target gene interactions for each of 15 GRNs.

GRN analysis of AOSS, TSS-Early, TSS-Late genes along with their co-expressed TFs revealed that the connectivity of each network has a positive correlation with the corresponding out-node degree, and that the out-node degree '*x*' in each of the networks follows a power-law distribution

y=ax^b with an average exponent *b* of 0.182±0.064, 0.169±0.064, 0.23±0.07 for AOSS, TSS-Early, TSS-Late, respectively (***Table 3***). The connectivity of each network has a positive correlation with the corresponding out-node degree (AOSS: 0.661±0.205, TSSE: 0.653±0.127, TSSL: 0.629±0.229). This suggests that the GRNs have the properties of robust biological networks, which tend to contain scale-free properties such as having few nodes with a very large number of regulatory edges, or "hubs." These hub genes are absent in random networks, for which the degree of all nodes is close to the average degree according to network theory (Cohen, Havlin and ben-Avraham, 2002; Barabási and Oltvai, 2004). Thus, the hubs in our GRNs potentially co-regulate many of the stem-expressed genes (***Figure 5***). In the stem GRNs, each TF is associated with an average of 4 genes during the early stages and 15 genes during the late stages (***Table 3***). Similarly, multiple TFs can regulate the expression of a single gene, and in the stem GRNs, each target gene is associated with an average of five regulatory TFs during the early stages compared to 18 regulatory TFs in the late stages. These results demonstrate that stem-expressed genes become increasingly co-regulated over development.

**TF Hubs enriched for known stem genes:** We evaluated the hubbiness of the TFs in our GRNs to identify TFs that exert a large amount of control over the entire network. Hubs are defined here as highly connected TF genes (top 25% h-index) (See ***Table S4*** for list of TF hubs in each network). The most frequent TF families present in the networks are bHLH, bZIP, ERF, C2H2, G2-like, GATA, HD-ZIP, NAC, MYB, MYB-related, TALE, Trihelix, and WRKY. TFs previously reported as important regulatory genes in stem tissue are present in the networks, including known stem TF hubs such as PHB, ATHB15, APL, VND7, KNAT1, MYB42, and MYB52 (McConnell *et al.*, 2001; Ye, 2002; Emery *et al.*, 2003; Kubo *et al.*, 2005; Floyd, Zalewski and Bowman, 2006; Hoon Jung and Mo Park, 2007; Zhong *et al.*, 2008; Zhou, Sebastian and Lee, 2011; De Rybel *et al.*, 2016; Hennet *et al.*, 2020). Radial patterning of the vascular bundles is established by TF hubs from class III HD-ZIPs, PHB, ATHB15 (McConnell *et al.*, 2001; Emery *et al.*, 2003; Floyd, Zalewski and Bowman, 2006). NACs are reported to be TF hubs that can induce xylem cell differentiation (Kubo *et al.*, 2005; Zhong *et al.*, 2008). For example, VND7 (NAC TF Family) initiates metaxylem vessel differentiation and induces hierarchical gene expression network by inducing other TFs such as MYBs, which in turn function as secondary and tertiary regulators of genes related to secondary cell wall thickening during xylem cell

differentiation (Ye, 2002; Hoon Jung and Mo Park, 2007; Zhou, Sebastian and Lee, 2011; De Rybel *et al.*, 2016).

Next, we compared hubbiness of TFs over development to identify significant TF genes that change their hub property status relevant to early and late developmental phases. Our expectation was there might be different set of critical TFs that drive stem-specific gene expression during early and late development. There were 22 and 22 TFs that acted as early TF hubs (act as a TF hub only during early stages) for AOSS and TSSE genes, respectively. Likewise, there were 19 and 21 late TF hubs (act as a TF hub only during late stages) for AOSS and TSSL genes, respectively (***Table S5***). We examined whether any of these early and late TF hubs are unique to AOSS vs. TSSE/TSSL genes. However, the results showed all early and late TF hubs are critical TF hubs for AOSS, TSSE, TSSL genes during early and late stages, respectively. Likewise, genes that remain a functionally conserved TF hub in all GRNs during early and late phases also tend to be TF hubs for all AOSS, TSSE, TSSL GRNs. This could mean stem-specific and transiently stem-specific genes are likely co-regulated by the same higher-level transcriptional program during respective early and late developmental phases.

The stem-preferred spatial-temporal GRNs exhibit different transcriptional regulatory programs that activate the stem-preferred genes over time (***Table 3***, ***Figure 5***), revealing active TFs that temporally drive stem-specific expression during early and late development. To further explore this possibility, we compared early (juvenile, vegetative) regulators to late (floral, anthesis, grain) regulators of stem-preferred genes to identify stage-specific TFs. There were six TFs predicted to regulate at least one AOSS or TSSE gene during the early stages (SP-Early regulators: AIB/Sobic.003G272200 (regulatory edges to 1 AOSS and 3 TSSE), Sobic.004G283300/Integrase-type DNA binding superfamily protein (regulatory edges to 3 TSSE), PDF2/ Sobic.007G065300 (regulatory edges to 3 AOSS and 2 TSSE), bHLH/Sobic.007G189400 (regulatory edges to 1 AOSS and 4 TSSE), MYB61/Sobic.009G036500 (regulatory edges to 2 TSSE), and WRKY33/Sobic.009G100500 (regulatory edges to 2 TSSE)). Only one TF was identified that uniquely regulates the expression of stem-specific genes during the late developmental stages in the AOSS and TSSL gene groups (SP-Late regulator: RSM2/ Sobic.009G235500 (regulatory edges to 6 AOSS and 4 TSSL)).

Next, we tested whether there are unique TFs belonging to different CREs that turn on AOSS genes vs. TSSE or TSSL genes. Therefore, we tested whether there are: a) unique TF regulators specific to AOSS genes during early stages; b) unique TF regulators specific to AOSS genes during late stages; c) unique TF regulators specific to TSSE genes during early stages; and d) unique TF regulators specific to TSSL during late stages. There were no TFs unique to the AOSS and TSSL networks over time. However, three TFs, MYB61, WRKY33, and Sobic.004G283300, are predicted to uniquely regulate TSSE genes during juvenile and vegetative networks.

**Comparison of gene regulatory networks of stem-preferred genes between BTx623 and Della cultivars**

To examine if there is conserved regulation of stem-specific gene expression among grain and sweet sorghum varieties we compared the temporal stem-specific GRNs between *cv.* BTx623 (grain) and *cv.* Della (sweet) using transcriptome profiles from McKinley et al. (2018) and McCormick et al. (2018). Comparative analysis revealed little consensus among regulatory edges (average consensus of 1.8% for AOSS, 1.8% for TSSE, 1.3% for TSSL) between BTx623 and Della GRNs at comparable developmental stages (***Figure 6, Figure 7, Figure 8***). Thus, GRN analysis suggests that there are different transcriptional regulatory programs regulating the expression of stem-specific genes in these two cultivars. While the majority of regulatory edges directly connected to the stem-specific genes are variable between the two cultivars, it is interesting to note that whole network analysis revealed that the same TF hubs are conserved, with an average consensus of 38.9%, 39.5%, and 46.0% in the AOSS, TSSE, and TSSL, respectively.

To further investigate the variation in regulatory edges in the stem GRNs between cultivars, we explored the genetic sequences of stem expressed GOIs. Various events can alter the expression of a gene, such as insertions, deletions, and point mutations in either the gene of interest or the transcription factors (TFs) that regulate that gene. Regulatory DNA variants adjacent to or within tissue-specific genes or their corresponding transcriptional regulators may influence gene expression patterns in different cultivars. While many mechanisms can regulate these genes differently (i.e. involvement of various cell- and species-specific non-coding cellular factors such as post-transcriptional regulators and post-translational modifications), it is reasonable to examine the contribution of cis-regulatory element variants for divergent regulation. This was elucidated by estimating genetic variation between BTx623 and Della *cvs*, for insertions, deletions, and point

mutations. Several single nucleotide variations (SNVs) and/or structural variations (SVs) between BTx623 and Della were found in the gene body, 1.5 kbp upstream, and/or 500 bp downstream regions for a number of stem-expressed genes in the AOSS, TSSE, and TSSL gene groups. The list of SNVs and SVs found in each of these groups is in ***File S1*** and ***File S2***, respectively. For example, within the AOSS group of genes, six out of ten genes had at least one SNV and/or SVs in the gene body, 1.5 kbp upstream, or 500 bp downstream region between BTx623 and Della. Of these genes, four contained these mutations within a CRE (***Table 4*** and ***Figure S3***) in their putative promoter regions. For example, both Sorghum homologous of KNAT1 (Sobic.001G106000 and Sobic.001G106200) had multiple CREs with SNVs, including *NAC, TBP, SBP, AT-Hook, GATA, ARR-B, bZIP, C2H2, MADS-Box* (***Table 4***).

**Genetic variation in hierarchical regulators may contribute to the differential activation of stem-specific genes among Sorghum cultivars**

In cases where stem-expressed GOIs have different GRNs but no genetic variations between the cultivars, we explored if genetic alterations exist in the genes encoding TFs predicted to regulate the stem-specific GOIs. Again, several single nucleotide variations (SNVs) and/or structural variations (SVs) between BTx623 and Della were found in the gene body, 1.5 kbp upstream, and/or 500 bp downstream regions for a number of TFs predicted to regulate stem-expressed genes in the AOSS, TSSE, and TSSL gene groups (***File S1*** and ***File S2***). For example, 271 out of 381 total TFs across the AOSS GRNs had at least one SNV and/or SV within the gene region under investigation. Of those, 167 TF genes (136 are considered network hubs) had genetic alterations inside a CRE within their putative promoter region. Key TF hubs with variations include PHB, ATHB15, KNAT1, GATA4, ABF4, AGL7, HB-7, TGA4, VIP1, MYB48, FMA, GBF3, and LHY1. TF hubs with genetic variations likely contribute to the observed global differences in stem GRNs between cultivars. In a specific example, Sobic.006G026300 is one of most stem-specific and highly expressed genes in Sorghum *cv.* BTx623. ***Figure 10*** shows TFs predicted to regulate Sobic.006G026300 over the five developmental stages. The network prediction for Sobic.006G026300 in *cv BTx623 and cv Della* showed different regulation patterns (***Figure 9***). The genomic region of this gene is highly conserved with no observed mutations within the gene body or up- or down-stream non-coding regulatory regions (***Figure S2***). One explanation could be genetic variations present in the genes encoding the TF regulators between

391 cultivars. ***Table 5*** contains list of SNVs and SVs for some of the predicted TF regulators of
392 Sobic.006G026300, including NAC1 (Sobic.008G164800), a regulator during the vegetative –
393 grain stages (***Figure 10***), that has several significant structural variation, namely TBP (INS), bZIP
394 (INS/RPL), AT-Hook (INS/RPL). Likewise, the TF bHLH93 (Sobic.001G513700), a regulator
395 during the juvenile-vegetative stages (***Figure 10***), has structural variations within the RAV (DEL),
396 MADS-Box (DEL), G2-like (DEL), ERF (INS), and C2H2 (INS) CREs, and the TF ABF4
397 (Sobic.002G225100), also a regulator during the juvenile stage, has a 24 bp and a 29 bp deletion
398 next to each other flanking WRKY, MYB, MYB-related CREs, and 21 bp insertions within the
399 bZIP and bHLH CREs in its putative promoter region. Because changes in stem-specific gene
400 expression between species could also arise by diversification of master-regulators, altering hub
401 expression can cause large-scale network changes, potentially driving the different tissue-specific
402 gene regulatory networks we observed.

403

404 **CONCLUSIONS**

405 Using *in-silico* analysis of spatial-temporal transcriptome datasets, we report a set of "stem-
406 specific" genes that are always expressed in the stem throughout development, and "stem-
407 preferred" genes that are transiently stem-specific over development. This includes genes known
408 to be involved in vascular tissue development and/or function such as KNAT1, HB7, MYC2,
409 CLV3, SUVR1, WRKY12, AGO1, PLC2 (Hoon Jung and Mo Park, 2007; Martínez-Navarro *et*
410 *al.*, 2013; De Rybel *et al.*, 2016; Hennet *et al.*, 2020). The identification of common cis-regulatory
411 elements within the promoters of stem-specific and stem-preferred genes suggest co-regulation,
412 which was further explored using predictive GRNs. Network analysis revealed that GRN topology
413 changes over time, indicating that different transcriptional regulatory programs are responsible for
414 the temporal activation of the stem-expressed genes. However, TF hubs appear to be conserved
415 across the temporal stem GRNs and include known regulators of stem genes as well as TFs not
416 previously implicated in stem development/function that merit further investigation.

417 The analysis of temporal stem GRNs between grain and sweet sorghum varieties revealed striking
418 differences in the regulatory landscape, which is not entirely surprising given the considerable
419 phenotypic differences between grain and sweet sorghums. The sorghum BTx623 cultivar is an
420 early maturing grain sorghum genotype with short stature, which is phenotypically distinct from

the tall, late-maturing Della cultivar (sweet sorghum genotype), typically grown for stem sugars or high biomass yield (Jiang *et al.*, 2013; Cooper *et al.*, 2019). BTx623 encodes recessive alleles of the dwarfing gene *Dw1* (Hilley *et al.*, 2016) that encodes a protein involved in brassinosteroid signaling (Hirano *et al.*, 2017) and *Dw3*, a gene that encodes an ABCB1 auxin efflux transporter (Multani *et al.*, 2003; Knöller *et al.*, 2010), whereas Della is dominant for both dwarfing loci. The difference in dwarfing genotype results in differences in internode lengths and the ratio of node/internode tissues in BTx623 and Della. Grain sorghums develop treachery elements during xylem formation earlier than sweet sorghums, and sweet sorghums accumulate more stem sugars compared to grain sorghum (Jiang *et al.*, 2013; Cooper *et al.*, 2019). Grain sorghums usually contain less carbohydrate in the upper internodes at grain maturity compared with sweet sorghum varieties, due to mobilization of carbon from stems during grain filling. The divergence of stem related genes between the two types has previously been suggested from genetic variation analysis (Jiang *et al.*, 2013; Cooper *et al.*, 2019), and the present study suggests genomic divergence in stem-expressed genes based on their developmental transcriptomes and regulatory programs.

Genomic perturbations such as single nucleotide variants and large structural variants can lead to an altered function or expression of TFs involved in stem-specific functions that lead to downstream changes in TF activity (e.g., HD-ZIP, MYB, NAC, TALE). Significant nucleotide sequence variation in a canonical TF binding motif in a promoter or enhancer region affects the prediction of TF-target edges in a regulatory network. Such perturbations in CRE sites within putative promoter regions of stem-expressed genes resulted in different GRNs for *BTx623* and *Della* cultivars. However, further studies are needed to confirm "real" TF-target interactions through experiments such as yeast-one hybrid (Y1H) (Arda and Walhout, 2009; Fuxman Bass, Reece-Hoyes and Walhout, 2016) or DAP-seq (O'Malley *et al.*, 2016). Nonetheless, the current study provides a starting point for more in-depth experiments to assess the extent to which these promoter elements are tissue-specific through targeted mutagenesis, for example, using tissue-specific genome modification (Schürholz *et al.*, 2018; Decaestecker *et al.*, 2019; Shi *et al.*, 2019; Ali, Mahfouz and Mansoor, 2020). This will greatly benefit efforts to fine-tune manipulation of tissue-specific and developmental-stage specific gene expression of transgenes required to genetically engineer sorghum stems to accumulate oil and elevate its status as a biofuel crop.

