DISSECTING THE GENETIC BASIS OF CARBOHYDRATE PARTITIONING IN MAIZE AND SUGARCANE

A Dissertation

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

SINGHA RAJ DHUNGANA

Dr. David M. Braun, Dissertation Supervisor

May, 2022

The undersigned, appointed by the dean of the Graduate School,

have examined the dissertation entitled

DISSECTING THE GENETIC BASIS OF CARBOHYDRATE PARTITIONING

IN MAIZE AND SUGARCANE

Presented by Singha Raj Dhungana,

A candidate for the degree of

Doctor of Philosophy,

And hereby certify that, in their opinion, it is worthy of acceptance.

_____

Dr. David M. Braun

_____

Dr. James A. Birchler

_____

Dr. Ruthie Angelovici

_____

Dr. Trupti Joshi

ACKNOWLEDGEMENTS

First, I would like to express my sincere gratitude to my advisor Dr. David Braun for his guidance and support throughout my graduate career. As my teacher and mentor, he has taught me more than I could ever give him credit for. I will forever cherish the experience of working with him and be grateful for his mentorship that has shaped me into the scientist that I am today.

I would like to thank my committee members Dr. Jim Birchler, Dr. Ruthie Angelovici, and Dr. Trupti Joshi for their support and advice. I would specially like to thank Dr. Jim Birchler for giving me opportunity to rotate in his lab and learn about maize cytogenetics, Dr. Ruthie Angelovici for advice on career choices after graduation, and Dr. Trupti Joshi for always being ready to help me whenever I had any questions.

I would like to thank all the current and former members of the Braun lab for the advice and technical assistance on my projects, as well as always being open to discussion about a variety of topics and projects. I would specially like to thank Tyler McCubbin for his assistance in various experiments. I would also like to thank Dr. Frank Baker for his advice and help with microscopy. I would like to thank Dr. Ben Julius, Dr. Rachel Mertz, Dr. Kristen Leach, and Dr. Thu Tran who were present in the lab when I joined and trained me and provided guidance in the first year of graduate school.

I would like to thank Dr. Diane Janick-Buckner and Dr. Brent Buckner who welcomed me to their labs at Truman State University and introduced me to the amazing crop maize for research. I cherish their mentorship, support and guidance that paved the path to where I am today.

I would like to thank all my friends and colleagues who made my stay in Columbia enjoyable and to my friends who despite the distance have been there for me.

I would like to thank my family here in the USA and back in Nepal for their support. My parents and my brothers have always supported me and encouraged me to pursue my ambitions with their endless love and care. They have always believed in me and have been my support to get me to where I am today and without their support it would not have been possible.

Finally, I would like to thank my wife, Deepti, for her love, support, and patience. She pushes me to be a better human and without her support it would not have been possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Carbohydrate partitioning, the process by which carbohydrates synthesized in the photosynthetic source tissues (mature leaves) are mobilized to non-photosynthetic (sink) tissues, such as roots, seeds, and developing organs is crucial for plant growth, development, and yield. Various physiological, biochemical, and anatomical studies have addressed this process, but the genetic control of carbohydrate partitioning is still not well understood. The main purpose of this dissertation is to elucidate aspects of the genetic control of carbohydrate partitioning in maize and sugarcane.

Chapter 1 summarizes our current knowledge of phloem loading in grasses, the principal families of sugar transporters involved in sucrose transport, and novel mechanisms by which the activities of these sugar transporters are modulated. It also elaborates some of the recent discoveries in some eudicot species.

Chapter 2 describes the genome of the wild ancestor of sugarcane (*Saccharum spontaneum*) and presents some ideas about the genes that may regulate sugar storage in the modern sugarcane varieties. For this study, I annotated *Sucrose Transporters (SUTs)* and *Tonoplast Sugar Transporters (TSTs)* in the *S. spontaneum* genome. Similarly, Chapter 3 describes the sugar transporter families; SUTs and TSTs in low (*Saccharum spontaneum*) and high sugar accumulating (*Saccharum officinarum*) sugarcane species and takes a comparative genomics approach to understand the ability of modern sugarcane cultivars to store huge amounts of sugars in their stem.