## Methods

### RNA-Seq data set representing different tissues over 5 development time points

We used a publicly available *Sorghum bicolor* transcriptomic dataset containing expression profiles of 38 tissue samples (McCormick *et al.*, 2018b) representing four plant organs (stem, leaf, root, and reproductive structures) over five developmental stages (juvenile, vegetative, floral initiation, anthesis, and grain maturity). This dataset was used to obtain a tissue-specific expression atlas (***Table 1***).

The quality cleaned raw reads of the 38 RNA-seq samples, *Sorghum bicolor (cv BTx623)* reference genome (S. bicolor v3.1.1), and genome annotations were downloaded from Phytozome (phytozome.jgi.doe.gov). The reads were then mapped to the reference genome using the STAR/2.6.0c-aligner (Dobin *et al.*, 2013). Read numbers that mapped to each gene were counted by featureCounts v.1.4.5-pl under the subread/1.5.2 package (Liao, Smyth and Shi, 2014). The normalized gene expression levels were estimated as in transcripts-per-million (TPM) using modules from edgeR package (Robinson, McCarthy and Smyth, 2010) in R (R Development Core Team, 2019). In a given tissue sample at a given developmental stage, genes having more than five counts per million (CPM) were considered as expressed. To visualize close grouping of spatial-temporal samples, hierarchical clustering and principal component analysis (PCA) of the log-transformed and quantile normalized TPM counts was performed and visualized using 'FactoMineR' R package (***Figure 1***). This revealed that roots, stems, and leaves formed distinct clusters based on gene expression, while stem samples formed a cluster with the peduncle and panicle (***Figure 1*).** Grass peduncles are the specialized stem internodes formed after floral initiation that support the panicle, consistent with co-clustering with other stem tissues. Peduncle tissue and the first panicle tissue sample was collected from BTx623 approximately 7 days after floral initiation when the peduncle/panicle was in an early stage of development. During early stages of inflorescence development, the rachis, the stem tissue of an inflorescence, constitutes a large portion of harvested panicle tissue. This helps explain why stem, peduncle and panicle tissues form a cluster.

### Identification of stem-preferred genes

We computed a tissue-specificity index to explicitly identify stem-specific genes expressed constitutively across all development stages or only during a specific developmental stage using tissue-specific RNA-seq gene expression profiles from McCormick et al. 2018. In this study, Tau Index (τ) (Kryuchkova-Mostacci and Robinson-Rechavi, 2016) was used to identify tissue-specific and broadly expressed genes. The index τ is defined as:

$$\tau = \frac{\sum_{i=1}^{N}(1 - x_i)}{N - 1}$$

Where, $N$ is the number of tissue types, and $x_i$ is the expression profile component for a given gene in a given tissue, normalized by the maximal expression value of the given gene from all tissue samples considered. A custom R script adapted from (Kryuchkova-Mostacci and Robinson-Rechavi, 2016) was used to perform Tau index analysis (https://github.com/Madara-Dilhani/Tissue_Specificity). The τ values can vary from 0-1; genes with a τ value of 1 being absolute tissue-specific genes that are only expressed in a single tissue, and genes with a τ value of 0 are expressed in all tissues. Generally, tissue-specificity measures produce a bimodal distribution enabling identification of these groups, a) 0-0.2 tissue specificity (non/low specificity, housekeeping genes), b) 0.2-0.8 tissue specificity (intermediate specificity), c) 0.8-1.0 tissue specificity (high/absolute specificity). For example, "Housekeeping Genes" are often expressed in many (or all) tissues with no enriched expression for a particular tissue, thus the reported Sorghum housekeeping genes PP2A (Sobic.004G092500, Sobic.010G174800, Sobic.001G492300, Sobic.001G049400, Sobic.001G251700) and EIF4-α (Sobic.004G039400, Sobic.004G356900, Sobic.006G121100, Sobic.010G251100, Sobic.001G224400, Sobic.002G335600, Sobic.003G442400) had τ values < 0.2 (Sudhakar Reddy *et al.*, 2016; Nguyen, Eamens and Grof, 2018). A few previously reported maize stem-preferred gene homologs in sorghum were identified as highly stem specific (τ = 1); Glycosyl Hydrolase (Sobic.002G273500), AP2 (Sobic.003G203800, Sobic.008G149400), KNAT (Sobic.001G106000, Sobic.001G106200, Sobic.001G075101), RNI (Sobic.001G075101, Sobic.006G022500, Sobic.006G088820), WRKY (Sobic.004G298400, Sobic.006G166300), SUV (Sobic.004G218000), CLAVATA (Sobic.004G207000), BRX (Sobic.006G203200), CDC2 (Sobic.006G202100) (Hoopes *et al.*, 2019).

The mean gene expression (quantile normalized log-TPM) across replicates within an organ at a single developmental stage was determined. The τ index was calculated per each expressed gene at each developmental stage. The τ index is associated with its respective tissue with maximum expression. In this analysis, a gene was considered tissue-specific if it had a $\tau = 1$. Following τ analysis, three groups were identified: 1) genes that are always tissue-specific throughout development, now called Always On Stem Specific (AOSS), 2) genes that are stem-specific during early developmental stages (juvenile and vegetative), now called Temporally Stem Specific Early (TSS-Early), and 3) genes that are tissue-specific during late developmental stages (floral initiation – grain maturity), now called Temporally Stem Specific Late (TSS-Late). The $\log_2$-TPM values of 3 categories of stem-specific genes were visualized in heatmaps generated using ggplot2 R package (***Figure 2***).

**Enrichment of biological functions and cis-regulatory elements**

We identified sets of enriched gene ontology terms (GO-terms), and cis-regulatory elements (CREs) for each stem-specific gene set separately (AOSS, TSS-Early, TSS-Late). The GO enrichment tool (p-value $p \leq 0.01$) at PlantTFDB v5.0 (http://planttfdb.gao-lab.org) was used to identify enriched GO terms (***Figure 3***). Furthermore, the promoters of tissue-specific gene sets were analyzed to identify enriched CREs. The PlantPan 4.0 TF binding site tool (http://plantpan.itps.ncku.edu.tw/index.html) was used to map a position weight matrix (PWM) of known plant TF binding motifs to promoters of all genes in the *Sorghum bicolor (cv BTx623)* genome. The putative promoter sequences of all sorghum genes were limited to 1500 bp upstream from −1 position from the start codon of the gene. The enrichment of CREs in the promoter regions of tissue-specific genes were calculated based on frequency of occurrence of CREs in the promoter regions compared to occurrence in all promoters in the genome (fold change ≥ 1.5, adjusted FDR ≤ 0.05). Here, CREs that were present in both (+) and (-) strand were considered. The enriched CREs in each stem-specific gene set (AOSS, TSS-Early, TSS-Late) is visualized in ***Figure 4***.

**Stem-preferred gene regulatory network (GRN) construction**

We constructed five gene regulatory networks, in which pairwise correlation analysis (Pearson correlation ≥ |0.7| and Mutual Rank ≥ |0.7|, p-value ≤ 0.05) was performed for all stem-preferred genes (AOSS, TSS-Early, TSS-Late) and stem-expressed TFs accounting for CREs enriched for

each stem-preferred gene and stem-expressed TFs (Obayashi and Kinoshita, no date). Gene-gene edges are defined by Pearson Correlation Coefficients (PCC) over 0.7 with a Mutual Rank (MR) > 0.7. The MR is another co-expression measure which takes a geometric average of the PCC rank from gene A to gene B and that of gene B to gene A. The use of MR measure makes sure only highly significant co-expressed gene pairs are considered. TF-target associations were predicted using known plant TF binding motifs mapped to 1.5 kbp upstream from the start of each gene, in which only enriched CREs were considered. The final GRN is a merged network that contain TF-to-gene connections, which have highly correlated expression. Networks were constructed using a custom R script (https://github.com/Madara-Dilhani/GeneReg_Network) and visualized using Cytoscape (version 3.8). Network analysis was performed in NetworkAnalyzer tool in Cytoscape to assess network properties.

**Variation in stem-preferred GRN of Grain Sorghum and Sweet Sorghum genotypes**

We used RNA-seq data of *Sorghum bicolor cv Della* stem tissue for 11 different time points covering five development stages from McKinley et al. (2018) to evaluate conserved gene regulation of stem-preferred genes between sweet and grain sorghum varieties. McKinley et al. (2018) had sampled *Sorghum bicolor cv Della* stem tissue samples by taking the 10th internode of the stem at the following time points: Days after emergence (DAE), days before Anthesis (DBA), and days after anthesis (DAA). At the juvenile stage: 14 DAE/-55 DBA; at the vegetative stage: 29 DAE/-40 DBA, 31 DAE/-38 DBA, 40 DAE/-29 DBA; at floral initiation: 53 DAE/ -16 DBA, 62 DAE/-7 DBA; at anthesis: 69 DAE/Anthesis; at grain maturity: 80 DAE/+11 DAA, 94 DAE/+25 DAA; at post-grain maturity: 102 DAE/+43 DAA, 137 DAE/+68 DAA. We used publicly available raw counts for all the samples obtained from Phytozome to normalize gene expression levels using modules from edgeR package (Robinson, McCarthy and Smyth, 2010) in R (R Core Team, 2019). We constructed eleven gene regulatory networks (GRNs) for 11 time points sampled representing 5 developmental stages. The co-expression analysis (Pearson correlation $\geq |0.7|$ and Mutual Rank $\geq |0.7|$, p-value $\leq 0.05$) was performed for all stem-preferred genes (AOSS, TSS-Early, TSS-Late) and expressed TFs accounting for CREs enriched for each stem-preferred gene and stem-expressed TFs (Obayashi & Kinoshita, 2009). Gene regulatory networks between BTx623 and Della were compared for comparable time points, which were selected based on time of sample collection in their respective developmental maturities and

anatomical similarity (McKinley et al., 2018). The comparable time points were as follows: a) Juvenile - BTx623-8 DAE with Della-14 DAE, b) Vegetative - BTx623-24 DAE with Della-29 DAE, c) Anthesis - BTx623-65 DAE with Della-69 DAE, d) Grain Maturity - BTx623-96 DAE with Della-94 DAE.

**Single nucleotide and structural variations in regulatory regions of stem-preferred genes and their regulatory TFs between *Sorghum bicolor cv BTx623* and *Sorghum bicolor cv Della***

We collected Della re-sequenced genome and variant reports (https://genome.jgi.doe.gov/portal/Sorbicsequencing_9/Sorbicsequencing_9.info.html), where authors used the following variant calling pipeline: a) The Della re-sequenced reads were mapped to the BTx623 reference genome using BWA (Li and Durbin, 2009), b) putative single nucleotide point (SNP) mutation and small indel sites were identified by using samtools (Li *et al.*, 2009), bcftools (Danecek, Schiffels and Durbin, 2016), vcfutils (Danecek *et al.*, 2011), picard (*Picard Tools - By Broad Institute*, no date), snpEff, (Cingolani *et al.*, 2012), c) putative large structural variant sites such as deletions, insertions, inversions, intra-chromosomal translocations, inter-chromosomal translocations using following tools; BreakDancer (Fan *et al.*, 2014), Pindel (Ye *et al.*, 2009). For this study, we predicted significant point mutations and structural variants within 1.5 kb upstream promoter region, gene body, and 500 bp downstream region of tissue-specific genes and their regulators in respective developmental stages from BTx623 gene regulatory network to identify variable and conserved regulatory regions.

**Data availability statement:** There is no data associated with this manuscript. R script for tissue-specificity analysis is available at https://github.com/Madara-Dilhani/Tissue_Specificity. Other scripts used for network analysis are available at https://github.com/Madara-Dilhani/GeneReg_Network.

**Acknowledgements:**

*This work was funded by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research*

595 *under Award Number DE-SC0018420) and in part by the DOE Great Lakes Bioenergy Research*
596 *Center (DOE BER Office of Science DE-FC02-07ER64494). Any opinions, findings, and*
597 *conclusions or recommendations expressed in this publication are those of the author(s) and do*
598 *not necessarily reflect the views of the U.S. Department of Energy.*
599
600 **Author contributions:** MHA and AMC conceived the project; MHA, AMC, and JEM designed
601 the experiments; JEM contributed the data; MHA performed all data analysis; MHA, AMC, and
602 JEM wrote the manuscript. All authors reviewed and approved of the manuscript.
603
604 **Conflicts of interest statement:** The authors of this manuscript claim no conflicts of interest.
605
606 **Supporting Information**

607 **Figure S1:** Expression dynamics of highly expressed most stem-specific AOSS genes in Sorghum
608 cv BTx623 and cv Della.

609 **Figure S2:** The view upstream non-coding region, gene body, downstream non-coding region of
610 3 highly expressed most stem-specific AOSS genes from Integrated Genomic Viewer (IGV).

611 **Figure S3:** The view upstream non-coding region, gene body, downstream non-coding region of
612 AOSS genes with significant genetic alterations from Integrated Genomic Viewer (IGV).

613 **Figure S4:** The Integrated Genomic Viewer (IGV) view upstream non-coding region, gene body,
614 downstream non-coding region of regulatory TFs of Sobic.006G024800 during a) Early stage and
615 b) Late stage.

616 **Table S1:** Previously known xylem and phloem specific cis-regulatory element in different
617 vascular specific promoters

618 **Table S2:** Statistics of tissue-specific genes across tissues during development

619 **Table S3:** Tissue-specificity Index by stages for AOSS, TSS-Early, TSS-Late genes

620 **Table S4:** Transcription factors which are TF hubs in at least one of the 15 GRNs

**Table S5:** Transcription factors which showed change of TF hub characteristics in early and late developmental stages

**File S1:** The list of predicted single nucleotide variations (SNVs) between BTx623 and Della found in the gene body, 1.5 kbp upstream (including predicted CREs) for stem-expressed genes in the AOSS, TSSE, and TSSL gene group and their TF regulators.

**File S2:** The list of predicted structural variations (SVs) between BTx623 and Della found in the gene body, 1.5 kbp upstream (including predicted CREs) for stem-expressed genes in the AOSS, TSSE, and TSSL gene group and their TF regulators.

**Data S1:** TF–target gene interactions for each of 15 GRNs generated for this study is presented in .sif files.

## References

Ali, Z., Mahfouz, M. M. and Mansoor, S. (2020) 'CRISPR-TSKO: A Tool for Tissue-Specific Genome Editing in Plants', *Trends in Plant Science*. Elsevier Ltd, pp. 123–126. doi: 10.1016/j.tplants.2019.12.002.

Arda, H. E. and Walhout, A. J. M. (2009) 'Gene-centered regulatory networks'. doi: 10.1093/bfgp/elp049.

Arendsee, Z. W., Li, L. and Wurtele, E. S. (2014) 'Coming of age: Orphan genes in plants', *Trends in Plant Science*. doi: 10.1016/j.tplants.2014.07.003.

Barabási, A.-L. and Oltvai, Z. N. (2004) 'Network biology: understanding the cell's functional organization', *Nature Reviews Genetics*, 5(2), pp. 101–113. doi: 10.1038/nrg1272.

Baud, S. (2018) 'Seeds as oil factories', *Plant Reproduction*. Springer Verlag, pp. 213–235. doi: 10.1007/s00497-018-0325-6.

Beechey-Gradwell, Z. *et al.* (2020) 'Storing carbon in leaf lipid sinks enhances perennial ryegrass carbon capture especially under high N and elevated CO 2', *Journal of Experimental Botany*, 71(7), pp. 2351–2361. doi: 10.1093/jxb/erz494.

Bensussan, M. *et al.* (2015) 'Suppression of dwarf and Irregular Xylem phenotypes generates low-acetylated biomass lines in arabidopsis', *Plant Physiology*. American Society of Plant Biologists, 168(2), pp. 452–463. doi: 10.1104/pp.15.00122.

Byrne, M. E., Simorowski, J. and Martienssen, R. A. (2002) 'ASYMMETRIC LEAVES1 reveals know gene redundancy in Arabidopsis', *Development*, pp. 1957–1965.

Byrt, C. S., Grof, C. P. L. and Furbank, R. T. (2011) 'C4 Plants as Biofuel Feedstocks: Optimising Biomass Production and Feedstock Quality from a Lignocellulosic Perspective', *Journal of Integrative Plant Biology*. J Integr Plant Biol, pp. 120–135. doi: 10.1111/j.1744-7909.2010.01023.x.

Caño-Delgado, A., Lee, J.-Y. and Demura, T. (2010) 'Regulatory Mechanisms for Specification and Patterning of Plant Vascular Tissues', *Annual Review of Cell and Developmental Biology*. Annual Reviews, 26(1), pp. 605–637. doi: 10.1146/annurev-cellbio-100109-104107.

Cao, X. *et al.* (2007) 'Abscisic acid and stress signals induce Viviparous1 expression in seed and vegetative tissues of maize', *Plant Physiology*. Plant Physiol, 143(2), pp. 720–731. doi: 10.1104/pp.106.091454.

Carlsson, A. S. *et al.* (2011) 'Replacing fossil oil with fresh oil - with what and for what?', *European Journal of Lipid Science and Technology*. Wiley-Blackwell, pp. 812–831. doi: 10.1002/ejlt.201100032.

Chen, L. *et al.* (2015) 'The characterization of GmTIP, a root-specific gene from soybean, and the expression analysis of its promoter', *Plant Cell, Tissue and Organ Culture*. Kluwer Academic Publishers, 121(2), pp. 259–274. doi: 10.1007/s11240-014-0682-2.

Cingolani, P. *et al.* (2012) 'A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3', *Fly*. Taylor and Francis Inc., 6(2), pp. 80–92. doi: 10.4161/fly.19695.

Cohen, R., Havlin, S. and ben-Avraham, D. (2002) 'Structural properties of scale-free networks', *Handbook of Graphs and Networks*. (Wiley Online Books), pp. 85–110. doi: doi:10.1002/3527602755.ch4.

Cooper, E. A. *et al.* (2019) 'A new reference genome for Sorghum bicolor reveals high levels of

676 sequence similarity between sweet and grain genotypes: Implications for the genetics of sugar
677 metabolism', *BMC Genomics*. BioMed Central Ltd., 20(1), p. 420. doi: 10.1186/s12864-019-
678 5734-x.

679 da Costa, R. M. F. *et al.* (2014) 'Genotype, development and tissue-derived variation of cell-wall
680 properties in the lignocellulosic energy crop Miscanthus', *Annals of Botany*, 114(6), pp. 1265–
681 1277. doi: 10.1093/aob/mcu054.

682 Danecek, P. *et al.* (2011) 'The variant call format and VCFtools', *BIOINFORMATICS*
683 *APPLICATIONS NOTE*, 27(15), pp. 2156–2158. doi: 10.1093/bioinformatics/btr330.

684 Danecek, P., Schiffels, S. and Durbin, R. (2016) *Multiallelic calling model in bcftools (-m)*.

685 Decaestecker, W. *et al.* (2019) 'CRISPR-Tsko: A technique for efficient mutagenesis in specific
686 cell types, tissues, or organs in Arabidopsis[open]', *Plant Cell*. American Society of Plant
687 Biologists, 31(12), pp. 2868–2887. doi: 10.1105/tpc.19.00454.

688 Deikman, J., Kline, R. and Fischer, R. L. (1992) 'Organization of ripening and ethylene
689 regulatory regions in a fruit-specific promoter from tomato (Lycopersicon esculentum)', *Plant*
690 *Physiology*. American Society of Plant Biologists, 100(4), pp. 2013–2017. doi:
691 10.1104/pp.100.4.2013.

692 Demirbas, A. (2007) 'Progress and recent trends in biofuels', *Progress in Energy and*
693 *Combustion Science*. Pergamon, 33(1), pp. 1–18. doi: 10.1016/j.pecs.2006.06.001.

694 Demirbas, T. and Demirbas, A. H. (2010) 'Bioenergy, Green Energy. Biomass and Biofuels',
695 *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 32(12), pp. 1067–
696 1075. doi: 10.1080/15567030903058600.

697 Dobin, A. *et al.* (2013) 'STAR: ultrafast universal RNA-seq aligner', *Bioinformatics*. Narnia,
698 29(1), pp. 15–21. doi: 10.1093/bioinformatics/bts635.

699 Donoghue, M. T. *et al.* (2011) 'Evolutionary origins of Brassicaceae specific genes in
700 Arabidopsis thaliana', *BMC Evolutionary Biology*, 11(1). doi: 10.1186/1471-2148-11-47.

701 Douglas, S. J. *et al.* (2002) 'KNAT1 and ERECTA regulate inflorescence architecture in
702 Arabidopsis', *Plant Cell*, 14(3), pp. 547–558. doi: 10.1105/tpc.010391.

703 Durrett, T. P., Benning, C. and Ohlrogge, J. (2008) 'Plant triacylglycerols as feedstocks for the
704 production of biofuels', *Plant Journal*. Plant J, pp. 593–607. doi: 10.1111/j.1365-
705 313X.2008.03442.x.

706 Dutt, M. *et al.* (2014) 'Temporal and spatial control of gene expression in horticultural crops',
707 *Horticulture Research*. Nature Publishing Group, pp. 1–17. doi: 10.1038/hortres.2014.47.

708 Dyer, J. M. and Mullen, R. T. (2005) 'Development and potential of genetically engineered
709 oilseeds', *Seed Science Research*. Cambridge University Press (CUP), 15(4), pp. 255–267. doi:
710 10.1079/ssr2005216.

711 Emery, J. F. *et al.* (2003) 'Radial Patterning of Arabidopsis Shoots by Class III HD-ZIP and
712 KANADI Genes', *Current Biology*. Cell Press, 13(20), pp. 1768–1774. doi:
713 10.1016/j.cub.2003.09.035.

714 Fan, J., Yan, C. and Xu, C. (2013) 'Phospholipid:diacylglycerol acyltransferase-mediated
715 triacylglycerol biosynthesis is crucial for protection against fatty acid-induced cell death in
716 growing tissues of Arabidopsis', *Plant Journal*. Plant J, 76(6), pp. 930–942. doi:
717 10.1111/tpj.12343.

718 Fan, X. *et al.* (2014) 'BreakDancer: Identification of genomic structural variation from paired-
719 end read mapping', *Current Protocols in Bioinformatics*. John Wiley and Sons Inc.,
720 2014(SUPPL.45). doi: 10.1002/0471250953.bi1506s45.

721 Feltus, F. A. and Vandenbrink, J. P. (2012) 'Bioenergy grass feedstock: Current options and
722 prospects for trait improvement using emerging genetic, genomic, and systems biology toolkits',
723 *Biotechnology for Biofuels*. Biotechnol Biofuels. doi: 10.1186/1754-6834-5-80.

724 Floyd, S. K., Zalewski, C. S. and Bowman, J. L. (2006) 'Evolution of class III homeodomain-
725 leucine zipper genes in streptophytes', *Genetics*, 173(1), pp. 373–388. doi:
726 10.1534/genetics.105.054239.

727 Fuxman Bass, J. I., Reece-Hoyes, J. S. and Walhout, A. J. M. (2016) 'Gene-centered yeast one-
728 hybrid assays', *Cold Spring Harbor Protocols*. Cold Spring Harbor Laboratory Press, 2016(12),
729 pp. 1039–1043. doi: 10.1101/pdb.top077669.

730 Gallego-Giraldo, L. *et al.* (2016) 'Combining enhanced biomass density with reduced lignin

731 level for improved forage quality', *Plant Biotechnology Journal*. Blackwell Publishing Ltd,
732 14(3), pp. 895–904. doi: 10.1111/pbi.12439.

733 Gao, J. *et al.* (2015) 'Evolution, diversification, and expression of KNOX proteins in plants',
734 *Frontiers in Plant Science*. Frontiers Media S.A., 6(OCTOBER). doi: 10.3389/fpls.2015.00882.

735 Goldshmidt, A. *et al.* (2008a) 'Signals derived from YABBY gene activities in organ primordia
736 regulate growth and partitioning of Arabidopsis shoot apical meristems', *Plant Cell*, 20(5), pp.
737 1217–1230. doi: 10.1105/tpc.107.057877.

738 Goldshmidt, A. *et al.* (2008b) 'Signals derived from YABBY gene activities in organ primordia
739 regulate growth and partitioning of Arabidopsis shoot apical meristems', *Plant Cell*. American
740 Society of Plant Biologists, 20(5), pp. 1217–1230. doi: 10.1105/tpc.107.057877.

741 Graham, I. A. (2008) 'Seed Storage Oil Mobilization', *Annual Review of Plant Biology*. Annual
742 Reviews, 59(1), pp. 115–142. doi: 10.1146/annurev.arplant.59.032607.092938.

743 Gujas, B. and Rodriguez-Villalon, A. (2016) 'Plant Phosphoglycerolipids: The Gatekeepers of
744 Vascular Cell Differentiation', *Frontiers in Plant Science*. Frontiers Media S.A., 7(FEB2016), p.
745 103. doi: 10.3389/fpls.2016.00103.

746 Harwood, J. L. *et al.* (1971) 'On the origin of oil droplets in maturing castor bean seeds, Ricinus
747 communis', *Lipids*. Springer-Verlag, 6(11), pp. 851–854. doi: 10.1007/BF02531217.

748 Heaton, E. A., Dohlemnn, F. G. and Long, S. P. (2008) 'Meeting US biofuel goals with less land:
749 the potential of Miscanthus', *Global Change Biology*. John Wiley & Sons, Ltd, 14(9), pp. 2000–
750 2014. doi: 10.1111/j.1365-2486.2008.01662.x.

751 Helizon, H. *et al.* (2014a) 'Class I KNOX transcription factors promote differentiation of
752 cambial derivatives into xylem fibers in the Arabidopsis hypocotyls', *Development (Cambridge)*.
753 Company of Biologists Ltd, 141(22), pp. 4311–4319. doi: 10.1242/dev.111369.

754 Helizon, H. *et al.* (2014b) 'Class I KNOX transcription factors promote differentiation of
755 cambial derivatives into xylem fibers in the Arabidopsis hypocotyls', *Development (Cambridge)*.
756 Company of Biologists Ltd, 141(22), pp. 4311–4319. doi: 10.1242/dev.111369.

757 Hennet, L. *et al.* (2020) 'Transcriptional Regulation of Sorghum Stem Composition: Key Players

Identified Through Co-expression Gene Network and Comparative Genomics Analyses', *Frontiers in Plant Science*. Frontiers Media S.A., 11, p. 224. doi: 10.3389/fpls.2020.00224.

Hill, J. *et al.* (2006) 'Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels', *Proceedings of the National Academy of Sciences of the United States of America*, 103(30), pp. 11206–11210. doi: 10.1073/pnas.0604600103.

Hilley, J. *et al.* (2016) 'Identification of Dw1, a regulator of sorghum stem internode length', *PLoS ONE*. Public Library of Science, 11(3), p. e0151271. doi: 10.1371/journal.pone.0151271.

Le Hir, R. and Bellini, C. (2013) 'The plant-specific Dof transcription factors family: New players involved in vascular system development and functioning in Arabidopsis', *Frontiers in Plant Science*. Frontiers Research Foundation. doi: 10.3389/fpls.2013.00164.

Hirakawa, Y. *et al.* (2008) *Non-cell-autonomous control of vascular stem cell fate by a CLE peptide/receptor system*. Available at: www.pnas.org/cgi/content/full/ (Accessed: 12 May 2020).

Hirano, K. *et al.* (2017) 'Sorghum DW1 positively regulates brassinosteroid signaling by inhibiting the nuclear localization of BRASSINOSTEROID INSENSITIVE 2', *Scientific Reports*. Nature Publishing Group, 7(1). doi: 10.1038/s41598-017-00096-w.

Hoffmann-Thoma, G. *et al.* (1996) 'Sucrose accumulation in sweet sorghum stem internodes in relation to growth', *Physiologia Plantarum*, 97(2), pp. 277–284. doi: 10.1034/j.1399-3054.1996.970210.x.

Hofvander, P. *et al.* (2016) 'Potato tuber expression of Arabidopsis WRINKLED1 increase triacylglycerol and membrane lipids while affecting central carbohydrate metabolism', *Plant Biotechnology Journal*. Blackwell Publishing Ltd, 14(9), pp. 1883–1898. doi: 10.1111/pbi.12550.

Hoon Jung, J. and Mo Park, C. (2007) *Vascular Development in Plants: Specification of Xylem and Phloem Tissues*, *Joumal of Plant Biology*.

Hoopes, G. M. *et al.* (2019) 'An updated gene atlas for maize reveals organ-specific and stress-induced genes', *The Plant Journal*. Blackwell Publishing Ltd, 97(6), pp. 1154–1167. doi: 10.1111/tpj.14184.

785 Huang, H. *et al.* (2016) 'Technoeconomic Analysis of Biodiesel and Ethanol Production from
786 Lipid-Producing Sugarcane and Sweet Sorghum', *Industrial Biotechnology*. Mary Ann Liebert
787 Inc., 12(6), pp. 357–365. doi: 10.1089/ind.2016.0013.

788 Huang, H., Long, S. P. and Singh, V. (2015) 'Ultra oil-producing'.

789 Huminiecki, L., Lloyd, A. T. and Wolfe, K. H. (2003) 'Congruence of tissue expression profiles
790 from Gene Expression Atlas, SAGEmap and TissueInfo databases', *BMC Genomics*. BioMed
791 Central, 4(1), p. 31. doi: 10.1186/1471-2164-4-31.

792 Ito, Y. *et al.* (2006) 'Dodeca-CLE as peptides as suppressors of plant stem cell differentiation',
793 *Science*. American Association for the Advancement of Science, 313(5788), pp. 842–845. doi:
794 10.1126/science.1128436.

795 Jang, G. *et al.* (2017) 'Antagonistic interaction between jasmonic acid and cytokinin in xylem
796 development', *Scientific Reports*. Nature Publishing Group, 7(1), pp. 1–13. doi: 10.1038/s41598-
797 017-10634-1.

798 Jiang, S. Y. *et al.* (2013) 'Genetic variation and expression diversity between grain and sweet
799 sorghum lines', *BMC Genomics*. BioMed Central, 14(1), p. 18. doi: 10.1186/1471-2164-14-18.

800 Julien, P. *et al.* (2012) 'Mechanisms and Evolutionary Patterns of Mammalian and Avian Dosage
801 Compensation', *PLoS Biology*. Edited by N. H. Barton. Public Library of Science, 10(5), p.
802 e1001328. doi: 10.1371/journal.pbio.1001328.

803 Kanehara, K. *et al.* (2015) 'Arabidopsis AtPLC2 Is a Primary Phosphoinositide-Specific
804 Phospholipase C in Phosphoinositide Metabolism and the Endoplasmic Reticulum Stress
805 Response', *PLoS Genetics*. Public Library of Science, 11(9). doi: 10.1371/journal.pgen.1005511.

806 Karp, A. and Shield, I. (2008) 'Bioenergy from plants and the sustainable yield challenge', *New*
807 *Phytologist*. John Wiley & Sons, Ltd, 179(1), pp. 15–32. doi: 10.1111/j.1469-
808 8137.2008.02432.x.

809 Kebrom, T. H., McKinley, B. and Mullet, J. E. (2017) 'Dynamics of gene expression during
810 development and expansion of vegetative stem internodes of bioenergy sorghum', *Biotechnology*
811 *for Biofuels*. BioMed Central Ltd., 10(1), p. 159. doi: 10.1186/s13068-017-0848-3.