Chapter 4 describes the characterization and cloning of maize *carbohydrate partitioning defective13 (cpd13)* and *carbohydrate partitioning defective35 (cpd35)* mutants, whereas

Chapter 5 describes the characterization and progress towards cloning of *carbohydrate partitioning defective60 (cpd60)* and *carbohydrate partitioning defective87 (cpd87)* mutants. All of these mutants are recessive and contribute to carbohydrate hyperaccumulation in the mature leaves of the mutant due to reduced sucrose export. The *cpd13* and *cpd35* mutations affect a Dna-J-thioredoxin-like protein, which is hypothesized to be involved in processing of proteins and have chaperone-like activities. The *cpd60* and *cpd87* mutations result in ectopic lignin in the phloem of mature leaves, but the mechanism and the gene involved remain to be identified.

Chapter 6 summarizes the discoveries made in the previous chapters and presents future research directions.

Appendix A is a research article that I collaborated on and contributed to the measurement of non-structural carbohydrates in *very oil yellow1 (vey1)* leaves introgressed to B73 and Mo17 inbred lines.

Collectively, the research presented here has enhanced our understanding of the genetic control of carbohydrate partitioning in maize and sugarcane. These studies establish the foundations for future experiments to determine the genetic architecture controlling carbohydrate partitioning in plants.

Chapter 1 : Sugar transporters in grasses: Function and modulation in source and storage tissues

Note: The information in this chapter was published under the title:

Dhungana, S.R., and Braun, D.M., 2021. Sugar transporters in grasses: Function and modulation in source and storage tissues. Journal of Plant Physiology 266, 153541.

https://doi.org/10.1016/j.jplph.2021.153541

ABSTRACT

Carbohydrate partitioning, the process of transporting carbohydrates from photosynthetic (source) tissues, such as leaves, to non-photosynthetic (sink) tissues, such as stems, roots, and reproductive structures, is vital not only for the growth and development of plants but also for withstanding biotic and abiotic stress. In many plants, sucrose is the primary form of carbohydrate loaded into the phloem for long-distance transport and unloaded into the sink tissues for utilization or storage. We highlight recent findings about 1) phloem loading in grasses, 2) the principal families of sugar transporters involved in sucrose transport, and 3) novel mechanisms by which the activities of sugar transporters are modulated. We discuss exciting discoveries from eudicot species that provide valuable insights regarding the regulation of these sugar transporters, which may be translatable to monocot species. As we better understand the intricate pathways that control the activities of various sugar transporters, we can utilize this knowledge for developing improved crop varieties.

Keywords: Apoplasmic loading, Phloem, Sugar Transporters, SUTs, TSTs

Abbreviations: BS Bundle Sheath, CC Companion Cells, EBEs effector-binding elements, INV, Invertase, M Mesophyll cells, PP Phloem Parenchyma cells, SP Storage Parenchyma cells, SUTs Sucrose Transporters, SWEETs Sugars Will Eventually be Exported Transporters, TALEs transcription-activator-like effectors, TSTs Tonoplast Sugar Transporters, VI Vacuolar Invertase

INTRODUCTION

Carbohydrates are the main source of energy required for the growth and development of plants. As all heterotrophic life on earth depend on plants as the primary source of food, understanding how plants synthesize and mobilize carbohydrates is essential. Carbohydrates, which are synthesized in the photosynthetic (source) tissues, such as mature leaves, must be transported to other non-photosynthetic (sink) organs, such as stems, roots, and reproductive structures. This process, termed carbohydrate partitioning, is not only indispensable for the growth and development of plants but also for withstanding biotic and abiotic stress (Braun and Slewinski, 2009; Julius et al., 2017; Lemoine et al., 2013). Hence, understanding the control of this process is critical to develop crop varieties producing high yield and withstanding harsher environmental conditions to feed the growing human population.

Over the years, multiple studies in crops plants, such as rice (*Oryza sativa*), maize (*Zea mays*), sorghum (*Sorghum bicolor*), sugarcane (*Saccharum officinarum*), and other eudicot species have been conducted to understand the process of carbohydrate partitioning at the genetic, biochemical, and physiological levels. However, we do not yet have a concrete understanding of how this process is regulated, especially in grasses like maize and rice, which are staple crops for the majority of the world.