812 Kehr, J. and Kragler, F. (2018) 'Long distance RNA movement', *New Phytologist*, 218(1), pp.
813 29–40. doi: 10.1111/nph.15025.

814 Kelly, A. A. *et al.* (2013) 'Suppression of the *SUGAR-DEPENDENT1* triacylglycerol lipase
815 family during seed development enhances oil yield in oilseed rape ( *Brassica napus* L.)', *Plant*
816 *Biotechnology Journal*. John Wiley & Sons, Ltd, 11(3), pp. 355–361. doi: 10.1111/pbi.12021.

817 Kidner, C. A. and Martienssen, R. A. (2005) 'The role of ARGONAUTE1 (AGO1) in meristem
818 formation and identity', *Developmental Biology*. Academic Press Inc., 280(2), pp. 504–517. doi:
819 10.1016/j.ydbio.2005.01.031.

820 Kim, H. U. *et al.* (2015) 'Senescence-inducible LEC2 enhances triacylglycerol accumulation in
821 leaves without negatively affecting plant growth', *Plant Biotechnology Journal*. Blackwell
822 Publishing Ltd, 13(9), pp. 1346–1359. doi: 10.1111/pbi.12354.

823 Knöller, A. S. *et al.* (2010) 'Brachytic2/ZmABCB1 functions in IAA export from intercalary
824 meristems', *Journal of Experimental Botany*, 61(13), pp. 3689–3696. doi: 10.1093/jxb/erq180.

825 Koçar, G. and Civaş, N. (2013) 'An overview of biofuels from energy crops: Current status and
826 future prospects', *Renewable and Sustainable Energy Reviews*. Pergamon, pp. 900–916. doi:
827 10.1016/j.rser.2013.08.022.

828 Kondo, Y. *et al.* (2011) 'CLE Peptides can Negatively Regulate Protoxylem Vessel Formation
829 via Cytokinin Signaling', *Plant and Cell Physiology*. Oxford University Press, 52(1), p. 37. doi:
830 10.1093/PCP/PCQ129.

831 Kryuchkova-Mostacci, N. and Robinson-Rechavi, M. (2016) 'A benchmark of gene expression
832 tissue-specificity metrics', *Briefings in Bioinformatics*. Narnia, 18(2), p. bbw008. doi:
833 10.1093/bib/bbw008.

834 Kubo, M. *et al.* (2005) 'Transcription switches for protoxylem and metaxylem vessel formation',
835 *Genes and Development*, 19(16), pp. 1855–1860. doi: 10.1101/gad.1331305.

836 Kumar, D., Long, S. P. and Singh, V. (2018) 'Biorefinery for combined production of jet fuel
837 and ethanol from lipid-producing sugarcane: a techno-economic evaluation', *GCB Bioenergy*.
838 Blackwell Publishing Ltd, 10(2), pp. 92–107. doi: 10.1111/gcbb.12478.

839 Kumar, N. (2020) 'Biotechnology for Biofuels: A Sustainable Green Energy Solution'. Springer.

840 Levine, M. T. *et al.* (2006) 'Novel genes derived from noncoding DNA in Drosophila
841 melanogaster are frequently X-linked and exhibit testis-biased expression', *Proceedings of the*
842 *National Academy of Sciences of the United States of America*, 103(26), pp. 9935–9939. doi:
843 10.1073/pnas.0509809103.

844 Li, H. *et al.* (2009) 'The Sequence Alignment/Map format and SAMtools', *BIOINFORMATICS*
845 *APPLICATIONS NOTE*, 25(16), pp. 2078–2079. doi: 10.1093/bioinformatics/btp352.

846 Li, H. and Durbin, R. (2009) 'Fast and accurate short read alignment with Burrows-Wheeler
847 transform', 25(14), pp. 1754–1760. doi: 10.1093/bioinformatics/btp324.

848 Li, N. *et al.* (2016) 'Fatty Acid and Lipid Transport in Plant Cells', *Trends in Plant Science*.
849 Elsevier Ltd, pp. 145–158. doi: 10.1016/j.tplants.2015.10.011.

850 Li, Y. *et al.* (2013) 'Isolation and characterization of two novel root-specific promoters in rice
851 (Oryza sativa L.)', *Plant Science*. Plant Sci, 207, pp. 37–44. doi: 10.1016/j.plantsci.2013.02.002.

852 Li, Y. *et al.* (2019) 'Genome-scale mining of root-preferential genes from maize and
853 characterization of their promoter activity', *BMC Plant Biology*. BioMed Central Ltd., 19(1), p.
854 584. doi: 10.1186/s12870-019-2198-8.

855 Liao, Y., Smyth, G. K. and Shi, W. (2014) 'featureCounts: an efficient general purpose program
856 for assigning sequence reads to genomic features', *Bioinformatics*. Narnia, 30(7), pp. 923–930.
857 doi: 10.1093/bioinformatics/btt656.

858 Lima, M. de F. *et al.* (2017) 'Molecular mechanisms of biomass increase in plants',
859 *Biotechnology Research and Innovation*. Elsevier BV, 1(1), pp. 14–25. doi:
860 10.1016/j.biori.2017.08.001.

861 Liu, Q. *et al.* (2017) 'Genetic enhancement of oil content in potato tuber (Solanum tuberosum L.)
862 through an integrated metabolic engineering strategy', *Plant Biotechnology Journal*. Blackwell
863 Publishing Ltd, 15(1), pp. 56–67. doi: 10.1111/pbi.12590.

864 Lung, S.-C. and Weselake, R. J. (2006) 'Diacylglycerol acyltransferase: A key mediator of plant
865 triacylglycerol synthesis', *Lipids*. John Wiley & Sons, Ltd, 41(12), pp. 1073–1088. doi:

866 10.1007/s11745-006-5057-y.

867 Maksakova, I. A. and Mager, D. L. (2005) 'Transcriptional Regulation of Early Transposon
868 Elements, an Active Family of Mouse Long Terminal Repeat Retrotransposons', *Journal of*
869 *Virology*. American Society for Microbiology, 79(22), pp. 13865–13874. doi:
870 10.1128/jvi.79.22.13865-13874.2005.

871 Mallory, A. C. *et al.* (2009) 'Redundant and Specific Roles of the ARGONAUTE Proteins
872 AGO1 and ZLL in Development and Small RNA-Directed Gene Silencing', *PLoS Genetics*.
873 Edited by T. Kakutani. Public Library of Science, 5(9), p. e1000646. doi:
874 10.1371/journal.pgen.1000646.

875 Martínez-Navarro, A. C. *et al.* (2013) 'Vascular gene expression: a hypothesis', *Frontiers in*
876 *Plant Science*. Frontiers Research Foundation, 4(JUL), p. 261. doi: 10.3389/fpls.2013.00261.

877 Mathur, S. *et al.* (2017) 'Sweet sorghum as biofuel feedstock: Recent advances and available
878 resources', *Biotechnology for Biofuels*. BioMed Central Ltd. doi: 10.1186/s13068-017-0834-9.

879 McConnell, J. R. *et al.* (2001) 'Role of PHABULOSA and PHAVOLUTA in determining radial
880 patterning in shoots', *Nature*, 411(6838), pp. 709–713. doi: 10.1038/35079635.

881 McCormick, R. F. *et al.* (2018a) 'The Sorghum bicolor reference genome: improved assembly,
882 gene annotations, a transcriptome atlas, and signatures of genome organization', *Plant Journal*.
883 Blackwell Publishing Ltd, 93(2), pp. 338–354. doi: 10.1111/tpj.13781.

884 McCormick, R. F. *et al.* (2018b) 'The *Sorghum bicolor* reference genome: improved assembly,
885 gene annotations, a transcriptome atlas, and signatures of genome organization', *The Plant*
886 *Journal*. Wiley/Blackwell (10.1111), 93(2), pp. 338–354. doi: 10.1111/tpj.13781.

887 McKinley, B. A. *et al.* (2018) 'Developmental dynamics of stem starch accumulation in
888 *Sorghum bicolor*', *Plant Direct*. John Wiley and Sons Inc., 2(8), p. e00074. doi:
889 10.1002/pld3.74.

890 McLaughlin, S. B. *et al.* (2002) 'High-value renewable energy from prairie grasses',
891 *Environmental Science and Technology*. American Chemical Society , 36(10), pp. 2122–2129.
892 doi: 10.1021/es010963d.

McLaughlin, S. B. and Kszos, L. A. (2005) 'Development of switchgrass (Panicum virgatum) as a bioenergy feedstock in the United States', *Biomass and Bioenergy*. Elsevier Ltd, 28(6), pp. 515–535. doi: 10.1016/j.biombioe.2004.05.006.

Mitchell, M. C. *et al.* (2020) 'Increasing growth and yield by altering carbon metabolism in a transgenic leaf oil crop', *Plant Biotechnology Journal*. Blackwell Publishing Ltd. doi: 10.1111/pbi.13363.

Miyashima, S. *et al.* (2013) 'Stem cell function during plant vascular development', *EMBO Journal*. European Molecular Biology Organization, pp. 178–193. doi: 10.1038/emboj.2012.301.

Molla, K. A. *et al.* (2013) 'Rice oxalate oxidase gene driven by green tissue-specific promoter increases tolerance to sheath blight pathogen (Rhizoctonia solani) in transgenic rice', *Molecular Plant Pathology*. Mol Plant Pathol, 14(9), pp. 910–922. doi: 10.1111/mpp.12055.

Moyle, R. L. and Birch, R. G. (2013) 'Sugarcane Loading Stem Gene promoters drive transgene expression preferentially in the stem', *Plant Molecular Biology*. Plant Mol Biol, 82(1–2), pp. 51–58. doi: 10.1007/s11103-013-0034-3.

Mudge, S. R. *et al.* (2013) 'Mature-stem expression of a silencing-resistant sucrose isomerase gene drives isomaltulose accumulation to high levels in sugarcane', *Plant Biotechnology Journal*. Plant Biotechnol J, 11(4), pp. 502–509. doi: 10.1111/pbi.12038.

Mullet, J. *et al.* (2014) 'Energy Sorghum-a genetic model for the design of C 4 grass bioenergy crops', *Journal of Experimental Botany*, 65(13), pp. 3479–3489. doi: 10.1093/jxb/eru229.

Multani, D. S. *et al.* (2003) 'Loss of an MDR transporter in compact stalks of maize br2 and sorghum dw3 mutants', *Science*. Science, 302(5642), pp. 81–84. doi: 10.1126/science.1086072.

Murray, S. C. *et al.* (2008) 'Genetic Improvement of Sorghum as a Biofuel Feedstock: I. QTL for Stem Sugar and Grain Nonstructural Carbohydrates', *Crop Science*. John Wiley & Sons, Ltd, 48(6), pp. 2165–2179. doi: 10.2135/cropsci2008.01.0016.

Napier, J. A. *et al.* (2014a) 'Understanding and manipulating plant lipid composition: Metabolic engineering leads the way', *Current Opinion in Plant Biology*. Elsevier Ltd, pp. 68–75. doi: 10.1016/j.pbi.2014.04.001.

Napier, J. A. *et al.* (2014b) 'Understanding and manipulating plant lipid composition: Metabolic engineering leads the way', *Current Opinion in Plant Biology*. Elsevier Ltd, pp. 68–75. doi: 10.1016/j.pbi.2014.04.001.

Nguyen, D. Q., Eamens, A. L. and Grof, C. P. L. (2018) 'Reference gene identification for reliable normalisation of quantitative RT-PCR data in Setaria viridis', *Plant Methods*. BioMed Central Ltd., 14(1). doi: 10.1186/s13007-018-0293-8.

O'Hagan, S. *et al.* (2018) 'GeneGini: Assessment via the Gini Coefficient of Reference "Housekeeping" Genes and Diverse Human Transporter Expression Profiles', *Cell Systems*. Cell Press, 6(2), pp. 230-244.e1. doi: 10.1016/J.CELS.2018.01.003.

O'Malley, R. C. *et al.* (2016) 'Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape', *Cell*. Cell Press, 165(5), pp. 1280–1292. doi: 10.1016/j.cell.2016.04.038.

Obayashi, T. and Kinoshita, K. (no date) 'Rank of Correlation Coefficient as a Comparable Measure for Biological Significance of Gene Coexpression'. doi: 10.1093/dnares/dsp016.

Olson, S. N. *et al.* (2012) 'High biomass yield energy sorghum: developing a genetic model for C4 grass bioenergy crops', *Biofuels, Bioproducts and Biorefining*. John Wiley & Sons, Ltd, 6(6), pp. 640–655. doi: 10.1002/bbb.1357.

Ottoline Leyser, H. M. and Furner, I. J. (1992) 'Characterisation of three shoot apical meristem mutants of Arabidopsis thaliana', *Development*, 116(2), pp. 397 LP – 403. Available at: http://dev.biologists.org/content/116/2/397.abstract.

P. Joshi, C. and Nookaraju, A. (2012) 'New Avenues of Bioenergy Production from Plants: Green Alternatives to Petroleum', *Journal of Petroleum & Environmental Biotechnology*. OMICS Publishing Group, 03(07). doi: 10.4172/2157-7463.1000134.

Papini-Terzi, F. S. *et al.* (2009) 'Sugarcane genes associated with sucrose content', *BMC Genomics*. BioMed Central, 10(1), p. 120. doi: 10.1186/1471-2164-10-120.

Para, A. *et al.* (2014) 'Hit-and-run transcriptional control by bZIP1 mediates rapid nutrient signaling in Arabidopsis', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 111(28), pp. 10371–10376. doi: 10.1073/pnas.1404657111.

948 Parajuli, S. *et al.* (2020) 'Towards oilcane: Engineering hyperaccumulation of triacylglycerol
949 into sugarcane stems', *GCB Bioenergy*. Wiley. doi: 10.1111/gcbb.12684.

950 Pauly, M. and Ramírez, V. (2018) 'New insights into wall polysaccharide o-acetylation',
951 *Frontiers in Plant Science*. Frontiers Media S.A. doi: 10.3389/fpls.2018.01210.

952 *Picard Tools - By Broad Institute* (no date). Available at: http://broadinstitute.github.io/picard/
953 (Accessed: 7 May 2020).

954 Qazi, H. A. *et al.* (2014) 'Alterations in stem sugar content and metabolism in sorghum
955 genotypes subjected to drought stress', *Functional Plant Biology*. CSIRO PUBLISHING, 41(9),
956 p. 954. doi: 10.1071/FP13299.

957 Qazi, H. A., Paranjpe, S. and Bhargava, S. (2012) 'Stem sugar accumulation in sweet sorghum -
958 Activity and expression of sucrose metabolizing enzymes and sucrose transporters', *Journal of*
959 *Plant Physiology*. J Plant Physiol, 169(6), pp. 605–613. doi: 10.1016/j.jplph.2012.01.005.

960 Rae, A. L. *et al.* (2009) 'Functional specialization of vacuoles in sugarcane leaf and stem',
961 *Tropical Plant Biology*. Springer, 2(1), pp. 13–22. doi: 10.1007/s12042-008-9019-9.

962 Ramachandran, P., Carlsbecker, A. and Etchells, J. P. (2016) 'Class III HD-ZIPs govern vascular
963 cell fate: an HD view on patterning and differentiation', *Journal of Experimental Botany*, 68(1),
964 pp. 55–69. doi: 10.1093/jxb/erw370.

965 Ramšak, Ž. *et al.* (2018) 'Network modeling unravels mechanisms of crosstalk between ethylene
966 and salicylate signaling in potato', *Plant Physiology*. American Society of Plant Biologists,
967 178(1), pp. 488–499. doi: 10.1104/pp.18.00450.

968 Robinson, M. D., McCarthy, D. J. and Smyth, G. K. (2010) 'edgeR: a Bioconductor package for
969 differential expression analysis of digital gene expression data.', *Bioinformatics (Oxford,*
970 *England)*. Oxford University Press, 26(1), pp. 139–40. doi: 10.1093/bioinformatics/btp616.

971 Rooney, W. L. *et al.* (2007) 'Designing sorghum as a dedicated bioenergy feedstock', *Biofuels,*
972 *Bioproducts and Biorefining*. John Wiley & Sons, Ltd, 1(2), pp. 147–157. doi: 10.1002/bbb.15.

973 Ruiz-Medrano, R. *et al.* (2011) 'Vascular expression in Arabidopsis is predicted by the
974 frequency of CT/GA-rich repeats in gene promoters', *The Plant Journal*. John Wiley & Sons,

975 Ltd, 67(1), pp. 130–144. doi: 10.1111/j.1365-313X.2011.04581.x.

976 Ruonala, R., Ko, D. and Helariutta, Y. (2017) 'Genetic Networks in Plant Vascular
977 Development', *Annual Review of Genetics*. Annual Reviews, 51(1), pp. 335–359. doi:
978 10.1146/annurev-genet-120116-024525.

979 Ruuska, S. A. *et al.* (2006) 'Genotypic variation in water-soluble carbohydrate accumulation in
980 wheat', *Functional Plant Biology*. CSIRO PUBLISHING, 33(9), p. 799. doi: 10.1071/FP06062.

981 Růžička, K. *et al.* (2015) 'Xylem development - from the cradle to the grave', *New Phytologist*.
982 Blackwell Publishing Ltd, 207(3), pp. 519–535. doi: 10.1111/nph.13383.

983 De Rybel, B. *et al.* (2016) 'Plant vascular development: From early specification to
984 differentiation', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, pp. 30–40.
985 doi: 10.1038/nrm.2015.6.

986 Sae-Lim, P. *et al.* (2019) 'Unraveling vascular development-related genes in laticifer-containing
987 tissue of rubber tree by high-throughput transcriptome sequencing', *Current Plant Biology*.
988 Elsevier B.V., 19, p. 100112. doi: 10.1016/j.cpb.2019.100112.

989 Sage, R. F. and Zhu, X.-G. (2011) 'Exploiting the engine of C 4 photosynthesis', *Journal of*
990 *Experimental Botany*, 62(9), pp. 2989–3000. doi: 10.1093/jxb/err179.

991 Sanchez, P., Nehlin, L. and Greb, T. (2012) 'From thin to thick: Major transitions during stem
992 development', *Trends in Plant Science*. Elsevier Current Trends, pp. 113–121. doi:
993 10.1016/j.tplants.2011.11.004.

994 Sanjaya *et al.* (2011) 'Increasing the energy density of vegetative tissues by diverting carbon
995 from starch to oil biosynthesis in transgenic Arabidopsis', *Plant Biotechnology Journal*, 9(8), pp.
996 874–883. doi: 10.1111/j.1467-7652.2011.00599.x.

997 Schnyder, H. (1993) 'The role of carbohydrate storage and redistribution in the source-sink
998 relations of wheat and barley during grain filling — a review', *New Phytologist*, 123(2), pp. 233–
999 245. doi: 10.1111/j.1469-8137.1993.tb03731.x.

1000 Schug, J. *et al.* (2005) 'Promoter features related to tissue specificity as measured by Shannon
1001 entropy', *Genome Biology*. BioMed Central, 6(4), p. R33. doi: 10.1186/gb-2005-6-4-r33.

1002 Schürholz, A. K. *et al.* (2018) 'A comprehensive toolkit for inducible, cell type-specific gene
1003 expression in Arabidopsis', *Plant Physiology*. American Society of Plant Biologists, 178(1), pp.
1004 40–53. doi: 10.1104/pp.18.00463.

1005 Scofield, S., Dewitte, W. and Murray, J. A. H. (2007) 'The KNOX gene SHOOT
1006 MERISTEMLESS is required for the development of reproductive meristematic tissues in
1007 Arabidopsis', *The Plant Journal*. John Wiley & Sons, Ltd, 50(5), pp. 767–781. doi:
1008 10.1111/j.1365-313X.2007.03095.x.

1009 Sekhon, R. S. *et al.* (2016) 'Stover Composition in Maize and Sorghum Reveals Remarkable
1010 Genetic Variation and Plasticity for Carbohydrate Accumulation', *Frontiers in Plant Science*.
1011 Frontiers Research Foundation, 7(June2016), p. 822. doi: 10.3389/fpls.2016.00822.

1012 Shi, D. *et al.* (2019) 'STEM CELLS AND REGENERATION Bifacial cambium stem cells
1013 generate xylem and phloem during radial plant growth'. doi: 10.1242/dev.171355.

1014 Shih, P. M. (2018) 'Towards a sustainable bio-based economy: Redirecting primary metabolism
1015 to new products with plant synthetic biology', *Plant Science*. Elsevier Ireland Ltd, 273, pp. 84–
1016 91. doi: 10.1016/j.plantsci.2018.03.012.

1017 Slewinski, T. L. (2012) 'Non-structural carbohydrate partitioning in grass stems: a target to
1018 increase yield stability, stress tolerance, and biofuel production', *Journal of Experimental*
1019 *Botany*, 63(13), pp. 4647–4670. doi: 10.1093/jxb/ers124.

1020 Slocombe, S. P. *et al.* (2009) 'Oil accumulation in leaves directed by modification of fatty acid
1021 breakdown and lipid synthesis pathways', *Plant Biotechnology Journal*, 7(7), pp. 694–703. doi:
1022 10.1111/j.1467-7652.2009.00435.x.

1023 Smith, A. D. *et al.* (2006) 'DNA motifs in human and mouse proximal promoters predict tissue-
1024 specific expression', *Proceedings of the National Academy of Sciences*. National Academy of
1025 Sciences, 103(16), pp. 6275–6280. doi: 10.1073/PNAS.0508169103.

1026 Stahle, M. I. *et al.* (2009a) 'YABBYs and the transcriptional corepressors LEUNIG and
1027 LEUNIG-HOMOLOG maintain leaf polarity and meristem activity in Arabidopsis', *Plant Cell*.
1028 American Society of Plant Biologists, 21(10), pp. 3105–3118. doi: 10.1105/tpc.109.070458.

1029 Stahle, M. I. *et al.* (2009b) 'YABBYs and the transcriptional corepressors LEUNIG and

LEUNIG-HOMOLOG maintain leaf polarity and meristem activity in Arabidopsis', *Plant Cell*. American Society of Plant Biologists, 21(10), pp. 3105–3118. doi: 10.1105/tpc.109.070458.

Sudhakar Reddy, P. *et al.* (2016) 'Evaluation of Sorghum [Sorghum bicolor (L.)] reference genes in various tissues and under abiotic stress conditions for quantitative real-time PCR data normalization', *Frontiers in Plant Science*. doi: 10.3389/fpls.2016.00529.

Sundaram, V. and Wysocka, J. (2020) 'Transposable elements as a potent source of diverse cis-regulatory sequences in mammalian genomes', *Philosophical Transactions of the Royal Society B: Biological Sciences*. Royal Society Publishing. doi: 10.1098/rstb.2019.0347.

Tarpley, L. and Vietor, D. M. (2007) 'Compartmentation of sucrose during radial transfer in mature sorghum culm', *BMC Plant Biology*. BioMed Central, 7(1), p. 33. doi: 10.1186/1471-2229-7-33.

Thelen, J. J. and Ohlrogge, J. B. (2002) 'Metabolic Engineering of Fatty Acid Biosynthesis in Plants', *Metabolic Engineering*. Academic Press Inc., 4(1), pp. 12–21. doi: 10.1006/mben.2001.0204.

Toll-Riera, M. *et al.* (2008) 'Origin of Primate Orphan Genes: A Comparative Genomics Approach'. doi: 10.1093/molbev/msn281.

Trindade, L. M. *et al.* (2003) 'Isolation and functional characterization of a stolon specific promoter from potato (Solanum tuberosum L.)', *Gene*. Elsevier, 303(1–2), pp. 77–87. doi: 10.1016/S0378-1119(02)01147-2.

Uhlen, M. *et al.* (2015) 'Tissue-based map of the human proteome', *Science*. American Association for the Advancement of Science, 347(6220), pp. 1260419–1260419. doi: 10.1126/science.1260419.

Vanhercke, T. *et al.* (2013) 'Synergistic effect of WRI1 and DGAT1 coexpression on triacylglycerol biosynthesis in plants', *FEBS Letters*. FEBS Lett, 587(4), pp. 364–369. doi: 10.1016/j.febslet.2012.12.018.

Vanhercke, T. *et al.* (2014) 'Metabolic engineering of biomass for high energy density: Oilseed-like triacylglycerol yields from plant leaves', *Plant Biotechnology Journal*, 12(2), pp. 231–239. doi: 10.1111/pbi.12131.

1058 Vanhercke, T. *et al.* (2019) 'Up-regulation of lipid biosynthesis increases the oil content in
1059 leaves of *Sorghum bicolor*', *Plant Biotechnology Journal*. Blackwell Publishing Ltd, 17(1), pp.
1060 220–232. doi: 10.1111/pbi.12959.

1061 Varvel, G. E. *et al.* (2008) 'Comparison of Corn and Switchgrass on Marginal Soils for
1062 Comparison of Corn and Switchgrass on Marginal Soils for Bioenergy Bioenergy'. doi:
1063 10.1016/j.biombioe.2007.07.003.

1064 Venglat, S. P. *et al.* (2002) 'The homeobox gene brevipedicellus is a key regulator of
1065 inflorescence architecture in Arabidopsis', *Proceedings of the National Academy of Sciences of*
1066 *the United States of America*, 99(7), pp. 4730–4735. doi: 10.1073/pnas.072626099.

1067 Waclawovsky, A. J. *et al.* (2010a) 'Sugarcane for bioenergy production: an assessment of yield
1068 and regulation of sucrose content', *Plant Biotechnology Journal*. John Wiley & Sons, Ltd, 8(3),
1069 pp. 263–276. doi: 10.1111/j.1467-7652.2009.00491.x.

1070 Waclawovsky, A. J. *et al.* (2010b) 'Sugarcane for bioenergy production: An assessment of yield
1071 and regulation of sucrose content', *Plant Biotechnology Journal*. Plant Biotechnol J, pp. 263–
1072 276. doi: 10.1111/j.1467-7652.2009.00491.x.

1073 Walsh, K. B., Sky, R. C. and Brown, S. M. (2005) 'The anatomy of the pathway of sucrose
1074 unloading within the sugarcane stalk', *Functional Plant Biology*. CSIRO PUBLISHING, 32(4),
1075 p. 367. doi: 10.1071/FP04102.

1076 Wang, H. *et al.* (2007) 'Developmental control of Arabidopsis seed oil biosynthesis', *Planta*.
1077 Planta, 226(3), pp. 773–783. doi: 10.1007/s00425-007-0524-0.

1078 Wang, H. *et al.* (2017) 'Isolation and characterization of a novel pollen-specific promoter in
1079 maize (Zea mays L.)', *Genome*. Canadian Science Publishing, 60(6), pp. 485–495. doi:
1080 10.1139/gen-2016-0089.

1081 Wang, W. C. (2016) 'Techno-economic analysis of a bio-refinery process for producing Hydro-
1082 processed Renewable Jet fuel from Jatropha', *Renewable Energy*. Elsevier Ltd, 95, pp. 63–73.
1083 doi: 10.1016/j.renene.2016.03.107.

1084 Weselake, R. J. (2016a) 'Engineering oil accumulation in vegetative tissue', in *Industrial Oil*
1085 *Crops*. Elsevier Inc., pp. 413–434. doi: 10.1016/B978-1-893997-98-1.00015-4.

1086 Weselake, R. J. (2016b) 'Engineering Oil Accumulation in Vegetative Tissue', in *Industrial Oil*
1087 *Crops*. Elsevier, pp. 413–434. doi: 10.1016/B978-1-893997-98-1.00015-4.

1088 Wu, D. D., Irwin, D. M. and Zhang, Y. P. (2011) 'De novo origin of human protein-coding
1089 genes', *PLoS Genetics*, 7(11). doi: 10.1371/journal.pgen.1002379.

1090 Xiao, S. J. *et al.* (2010) 'TiSGeD: a database for tissue-specific genes', *Bioinformatics*. Narnia,
1091 26(9), pp. 1273–1275. doi: 10.1093/bioinformatics/btq109.

1092 Xiong, G., Cheng, K. and Pauly, M. (2013) 'Xylan O-Acetylation Impacts Xylem Development
1093 and Enzymatic Recalcitrance as Indicated by the Arabidopsis Mutant tbl29'. doi:
1094 10.1093/mp/sst014.

1095 Xu, C. and Shanklin, J. (2016) 'Triacylglycerol Metabolism, Function, and Accumulation in
1096 Plant Vegetative Tissues', *Annual Review of Plant Biology*. Annual Reviews, 67(1), pp. 179–
1097 206. doi: 10.1146/annurev-arplant-043015-111641.

1098 Xu, Xuan *et al.* (2019) 'Cell wall composition and transcriptomics in stem tissues of stinging
1099 nettle ( *Urtica dioica* L.): Spotlight on a neglected fibre crop', *Plant Direct*. John Wiley and
1100 Sons Inc., 3(8), p. e00151. doi: 10.1002/pld3.151.

1101 Xu, Xiaoyu *et al.* (2019) 'Upregulated Lipid Biosynthesis at the Expense of Starch Production in
1102 Potato (Solanum tuberosum) Vegetative Tissues via Simultaneous Downregulation of ADP-
1103 Glucose Pyrophosphorylase and Sugar Dependent1 Expressions', *Frontiers in Plant Science*.
1104 Frontiers Media S.A., 10. doi: 10.3389/fpls.2019.01444.

1105 Xu, X. Y. *et al.* (2020) 'A Synergistic Genetic Engineering Strategy Induced Triacylglycerol
1106 Accumulation in Potato (Solanum tuberosum) Leaf', *Frontiers in Plant Science*. Frontiers Media
1107 S.A., 11. doi: 10.3389/fpls.2020.00215.

1108 Yanai, I. *et al.* (2005) 'Genome-wide midrange transcription profiles reveal expression level
1109 relationships in human tissue specification', *Bioinformatics*. Narnia, 21(5), pp. 650–659. doi:
1110 10.1093/bioinformatics/bti042.

1111 Yang, J. H. and Wang, H. (2016) 'Molecular mechanisms for vascular development and
1112 secondary cell wall formation', *Frontiers in Plant Science*. Frontiers Media S.A. doi:
1113 10.3389/fpls.2016.00356.

1114 Yang, L. *et al.* (2016) 'PtrWRKY19, a novel WRKY transcription factor, contributes to the
1115 regulation of pith secondary wall formation in Populus trichocarpa', *Scientific Reports*. Nature
1116 Publishing Group, 6(1), pp. 1–12. doi: 10.1038/srep18643.

1117 Yang, Y. *et al.* (2015) 'Ectopic expression of WRINKLED1 affects fatty acid homeostasis in
1118 Brachypodium distachyon vegetative tissues', *Plant Physiology*. American Society of Plant
1119 Biologists, 169(3), pp. 1836–1847. doi: 10.1104/pp.15.01236.

1120 Ye, K. *et al.* (2009) 'Pindel: a pattern growth approach to detect break points of large deletions
1121 and medium sized insertions from paired-end short reads', 25(21), pp. 2865–2871. doi:
1122 10.1093/bioinformatics/btp394.