In plants that utilize sucrose as the primary sugar for long-distance transport, it must be loaded into the phloem sieve tubes, which are composed of sieve elements (SE) connected end to end (Julius et al., 2017; Slewinski and Braun, 2010; Zimmermann and Ziegler, 1975). Mature SE are enucleate and rely on metabolic support from companion cells (CC), and hence these cells are referred to as SE-CC complexes (Esau, 1977). These SE-CC complexes have very few plasmodesmatal (cellular conduits connecting the cytosol of adjacent cells) connections to neighboring cells in most crop species (Evert et al., 1978; Evert et al., 1996; Gamalei, 1989; Haritatos et al., 2000; Robinson-Beers and Evert, 1991). Therefore, they must rely on transporters to move metabolites into these cells. Sucrose is synthesized in the mesophyll (M) cells, moves symplasmically (cell to cell through plasmodesmata) into neighboring bundle sheath (BS) cells, and then on into phloem parenchyma (PP) cells. Sucrose is exported into the apoplast by Sugars Will Eventually be Exported Transporters (SWEETs) (Fig. 1.1) (Chen et al., 2012; Eom et al., 2015). Sucrose Transporters (SUTs, also called SUCs) function to load sucrose into the SE-CC complex, which contains a higher concentration of sugar compared to neighboring cells (Evert et al., 1978; Geiger et al., 1973; Julius et al., 2017; Lalonde et al., 2004; Sauer, 2007). This process of actively loading sucrose into the SE-CC is known as apoplasmic (from the cell wall space) phloem loading. Sucrose is then delivered through the sieve tubes to distant sink tissues, where it is stored or metabolized into other compounds (Ruan, 2014). Another group of sugar transporters, Tonoplast Sugar Transporters (TSTs), previously referred to as Tonoplast Monosaccharide Transporters (TMTs), function in sucrose and hexose storage in various tissues in most monocots and eudicots by importing sugars into the vacuole (Cheng et al., 2018; Jung et al., 2015; Ren

et al., 2021; Wingenter et al., 2010; Wormit et al., 2006). These TSTs can function in both source tissues such as leaves and sink tissues such as roots or stems to sequester sugar, primarily sucrose (Xu and Liesche, 2021). In source tissues, the transient storage of sucrose in the vacuole can be remobilized at night to enable continued sucrose export to sink tissues. In sink tissues, TSTs can lead to sugar accumulation and storage in the vacuoles.

We present a conceptual framework of how apoplasmic sucrose loading occurs in the source tissues and how unloading occurs in sink and storage tissues of many annual crop species (Fig. 1.1). Sucrose is delivered to distant tissues by bulk flow caused by a hydrostatic pressure differential between the source and sink ends of the sieve tube (Knoblauch et al., 2016; Münch, 1930). Once sucrose reaches the sink/storage tissues, sucrose is unloaded from the sieve tube into the parenchyma cells either apoplasmically or symplasmically, depending on the location, age, and type of tissue. In symplasmic unloading, sucrose flows through plasmodesmata down its concentration gradient into adjacent cells until it reaches the storage parenchyma cells (SP). Here, sucrose can be imported into vacuoles by TSTs for storage (e.g., in sugar beet (*Beta vulgaris*) taproots or the stems of sugarcane or sweet sorghum) or converted into hexoses and stored as starch in plastids in grain crops (Fig. 1.1) (Ruan, 2014). In apoplasmic phloem unloading, sucrose is likely exported from the sieve tube by SWEETs and taken up into adjacent cells by SUTs. Sucrose then moves symplasmically through plasmodesmata into nearby parenchyma cells and is transported into vacuoles by TSTs (Fig. 1.1). In many cell types of both source and sink tissues, a distinct member of the SUT family (e.g., ZmSUT2 in maize (Leach et al., 2017)) is present on the vacuolar membrane and functions to export

sucrose out of the vacuole (Fig. 1.1). In addition, we note that there are other factors and processes not discussed in this review that influence sucrose availability and indirectly affect sucrose phloem loading and unloading processes, such as enzymes involved in its breakdown (invertases (INV) or sucrose synthases) or the H^+-ATPase localized to the CC-SE plasma membrane that impact source and sink strengths. For reviews discussing additional factors affecting sugar allocation in plants, please see (Bihmidine et al., 2013; Braun et al., 2014; Julius et al., 2017; Ruan, 2014; Stein and Granot, 2019; Xu and Liesche, 2021).