1123 Ye, Z.-H. (2002) ' V ASCULAR T ISSUE D IFFERENTIATION AND P ATTERN F
1124 ORMATION IN P LANTS ', *Annual Review of Plant Biology*. Annual Reviews, 53(1), pp. 183–
1125 202. doi: 10.1146/annurev.arplant.53.100301.135245.

1126 YH, L. and SR, J. (2003) '[Vascular-specific promoters and cis-regulatory elements].', *Sheng wu
1127 Gong Cheng xue bao = Chinese Journal of Biotechnology*, 19(2), pp. 131–135.

1128 Yu, X. *et al.* (2006) 'Computational analysis of tissue-specific combinatorial gene regulation:
1129 predicting interaction between transcription factors in human tissues', *Nucleic Acids Research*.
1130 Narnia, 34(17), pp. 4925–4936. doi: 10.1093/nar/gkl595.

1131 Zale, J. *et al.* (2016) 'Metabolic engineering of sugarcane to accumulate energy-dense
1132 triacylglycerols in vegetative biomass', *Plant Biotechnology Journal*. Blackwell Publishing Ltd,
1133 14(2), pp. 661–669. doi: 10.1111/pbi.12411.

1134 Zhang, L. *et al.* (2010) 'Altered xylem-phloem transfer of amino acids affects metabolism and
1135 leads to increased seed yield and oil content in Arabidopsis', *Plant Cell*. American Society of
1136 Plant Biologists, 22(11), pp. 3603–3620. doi: 10.1105/tpc.110.073833.

1137 Zhang, L. *et al.* (2014) 'Cloning and characterization of a novel Athspr promoter specifically
1138 active in vascular tissue', *Plant Physiology and Biochemistry*. Elsevier Masson, 78, pp. 88–96.
1139 doi: 10.1016/j.plaphy.2014.02.019.

1140 Zhang, Z. and Zhang, X. (2012) 'Argonautes compete for miR165/166 to regulate shoot apical
1141 meristem development', *Current Opinion in Plant Biology*. NIH Public Access, 15(6), pp. 652–

658. doi: 10.1016/j.pbi.2012.05.007.

Zhong, R. *et al.* (2008) 'A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in Arabidopsis', *Plant Cell*, 20(10), pp. 2763–2782. doi: 10.1105/tpc.108.061325.

Zhou, J., Sebastian, J. and Lee, J.-Y. (2011) 'Signaling and gene regulatory programs in plant vascular stem cells', *genesis*. John Wiley & Sons, Ltd, 49(12), pp. 885–904. doi: 10.1002/dvg.20795.

Zhu, Y. *et al.* (2013) 'PtrHB7, a class III HD-Zip Gene, Plays a Critical Role in Regulation of Vascular Cambium Differentiation in Populus', *Molecular Plant*, 6(4), pp. 1331–1343. doi: https://doi.org/10.1093/mp/sss164.

# Tables

## Table 1: The information of tissues sampled by development stage used in the transcriptome dataset used in this analysis from [McCormick et al., 2018].

| Tissue Type | Stage | Time of sample | Tissue Description | Sample ID |
|---|---|---|---|---|
| Leaves | Juvenile | DAE 8 | Leaf blade of first emerged leaf | leaf.blade.juv |
| | | | Lower portion (proximal) of third emerging leaf blade (in whorl) | leaf.lower.juv |
| | | | Upper portion (distal) of third emerging leaf blade | leaf.upper.juv |
| Roots | | | Central section of roots | root_bottom.juv |
| | | | Distal 1cm of root tips | root_top.juv |
| Stem | | | Entire stem (approximately 2mm in size) | stem_2mm.juv |
| Leaves | Vegetative | DAE 24 | Lower portion (proximal) of shoot, including sheaths, after removing first through fourth emerged leaves | leaf_lower_whorl.veg |
| | | | Middle portion of shoot after removing first through fourth emerged leaves | leaf_middle_whorl.veg |
| | | | Upper portion (distal) of shoot, including leaves emerged from whorl, after removing first through fourth emerged leaves | leaf_upper_whorl.veg |
| Roots | | | Central section of roots | root_bottom.veg |
| | | | Distal 1cm of root tips | root_top.veg |
| Stem | | | Entire stem (approximately 1cm in size) | stem_1cm.veg |
| Leaves | Floral initiation | DAE 44 | Leaf sheath of last fully emerged (ligulated) leaf | leaf_sheath_growing.floral |
| | | | Lower portion (proximal) of leaf blade of last fully emerged leaf (ligulated) | leaf_lower_growing.floral |
| | | | Upper portion (proximal) of leaf blade of last fully emerged leaf (ligulated) | leaf_upper_growing.floral |
| Roots | | | Central section of roots | root_middle.floral |
| | | | Distal 1cm of root tips | root_bottom.floral |
| Stem | | | Mature internode from middle of stem | internode_mature.floral |
| | | | 2cm long growing internode near apex (further from apex than 1cm growing internode) | internode_growing.floral |
| | | | 1cm long growing internode near apex (closer to apex than 2cm growing internode) | internode_growing_upper.floral |
| Reproductive Structures | | | Region of developing panicle in apex | panicle.floral |
| | | | Region of developing peduncle in apex | peduncle.floral |
| Leaves | Anthesis | DAE 65 | Leaf sheath of leaf below flag leaf | leaf_sheath_growing.anth |
| | | | Lower portion (proximal) of leaf blade of leaf below flag leaf | leaf_lower_growing.anth |
| | | | Upper portion (distal) of leaf blade of leaf below flag leaf | leaf_upper_growing.anth |
| Roots | | | Central section of roots | root_bottom.anth |
| | | | | |
| Stem | | | Mature internode from middle of stem | stem_mid_internode.anth |
| | | | Mature internode immediately below peduncle | Internode_flagleaf.anth |
| Reproductive Structures | | | Lower (proximal) half of panicle | panicle_lower.anth |
| | | | Upper (distal) half of panicle | panicle_upper.anth |
| Leaves | Grain maturity | DAE 96 | Leaf sheath of flag leaf | leaf_sheath_growing.grain |
| | | | Lower portion (proximal) of flag leaf blade | leaf_lower_growing.grain |
| Roots | | | Central section of roots | root_bottom.grain |
| | | | | |
| Stem | | | Mature internode from middle of stem | Stem_mid_internode.grain |
| | | | Mature internode immediately below peduncle | Internode_flagleaf.grain |
| Reproductive Structures | | | Mature, dormant, dry seed | Seed_dry.grain |
| | | | Developing seed, imbibed overnight | Seed_imbibed.grain |

## Table 2: Stem-related cellular functions of identified stem-specific and stem-preferred genes.

| Gene ID | TF Family | Ath Ortholog | Function | [Ref.] |
|---|---|---|---|---|
| *AOSS* | | | | |
| Sobic.001G106200<br>Sobic.001G106000 | *TALE* | *KNAT1* | Meristem maintenance through cell-to-cell communication, regulates internode development | (Byrne, Simorowski and Martienssen, 2002; Douglas *et al.*, 2002; Venglat *et al.*, 2002; Scofield, Dewitte and Murray, 2007; Helizon *et al.*, 2014b; Gao *et al.*, 2015; Kehr and Kragler, 2018) |
| *TSSE* | | | | |
| Sobic.001G075101 | *TALE* | *KNAT1* | Meristem maintenance through cell-to-cell communication, regulates internode development | et al. 2015; Venglat et al. 2002; Byrne, Simorowski, and Martienssen 2002; Helizon et al. 2014; Kehr and Kragler 2018) |
| Sobic.001G199200 | *YABBY* | *YAB1* | Expressed in primordia of all lateral organs and act to define the organ primordia domain and stimulate signals necessary for the dynamic partitioning of SAMs | (Goldshmidt *et al.*, 2008b; Stahle *et al.*, 2009b) |
| Sobic.001G286700 | *HD-ZIP* | *HB7* | preferentially expressed in cambial zone. Suppression of HB7 in poplar displayed significant reduction in xylem but increase in phloem and overexpression enhanced differentiation of cambial cells toward xylem cells but inhibited phloem differentiation. | (Zhu *et al.*, 2013; Ramachandran, Carlsbecker and Etchells, 2016) |
| Sobic.005G018900 | - | *PLC2* | Phospholipase C2 is required during the early stages of xylem differentiation, as indicated by the reduction of differentiated xylem cells upon application of PLC inhibitors | (Kanehara *et al.*, 2015; Gujas and Rodriguez-Villalon, 2016) |
| Sobic.007G183200 | *bHLH* | *MYC2* | JA-mediated xylem development, promotes cambial activities | (Miyashima *et al.*, 2013; Jang *et al.*, 2017; Sae-Lim *et al.*, 2019) |
| Sobic.008G151600 | | *TBL29* | trichome birefringence-like (TBL29) mutant phenotype in Arabidopsis had collapsed xylem and reduced plant growth | (Xiong, Cheng and Pauly, 2013; Bensussan *et al.*, 2015; Pauly and Ramírez, 2018; Xuan Xu *et al.*, 2019) |
| Sobic.009G189900 | | *CLV3* | suppression of stem cell proliferation in the meristem and Excess CLV3 promote differentiation from stem cells | (Ito *et al.*, 2006; Hirakawa *et al.*, 2008; Kondo *et al.*, 2011; Růžička *et al.*, 2015) |
| Sobic.004G298400 | *WRKY* | *WRKY12* | WRKY12 negatively regulate the expression of NST2, which acts as master regulators play a critical role in switching on and off secondary wall biosynthesis and represses cell wall biosynthetic genes in pith cells | (Sanchez, Nehlin and Greb, 2012; Gallego-Giraldo *et al.*, 2016; Yang and Wang, 2016; Yang *et al.*, 2016) |
| *TSSL* | | | | |
| Sobic.002G299200 | | *AGO1* | AGO1 is required for the function of most miRNAs including miR165/166 to repress the HD-ZIP TFs such as PHB, PHV, REV related to developmental regulators required for SAM establishment, vascular development | (Kidner and Martienssen, 2005; Mallory *et al.*, 2009; Zhang and Zhang, 2012) |
| Sobic.006G166300 | *WRKY* | *WRKY12* | WRKY12 negatively regulate the expression of NST2, which acts as master regulators play a critical role in switching on and off secondary wall biosynthesis and represses cell wall biosynthetic genes in pith cells | (Sanchez, Nehlin and Greb, 2012; Gallego-Giraldo *et al.*, 2016; Yang and Wang, 2016; Yang *et al.*, 2016) |

**Table 3:Network characteristics in the Sorghum BTx623 spatial-temporal stem-preferred gene regulatory network.** To characterize the topology of the stem-preferred gene's temporal GRNs, each network's topological characteristics were computed by NetworkAnalyzer in Cytoscape.

| **AOSS** | | | | | |
|---|---|---|---|---|---|
| | Juvenile | Vegetative | Floral Initiation | Anthesis | Grain Maturity |
| Total No. of genes in the network (stem genes + TF) | 7+751 | 7+731 | 8+871 | 751+8 | 8+724 |
| Total No. of regulatory edges | | | | | |
| Median no. of connections per gene | | | | | |
| In-Degree | 6 | 4 | 28 | 14 | 12 |
| Out-Degree | 5 | 4 | 21 | 15 | 11 |
| Shortest path length | 13.4 | 16.6 | 3.7 | 4.5 | 4.7 |
| Clustering coefficient | 0.306 | 0.292 | 0.215 | 0.234 | 0.225 |
| Connectivity of out-node distribution | | | | | |
| Degree exponent | 0.182 | 0.082 | 0.248 | 0.211 | 0.217 |
| R=Squared | 0.661 | 0.203 | 0.679 | 0.653 | 0.653 |
| | | | | | |
| **TSS-Early** | | | | | |
| Total No. of genes in the network (stem genes + TF) | 42+779 | 38+740 | 38+875 | 15+751 | 13+724 |
| Total No. of regulatory edges | | | | | |
| Median no. of connections per gene | | | | | |
| In-Degree | 6 | 4 | 29 | 14 | 12 |
| Out-Degree | 5 | 4 | 20 | 15 | 11 |
| Shortest path length | 13.4 | 16.8 | 3.7 | 4.5 | 4.7 |
| Clustering coefficient | 0.305 | 0.285 | 0.211 | 0.233 | 0.223 |
| Connectivity of out-node distribution | | | | | |
| Degree exponent | 0.169 | 0.092 | 0.252 | 0.204 | 0.237 |
| R=Squared | 0.653 | 0.392 | 0.675 | 0.589 | 0.711 |
| | | | | | |
| **TSS-Late** | | | | | |
| Total No. of genes in the network (stem genes + TF) | 2+460 | 3+729 | 4+870 | 4+750 | 5+723 |
| Median no. of connections per gene | | | | | |
| In-Degree | 6 | 4 | 29 | 14 | 12 |
| Out-Degree | 5 | 4 | 21 | 14 | 11 |
| Clustering coefficient | 0.290 | 0.291 | 0.214 | 0.233 | 0.225 |
| Connectivity of out-node distribution | | | | | |
| Degree exponent | 0.230 | 0.071 | 0.245 | 0.215 | 0.221 |
| R=Squared | 0.629 | 0.150 | 0.674 | 0.649 | 0.686 |

**Table 4:A table of SNVs and SVs within cis-regulatory elements in 1.5 Kbp upstream non-coding region of stem-specific genes with significant genetic alterations.**

| GENE_ID | CHR | POS | REF | ALT | TYPE | CRE Family |
|---|---|---|---|---|---|---|
| Sobic.001G106000 | Chr01 | 8142554 | A | G | SNV | NAC, TBP |
| | | 8142824 | T | C | SNV | SBP |
| | | 8142859 | T | C | SNV | AT-Hook |
| | | 8142864 | T | C | SNV | GATA, ARR-B |
| | | 8142891 | C | G | SNV | bZIP |
| | | | | | | |
| Sobic.001G106200 | Chr01 | 8178898 | TAAAAAAAA | TAAAAAAAAAA | INS | AT-Hook, C2H2 |
| | | 8179306 | A | G | SNV | AT-Hook |
| | | 8179602 | T | G | SNV | AT-Hook |
| | | 8179836 | G | T | SNV | MADS Box, AT-Hook |
| | | 8179837 | T | G | SNV | MADS Box, AT-Hook |
| | | | | | | |
| Sobic.006G147500 | Chr06 | 50913596 | GTTT | GTT | DEL | C2H2 |
| | | 50913790 | TAA | TAAA | INS | ZF-HD |
| | | 50914371 | A | C | SNV | GATA |
| | | 50914827 | A | G | SNV | MADS Box |
| | | | | | | |
| Sobic.002G419500 | Chr02 | 76700088 | GTTTTTT | GTTTTTTT | INS | SBP, AP2 |
| | | 76700683 | CTAT | CTATTAT | INS | WRKY |
| | | 76700882 | C | G | SNV | ERF, bHLH |
| | | 76700897 | GGCGCGCGCGCGCGCG | GGCGCGCG | DEL | E2F/DP |

**Table 5:Genetic alterations within cis-regulatory elements in 1.5 Kbp upstream non-coding region between Della and BTx623 cultivars for predicted regulatory TFs of Sobic.006G024800 during a) Early stage and b) Late stage.**

| GENE_ID | CHR | POS | REF | ALT | TYPE | CRE Family |
|---|---|---|---|---|---|---|
| Sobic.001G513700 | Chr01 | 78104570 | CAAAAAAAAA | CAAAAAAAA | DEL | RAV, MADS Box |
| | | 78104813 | ATATTCTTCGT | AT | DEL | G2-like |
| | | 78105411 | CGTAGTGTAGTGTAGTGTAGTGTAG | CGTAGTGTAGTGTAGTGTAGTGTAGTGTAG | INS | ERF, C2H2 |
| | | | | | | |
| Sobic.002G225100 | Chr02 | 61694212 | ATTTTT | ATTTTTT | INS | AT-Hook |
| | | 61694222 | C | T | SNV | TALE |
| | | 61694225 | A | G | SNV | TALE |
| | | 61694217 | T | <RPL> | RPL | TALE |
| | | 61694876 | C | A | SNV | bZIP, bHLH |
| | | 61694888 | G | A | SNV | bZIP, bHLH |
| | | 61694950 | T | G | SNV | bZIP |
| | | 61695090 | C | G | SNV | bZIP |
| | | 61695217 | AAGGAGGAGGA | AAGGAGGAGGAGGAGGAGGAGGAGGAGGA | INS | bHLH, bZIP |
| | | | | | | |
| Sobic.008G164800 | Chr08 | 59819801 | A | ATATAT | INS | TBP, bZIP, AT-Hook |
| | | 59819807 | TTCAGGGAAATTTCCTTGTGGTGCCATATATATGTGGAGATGGAGAGAGCTGTGAACAACACATTG | TATATTTCAGGGAAATTTCCTTGTGGTGCCATATATATGTGGAGATGGAGAGAGCTGTGAACAACACATTGT | RPL | TBP, AT-Hook |
| | | 59820200 | ACTTGTGAAAAGAAACTGAAATACAGAGAGAT | AGACTTGTGAAAAGAAACTGAAATAGAGAGAGATAGAGAGCAGGCTGAG | RPL | bZIP |
| | | 59819801 | ATATATT | ATATATTATATT | INS | TBP, AT-Hook |

## Figure Legends

**Figure 1:** Clustering of tissues.

**Figure 2:** Global gene expression of stem-specific genes across stem and non-stem tissues through all developmental stages.

**Figure 3:** A set of GO terms enriched in AOSS, TSS-Early, and TSS-Late stem-preferred genes.

**Figure 4:** A set of unique cis-element motifs enriched in AOSS, TSS-Early, and TSS-Late gene promoters.

**Figure 5:** Temporal gene regulatory network of a) AOSS, b) TSSE, c) TSSL genes through five developmental stages in BTx623 sorghum.

**Figure 6:** Temporal gene regulatory network of AOSS genes through development in BTx623 and Della cultivars.

**Figure 1:** Temporal gene regulatory network of TSSE genes through development in BTx623 and Della cultivars.

**Figure 8:** Temporal gene regulatory network of TSSL genes through development in BTx623 and Della cultivars.

**Figure 9:** Predicted gene regulatory networks of Sobic.006G26300 (one of most stem-specific and highly expressed gene) for BTx623 and Della for comparable time points.

**Figure 10:** Predicted regulatory TFs of Sobic.006G024800 during a) Early stage and b) Late stage. The regulatory interactions shown here are predicted at PCC & MR >= 0.9.

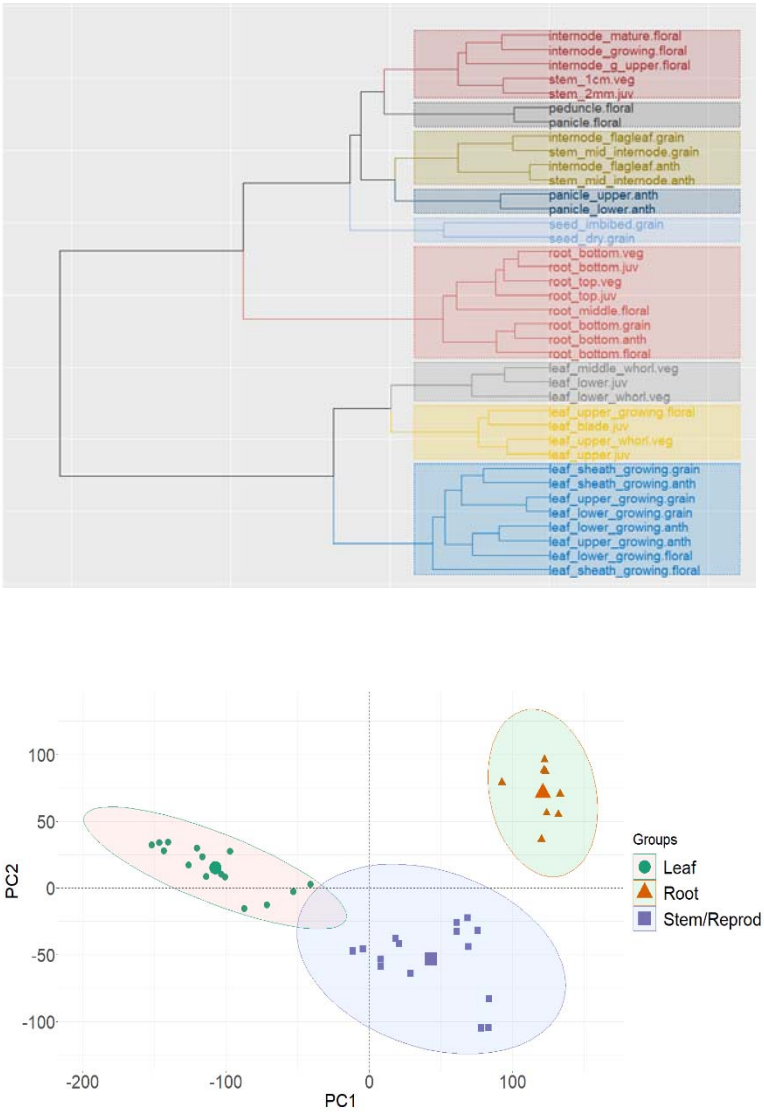


**Figure 1: Clustering of tissues.** A) Hierarchical Clustering analysis of the 38 tissue samples based on the expression of 23,674 genes (expressed in at least one sample, TPM ≥ 5). B) First two principal components of the spatial-temporal samples colored based on 4 major tissue types using expression values of all genes.

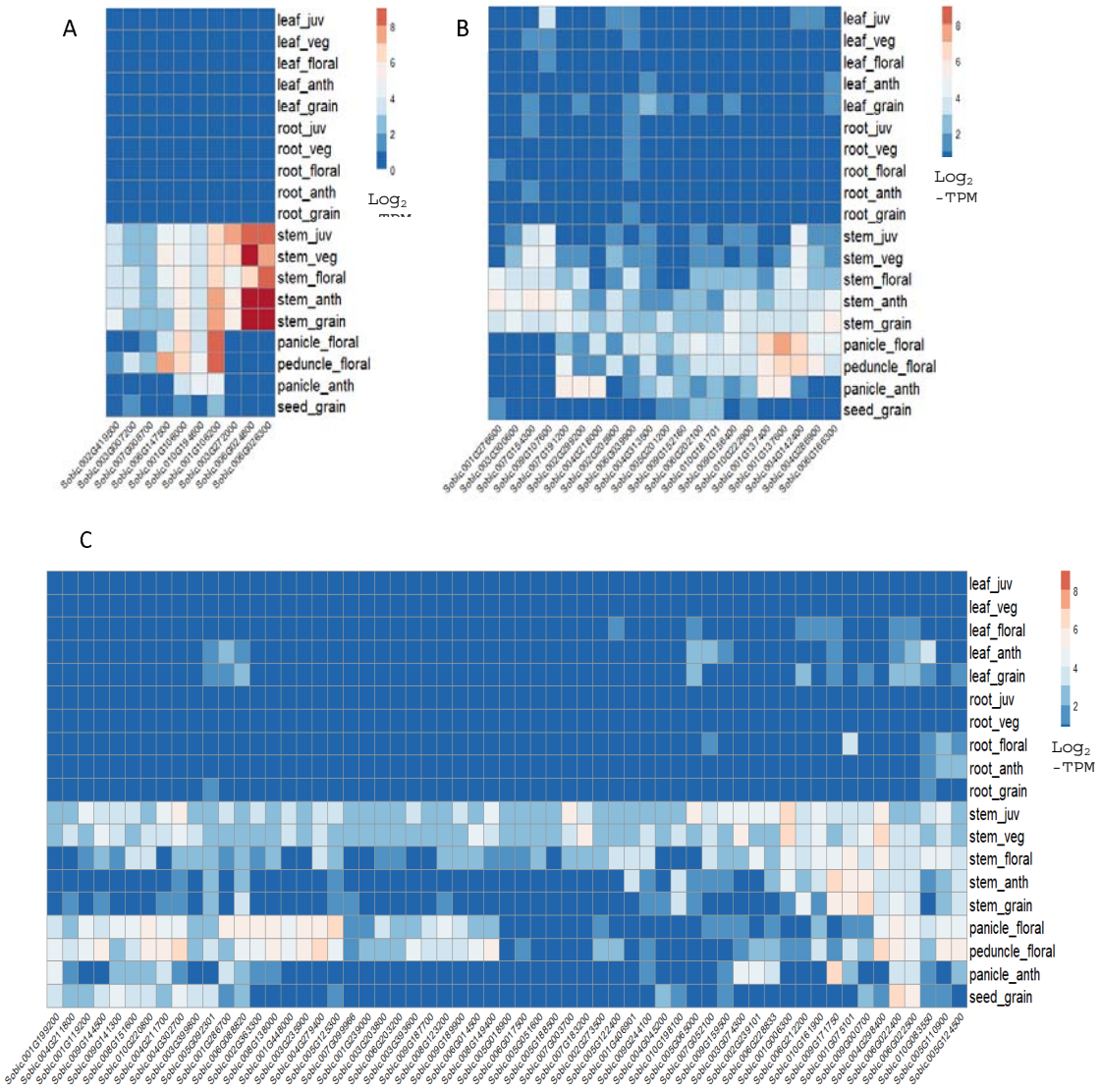


**Figure 2:Global gene expression of stem-specific genes across stem and non-stem tissues through all developmental stages.** A) Gene expression profile of 10 AOSS (Always-On-Stem-Specific) genes**,** B) Gene expression profile of 59 TSS-Early (Temporal stem-specificity during Early development), B) Gene expression profile of 21 TSS-Late (Temporal Stem-Specificity during Late development). Each row represents the mean gene expression values/ quantile normalized log2-TPM values. The mean expression values for each gene in each stage are represented by the intensity of color (blue representing no expression and orange representing maximum expression).

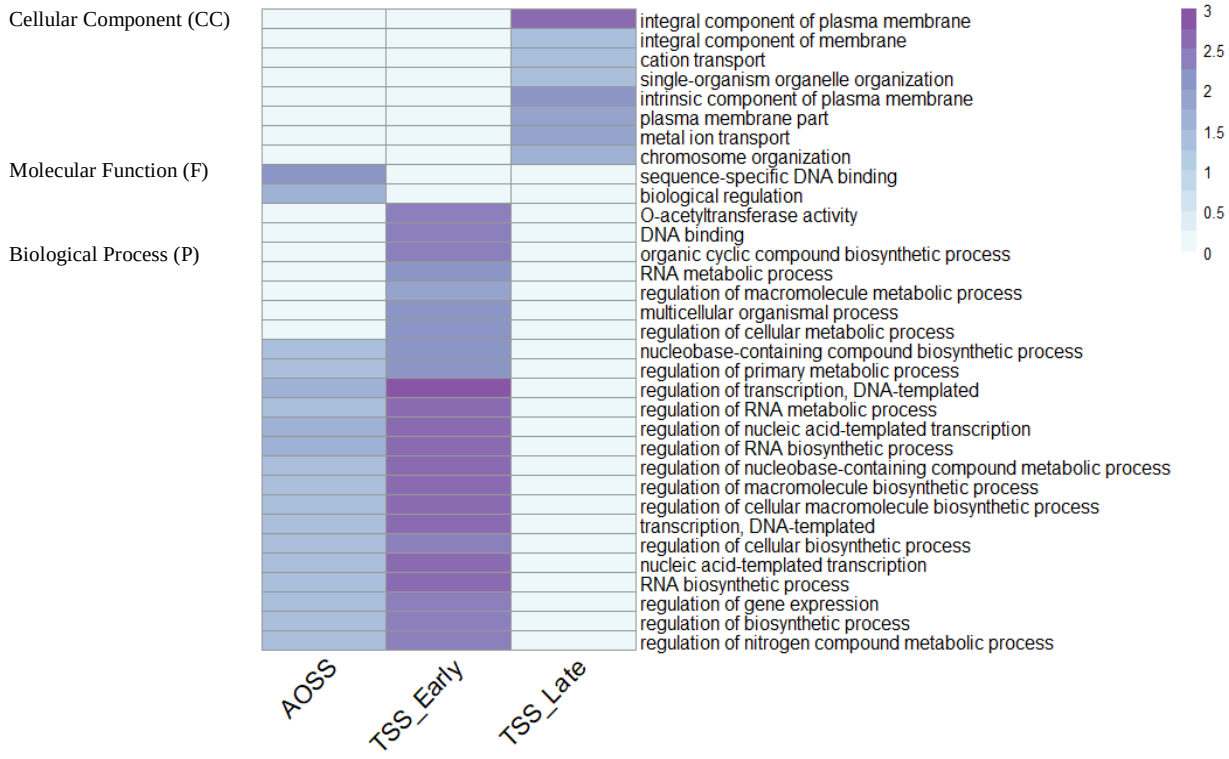


**Figure 3: A set of GO terms enriched in AOSS, TSS-Early, and TSS-Late stem-preferred genes.** A) Biological Processed GO Terms. B) Molecular Function and Cellular Component GO Terms. Significantly enriched GO terms are selected based on p-value < 0.05.

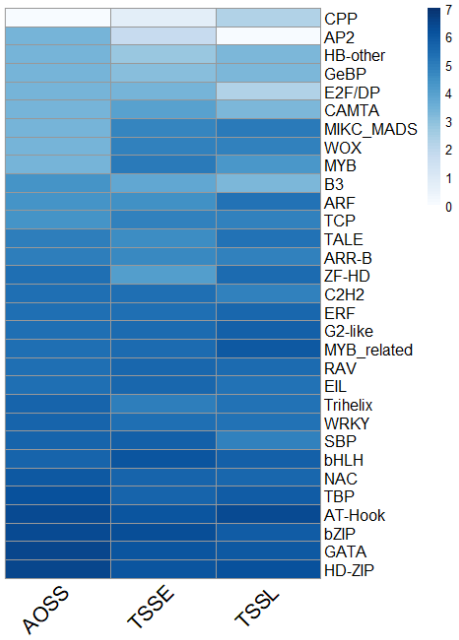


**Figure 4: A set of unique cis-element motifs enriched in AOSS, TSS-Early, and TSS-Late gene promoters.** The enrichment of cis-element motifs is identified based on over-representation of CREs in the 1.5 kb upstream promoters of stem-preferred genes, comparing to the whole genome (fold change > 1.1). The color intensity in the heatmap refers to percentage of stem-preferred sequences with enriched CRE.