In this review, we will focus on sugar transporters that are involved in loading sucrose into the phloem in the source tissues and unloading and storing it in the sink/storage tissues, with an emphasis on grass species. We will highlight recent studies emphasizing the route for phloem loading of sucrose in maize and rice. Furthermore, we will discuss recent findings that have identified various mechanisms involving transcriptional, post-transcriptional, and post-translational modifications of sugar transporters.

Phylogenetic analyses of SUTs and TSTs

We provide a phylogenetic analysis of SUT and TST family members in select grass and eudicot species to illustrate the evolutionary relationships among sugar transporter family members (Figs. 2 and 3). Based on a phylogenetic analysis of amino acid sequences, SUTs can be divided into five groups, with two groups containing sequences found only in monocots (salmon and orange clades), one group in eudicots only (yellow), and two groups present in both eudicots and monocots (blue and green) (Fig. 1.2, See Supplementary Table S1.1 for accession IDs and sequences analyzed)

(Braun and Slewinski, 2009; Kumar et al., 2018; Letunic and Bork, 2021). AtSUC2, the most well-characterized SUT in *Arabidopsis thaliana* belongs to the eudicot-specific clade (yellow), whereas OsSUT1 and ZmSUT1, the first SUTs characterized in rice and maize, respectively, belong to the monocot-specific group (orange) containing SUT1 and SUT3 sequences (Fig. 1.2) (Aoki et al., 1999; Hirose et al., 1997). Members of the Group 4 SUTs (Fig. 1.2, blue), mostly localize to the tonoplast and are proposed to export transiently stored sucrose from the vacuole to the cytoplasm (Endler et al., 2006; Eom et al., 2011; Leach et al., 2017). One Group 4 SUT has also been localized to the plasma membrane (Chincinska et al., 2013).

Similarly, TSTs can be divided into monocot-specific and eudicot-specific clades based on a phylogeny of amino acid sequences (Fig. 1.3, See Supplementary Table S1.1 for accession IDs and sequences analyzed). We identified three clades specific to monocots (light blue, purple and brown) and a clade (light green) consisting of only eudicot sequences. Several of the TSTs from these clades have been well characterized to function in transport of sugars (sucrose and glucose) into storage vacuoles in sugar beet taproots, Arabidopsis, sorghum, rice, and many other species (Bihmidine et al., 2016; Cho et al., 2010; Jung et al., 2015; Wormit et al., 2006). Because OsTMT3, and OsTMT4 sequences (grey) share less than 60% identity to other rice, maize, or sorghum TST sequences, we tentatively classified these as TST-like sequences (Supplementary Table S1.2; Fig.3). Additional molecular and biochemical characterization of these TST-like proteins is needed to determine if they function as bona fide TST proteins.

Roles of sugar transporters in phloem loading in source tissues and long-distance transport

In many grasses, such as wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), maize, sugarcane, and sorghum, the proposed mechanism for sucrose entry into the phloem is apoplasmic loading (Bihmidine et al., 2015; Evert et al., 1978; Evert et al., 1996; Robinson-Beers and Evert, 1991; Thompson and Dale, 1981). In these species, SUTs play a crucial role in actively loading sucrose into the phloem cells. Here, we discuss the phloem loading mechanisms as observed in these major cereal crops, as typified by maize, and in rice, where the loading path has been debated (Braun et al., 2014; Eom et al., 2011; Eom et al., 2012; Eom et al., 2016; Julius et al., 2017; Scofield et al., 2007).

Sucrose phloem loading in maize

Based on anatomical, physiological, and genetic data, the path for sucrose phloem loading in maize has been proposed to be apoplasmic (see (Braun et al., 2014) for review). Recent data lend additional support to this hypothesis. In maize ZmSWEET13a, b, and c have been shown to be involved in exporting sucrose to the cell wall space (Bezrutczyk et al., 2018). By using a combination of single-cell RNA sequencing, in situ hybridization, and reporter gene analyses, Bezrutczyk et al. found that ZmSWEET13 transcripts are expressed in the abaxial BS cells in the smallest leaf veins (intermediate rank-2 veins) where the proteins function to export sucrose to the apoplasm (Bezrutczyk et al., 2021). This specialization of the abaxial BS surrounding the smallest veins in maize leaves is similar to the architecture in barley leaf veins (Williams et al., 1989). From the apoplasm, sucrose is transported into the CC by ZmSUT1 before moving

symplasmically into the SE for long-distance transport (Slewinski et al., 2010; Slewinski et al., 2009). Furthermore, *ZmSUT1* is expressed in other cell types in leaves, where it functions to retrieve sucrose from the apoplasm to prevent its loss to the transpiration stream (Baker et al., 2016; Slewinski et al., 2009). Collectively, this evidence further solidifies the hypothesis that maize uses apoplasmic loading of sucrose in the phloem.

Recent progress in deciphering the path of phloem loading in rice

Like many grass species, rice uses sucrose as the primary sugar for long-distance transport (Fukumorita and Chino, 1982; Hayashi and Chino, 1990; Scofield et al., 2007). Several studies in rice examined the role of OsSUTs in sucrose transport and sought to identify whether rice uses a symplasmic or apoplasmic phloem loading mechanism. Previous studies looking at the functions of OsSUTs have suggested that rice might use a modified form of symplasmic loading, mediated by the tonoplast-localized OsSUT2 (Eom et al., 2011; Eom et al., 2012), or alternatively suggested that rice may use a different SUT for apoplasmic loading (Braun et al., 2014; Eom et al., 2016; Julius et al., 2017; Scofield et al., 2007).

Recently, Wang et al. (2021) proposed that rice uses apoplasmic loading as a major phloem loading strategy. In their study, yeast INV was overexpressed under the control of the constitutive 35S promoter from the *Cauliflower Mosaic Virus* and targeted to the cell wall to digest any apoplasmic sucrose to prevent it from being loaded into phloem cells (Wang et al., 2021). The INV-expressing transgenic lines accumulated hexoses, sucrose, and starch in leaves, displayed restricted vegetative growth, and had decreased grain yields, which are similar to the phenotypes shown by the *Zmsut1* maize mutant, which is defective in phloem loading (Slewinski et al., 2010; Slewinski et al.,

2009). Furthermore, *OsSUT1* was highly expressed in source leaves, and analyzing transgenic plants containing the promoter of *OsSUT1* driving the expression of the *GUS* reporter gene showed that *OsSUT1* was expressed in vascular parenchyma cells and CC. Additionally, CRISPR/Cas9 gene-edited *Ossut1* mutants showed reduced growth and grain yield, suggesting that phloem loading is affected. These results support the hypothesis that rice uses apoplasmic phloem loading.

In further agreement with this hypothesis, the transgenic expression of *OsSUT1* in Arabidopsis was able to complement the *Atsuc2* mutant, indicating that OsSUT1 is a functional homolog of the Arabidopsis AtSUC2 protein and can transport sucrose into the phloem (Eom et al., 2016). If rice does use apoplasmic phloem loading mediated by OsSUT1, it is curious that no visible phenotype was previously observed in *Ossut1* mutants (Hirose et al., 2010; Ishimaru et al., 2001; Scofield et al., 2002). This lack of sugar accumulation within leaves and retarded plant-growth phenotype could be due to the environmental conditions that the plants were grown under, in particular low light (Julius et al., 2017; Srivastava et al., 2009). Consistent with this idea, Wang et al., (2021) grew the *Ossut1* mutant plants in the field and observed a mutant phenotype. However, other environmental factors might also contribute to the phenotype of the mutant plants.

Additional evidence in support of this hypothesis is provided by a recent study (Singh et al., 2021) who simultaneously increased the expression of OsSUT1 and two sucrose-transporting SWEETs, OsSWEET11 and OsSWEET14, under the control of their native promoters in rice. The authors reported several pleiotropic phenotypes involving decreased sucrose transport, consistent with proposed roles of these genes in apoplasmic phloem loading (Singh et al., 2021). The sugar transporter-overexpressing plants showed

an increased starch level in leaves but decreased soluble sugars, which were attributed to decreased expression of genes functioning in membrane transport and sucrose synthesis. We urge readers to interpret these data with appropriate caution as these results are taken from a recent preprint, which currently (Sept. 10, 2021) awaits publication in a peer reviewed journal.