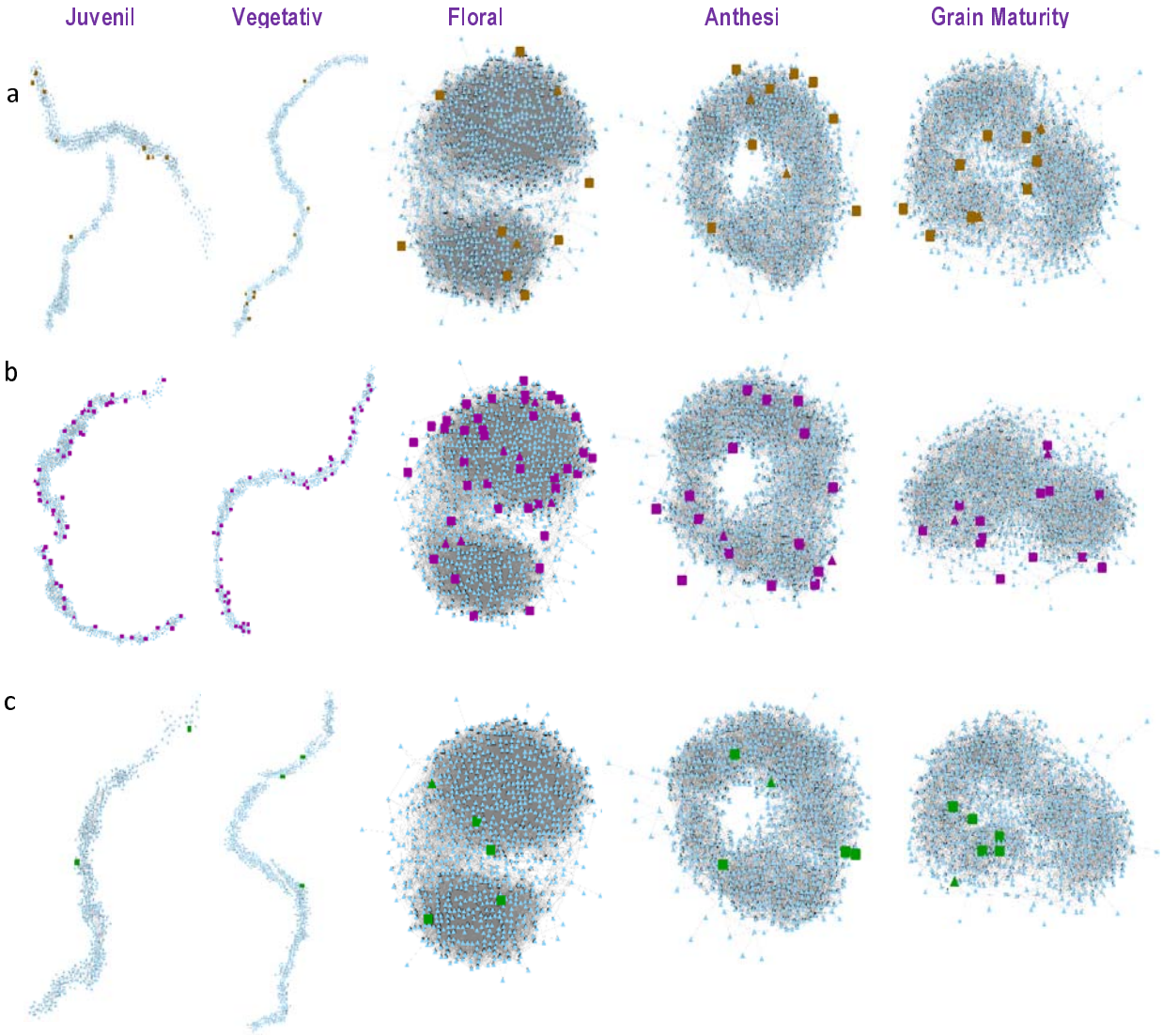


**Figure 5: Temporal gene regulatory network of a) AOSS, b) TSSE, c) TSSL genes through five developmental stages in BTx623 sorghum.** Edges with Pearson Correlation >= 0.7 and MR >= 0.7 is shown here.

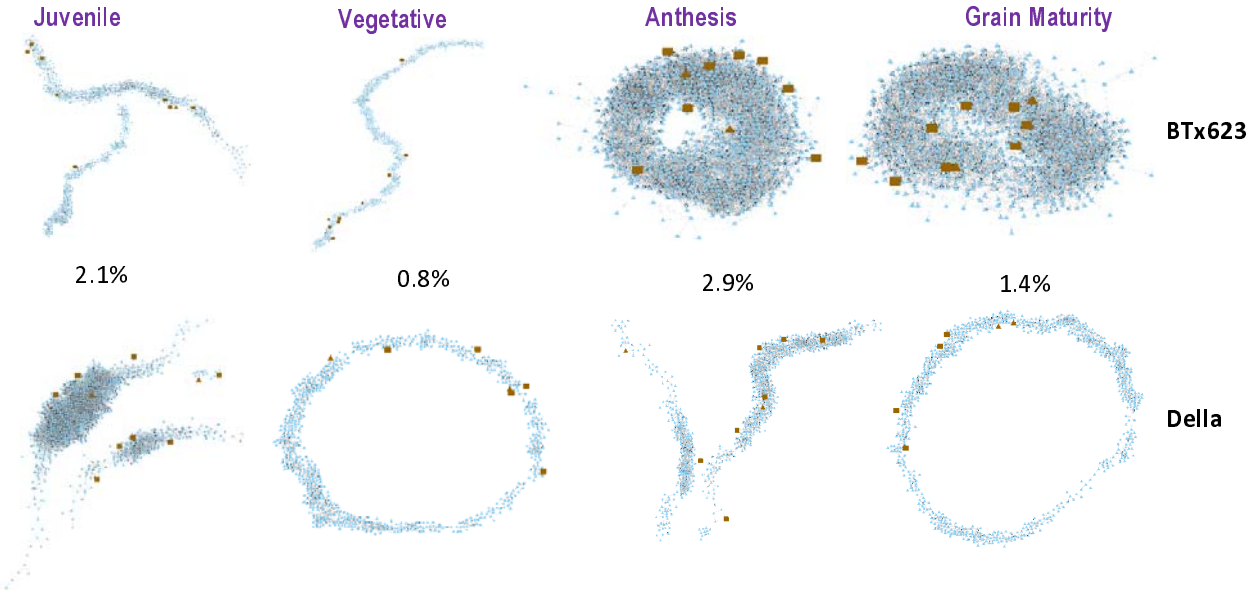


**Figure 6: Temporal gene regulatory network of AOSS genes through development in BTx623 and Della cultivars.** Edges with Pearson Correlation >= 0.7 and MR >= 0.7 is shown here. The percentage of shared regulatory edges are listed.

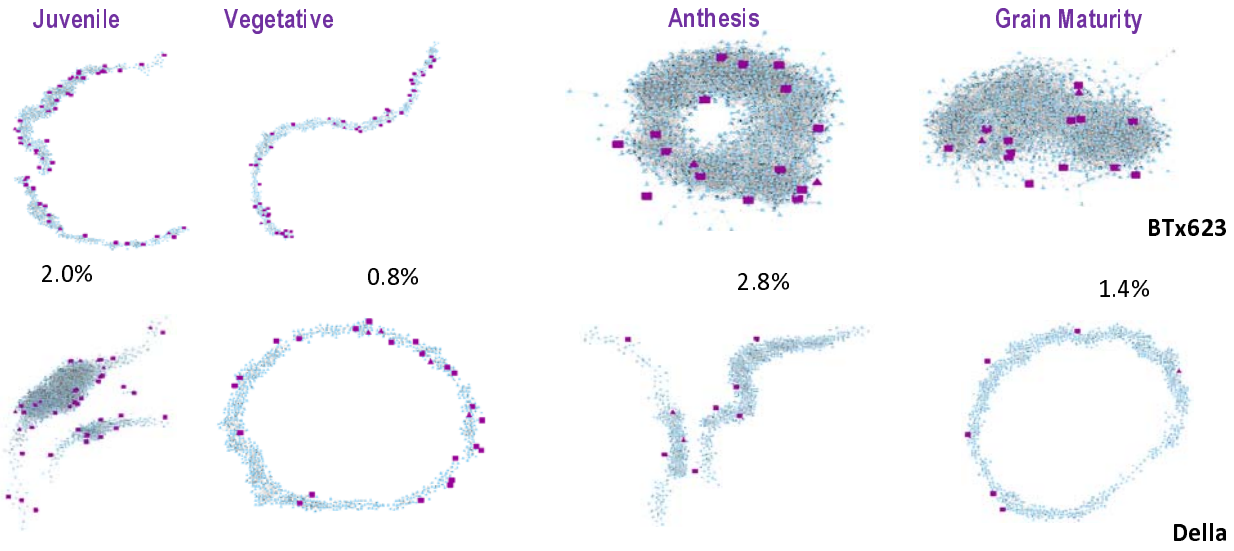


**Figure 7: Temporal gene regulatory network of TSSE genes through development in BTx623 and Della cultivars.** Edges with Pearson Correlation >= 0.7 and MR >= 0.7 is shown here. The percentage of shared regulatory edges are listed.

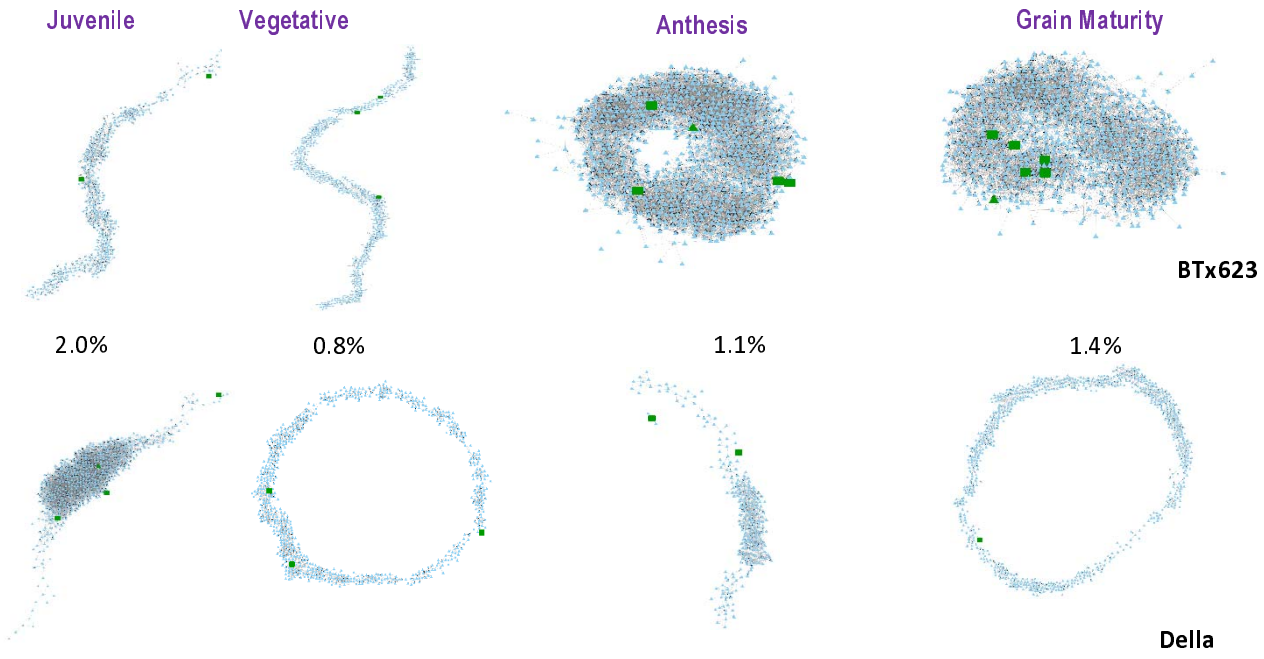


**Figure 8: Temporal gene regulatory network of TSSL genes through development in BTx623 and Della cultivars.** Edges with Pearson Correlation >= 0.7 and MR >= 0.7 is shown here. The percentage of shared regulatory edges are listed.


**Figure 9: Predicted gene regulatory networks of Sobic.006G26300 (one of most stem-specific and highly expressed gene) for BTx623 and Della for comparable time points.**

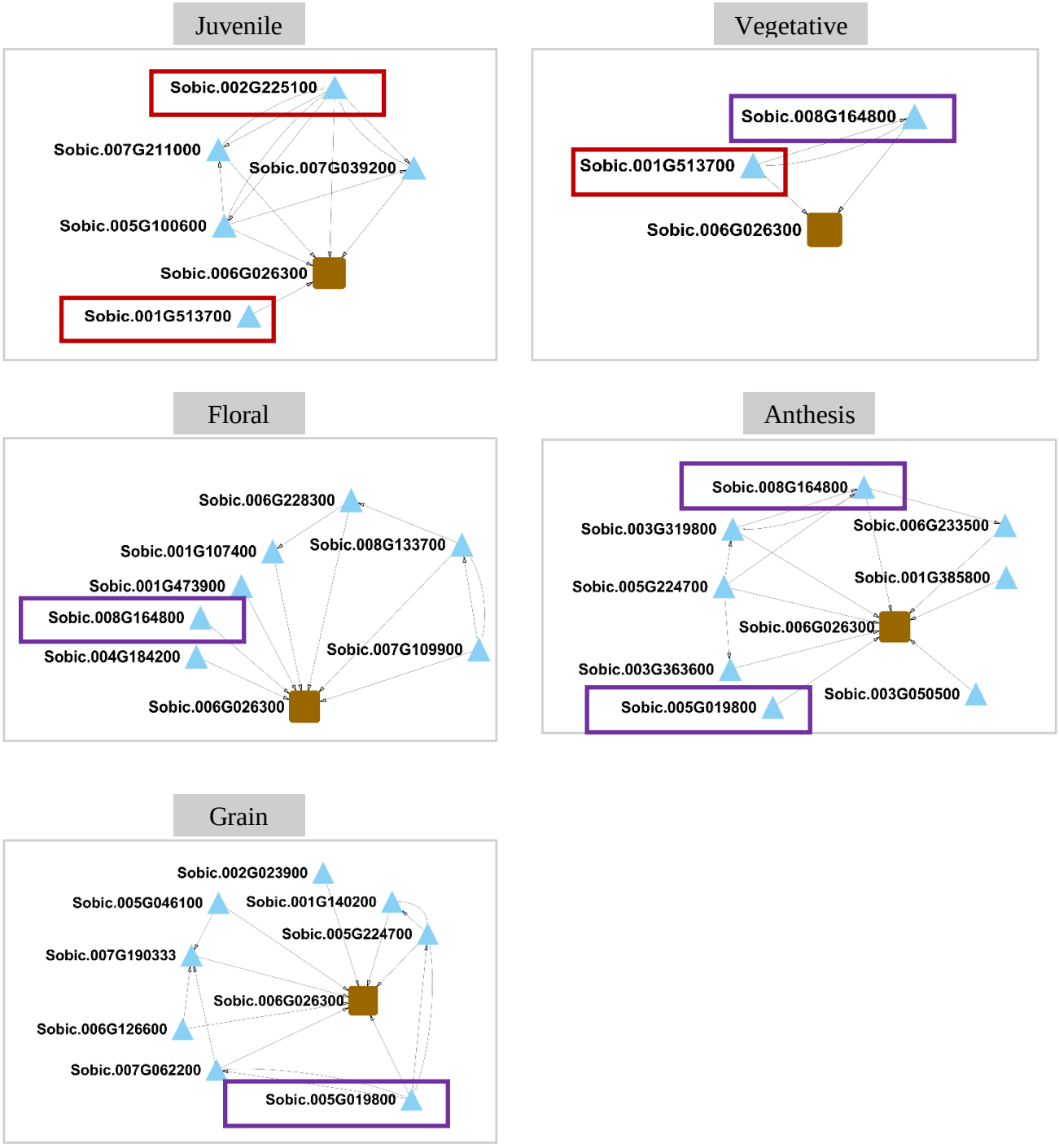


**Figure 10: Predicted regulatory TFs of Sobic.006G024800 during a) Early stage and b) Late stage. The regulatory interactions shown here are predicted at PCC & MR >= 0.9.**