Interestingly, the overexpression lines also showed resistance to the bacterial pathogen *Xanthomonas oryzae* pathovar *oryzae* (*Xoo*), which is hypothesized to be due to the transient increase of sugar in the apoplasm due to OsSWEET11;14 functions triggering defense mechanisms for pathogen starvation (Singh et al., 2021). OsSWEET11, OsSWEET13, and OsSWEET14 are known targets of *Xoo,* whose transcription-activator-like effectors (TALEs) bind to specific promoter regions within these genes, known as effector-binding elements (EBEs) (Blanvillain-Baufumé et al., 2017; Oliva et al., 2019). Oliva and colleagues (2019) used multiplexed CRISPR–Cas9 genome editing to mutate the EBEs of these *SWEETs* and were able to engineer a rice variety resistant to all currently known strains of *Xoo* (Oliva et al., 2019). This approach has exciting implications for increasing global rice production and food security.

Summing up, these results support the thesis that rice uses apoplasmic phloem loading, with OsSUT1 being the primary SUT involved in loading sucrose. Overall, in many if not all grass species, it is very likely that a similar mechanism of apoplasmic phloem loading of sucrose operates. Furthermore, the retrieval of sucrose from the apoplasm by SUTs expressed in cells besides CC might be a conserved function in many plants.

Regulation of sugar transporters

Sugar transporters serve crucial roles to load, unload, and store sucrose in plants. Even though the sugar transporters involved in sucrose transport of many crops and model species are known, very little is known about how they are regulated. Here, we discuss recent studies which have proposed various post-translational and transcriptional mechanisms by which the activities of sugar transporters are controlled.

Modulation of sugar transporters in rice

In considering the above studies that support that rice utilizes an apoplasmic phloem loading strategy, it is therefore relevant to understand how the functions of key sugar transporters, such as SUTs and SWEETs, are modulated. Exciting recent reports indicate that three transcription factors, DNA BINDING WITH ONE FINGER11 (OsDOF11) and two Nuclear Factor Y family members, OsNF-YB1 and OsNF-YC12, regulate the expression of several rice sugar transporter genes and thereby control the rate of sugar transport into various tissues (Fig. 1.4) (Bai et al., 2016; Wu et al., 2018; Xiong et al., 2019).

Chromatin immunoprecipitation (ChIP) assays showed that the transcription factor OsDOF11 directly binds to the promoter regions of *OsSUT1*, *OsSWEET11*, and *OsSWEET14* (Wu et al., 2018). Gene expression analyses of these sugar transporter genes in *Osdof11* mutants revealed that OsDOF11 coordinates the expression of the transporters functioning in apoplasmic phloem loading of sucrose. Further, in *Osdof11* mutant plants, the rate of sucrose transport was reduced compared to the wild type. Also, *Osdof11* mutant plants were less susceptible to infection by *Xoo*, suggesting that OsDOF11 influences sugar distribution during pathogenic invasion. In a follow up study, Kim, P. et

al. (2021) identified that overexpressing *OsDof11* provides resistance against *Rhizoctonia solani*, the causative agent of sheath blight disease (ShB) but decreased yield. The authors circumvented this by creating transgenic rice with *OsDOF11* fused to a transcriptional activation domain (Li et al., 2013), to obtain controlled tissue-specific expression of *OsDOF11* and *OsSWEET14*, resulting in increased yield as well as resistance to ShB (Kim et al., 2021b). These exciting results further prove that OsDOF11 and OsSWEET14 play an important role in sugar regulation during pathogen attack and are thus good candidates for developing pathogen-resistant varieties of crops.

Similarly, in the aleurone layer in rice seeds, the transcription factor OsNF-YB1 regulates grain filling through activating the expression of *OsSUT1*, *OsSUT3*, and *OsSUT4* (Bai et al., 2016). RNA in situ hybridization assays showed that OsNF-YB1 was localized to the aleurone layer, and Electrophoretic Mobility Shift Assays (EMSA) demonstrated that OsNF-YB1 bound directly to CCAAT box sequences located upstream of the *OsSUT1*, *OsSUT3*, and *OsSUT4* coding sequences (Fig. 1.4). Mutant phenotypes of defective grain-filling and a chalky endosperm were observed in plants with knockout or down-regulation of *OsNF-YB1*. As phloem unloading into developing cereal seeds follows an apoplasmic pathway (Bihmidine et al., 2013), the *Osnf-yb1* mutant plants displayed a phenotype similar to that observed in rice when *Cell Wall Invertase2 (OsCIN2)* is mutated or down-regulated (Wang et al., 2008). Another recent study showed that OsNF-YC12 directly binds to the promoter of *OsSUT1* using various ChIP and yeast one-hybrid assays (Xiong et al., 2019). OsNF-YC12 was also found to physically interact with OsNF-YB1, and it is proposed that the OsNF-YB1–OsNF-YC12

dimer regulates the expression of *OsSUTs* in the aleurone layer during the grain-filling stage (Fig. 1.4).

A recent study identified a different mechanism for the regulation of rice sugar transporter genes expression. Liu et al. identified a rice RNA Recognition Motif (OsRRM)-containing protein that interacted with messenger RNAs of sugar transporter genes affecting their expression levels (Liu et al., 2020). Since the full length OsRRM protein could not be expressed in *E.coli*, RNA EMSAs were performed with the truncated version of the protein containing both of the RNA recognition motifs, which showed that the OsRRM protein bound to *OsSUT2*, *OsTMT1*, and *OsTMT2* RNAs. Though OsRMM was originally described as a gene specifically expressed in the endosperm, the authors reported higher expression in stems and leaves, based on qRT-PCR data. Additionally, they observed expression in almost all tissues in a transgenic plant expressing the *GUS* reporter gene under the control of the native promoter, exons, introns, and downstream sequences. Interestingly, all three genes, *OsSUT2*, *OsTMT1* and OsTMT2, are involved in sucrose regulation into and out of vacuoles, and it will be fascinating to explore if OsRRM regulates other vacuolar functions along with sugar transport.

Adding up these findings, the control of *SUTs* and *TSTs* expression in rice is very dynamic and involves multiple transcription factors and RNA binding proteins in different tissues to fine tune sugar homeostasis (Fig. 1.4). Further investigations are needed to assess similar mechanisms in other crop species.

Modulation of SUTs in Arabidopsis

Arabidopsis has been a model plant species for many decades, and numerous studies have explored its sugar transporters and their functions (Chandran et al., 2003;

Gottwald et al., 2000; Srivastava et al., 2008; Xu and Liesche, 2021). Among the nine SUT family members in Arabidopsis, AtSUC2 is the best characterized SUT and is critical for sucrose phloem loading. However, we previously did not know much about its regulation. But new research from Xu et al. (2020) showed that AtSUC2 is regulated post-translationally via its protein turnover rate and phosphorylation state. Based on large-scale membrane protein interaction studies (Jones et al., 2014), Xu and colleagues confirmed that AtUBIQUITIN-CONJUGATING ENZYME34 (AtUBC34) and AtWALL ASSOCIATED KINASE LIKE8 (AtWAKL8) were interacting partners of AtSUC2 (Xu et al., 2020). AtUBC34 was found to trigger turnover of AtSUC2 in a light-dependent manner via ubiquitin-dependent degradation. In *Atubc34* mutants, increases in phloem loading, biomass, and yield were observed. On the contrary, in mutants of the other recently identified AtSUC2-interaction partner, *Atwakl8*, decreased phloem loading and growth were observed. Further, AtWAKL8 phosphorylates AtSUC2 *in vivo* and *in vitro* based on various phosphorylation assays.

To check whether a similar phosphorylation event may occur in other Arabidopsis SUTs, or another species, we searched for evidence that SUTs were phosphorylated in phosphoproteomic datasets of maize and Arabidopsis proteins (Durek et al., 2010; Heazlewood et al., 2008; Walley et al., 2016; Wang et al., 2017). We found multiple phosphorylation sites based on Mass Spectrometry (MS) and other experimental evidence in ZmSUT1 in maize, and AtSUC1, AtSUC5, and AtSUC9 in Arabidopsis (Supplementary Table S1.3). Interestingly, in the PhosphAt Database, we also found single serine residues that have been experimentally determined to be phosphorylated in AtSUC4, AtSUC6, AtSUC7, and AtSUC8, but surprisingly, not AtSUC2 (Durek et al.,

2010; Heazlewood et al., 2008). The fact that these SUTs are phosphorylated *in vivo* does not mean that these phosphorylation events regulate protein activity or stability, but they provide starting points to investigate whether the functions of these sugar transporters are controlled by protein phosphorylation.

Intron control of SUT tissue-specific expression

To study gene expression in tissues and at the cellular level, many studies have used the native gene promoter to drive expression of a reporter gene under the control of the endogenous promoter, omitting the introns. However, these constructs sometimes do not fully recapitulate the expression of the endogenous gene (Sieburth and Meyerowitz, 1997). Relatedly, Lasin et al. (2020) highlighted the importance of introns in controlling the expression of the *AtSUC1* gene. *AtSUC1* is expressed in roots, pollen, and trichomes. However, the authors found that root expression is controlled by an interaction between the *AtSUC1* promoter and two short introns. Without the introns, no expression was observed in roots (Lasin et al., 2020). Similarly, in tomato (*Solanum lycopersicum*), the presence of an intron in a transgene sequence of *SlSUT1* (formerly *LeSUT1*) determined whether the gene was expressed in trichomes, or CC and guard cells (Weise et al., 2008). These studies illuminate the importance of intronic sequences in regulating the spatial expression of *SUT* genes, potentially raising questions about previous conclusions based only on promoter: *GUS* reporter constructs, for example. Hence, for an accurate characterization of gene expression *in planta*, intronic sequences must be included in creating transgenic plants.

Overall, the above findings in Arabidopsis and rice present novel discoveries for the regulation of sugar transporters (e.g., SUTs, TSTs and SWEETs) in plants. It is likely

that conserved processes control the expression of sugar transporters in other plant species. Understanding how these sugar transporters are regulated will enable us to develop more efficient and vigorous crops by modifying the functions of these genes.

Roles of sugar transporters in storage organs and sink tissues

In many grass species, such as maize, rice, and wheat, carbohydrates are principally partitioned to, and stored as, starch in seeds. In fact, by overexpressing the Arabidopsis *SUC2* gene under a phloem-specific promoter in rice, grain yields were improved (Wang et al., 2015). Additionally, there are many enzymes and proteins involved in converting sucrose into starch and affecting sink strength. For details on different enzymes and proteins involved in these processes, please see (Pfister and Zeeman, 2016; Ruan, 2014; Smith and Zeeman, 2020). Furthermore, there are vacuolar-localized INV that cleave sucrose and affect the availability of sucrose for storage in the vacuole (Ruan, 2014). However, other grasses, such as sweet sorghum and sugarcane, partition the majority of their phloem-translocated sucrose to the stem for storage (Bihmidine et al., 2013). Here, we will focus on the mechanisms by which TSTs and SUTs function to transport sucrose in sink tissues in both monocot and eudicot species.

In sugar beet taproots, sweet sorghum stems, and sugarcane stems, TSTs are hypothesized to be responsible for vacuolar sucrose accumulation (Table 1.1). BvTST2.1, the key transporter responsible for sucrose accumulation in the storage roots of sugar beet, exhibits sucrose: H^+ antiport activity with substrate preference for sucrose over glucose (Jung et al., 2015). In sweet sorghum, SbTST1 and SbTST2 were proposed to be responsible for the high levels of sugar accumulation in stem tissues based on gene expression levels in mature leaves and stems (Bihmidine et al., 2016). In another study

looking at the expression levels of SUTs and TSTs in sorghum stem internodes at various developmental phases, SbTST1 and SbTST2 were upregulated in the mature internodes, in agreement with previous findings (Milne et al., 2016). On the other hand, SbSUT4 and SbSUT5, were highly expressed in the transition internodes and had reduced expression levels in mature internodes. Also, SbSUT1 was found to be expressed in transition and mature internodes, where it could function in uptake of sucrose into post-phloem cells, with sucrose subsequently concentrated in storage vacuoles by TSTs (See Milne et al., 2016, Fig. 7 for details). Likewise, in a sugarcane hybrid, ShPST2a and ShPST2b genes, which are orthologous to SbTST2 and SbTST1, respectively, are strongly expressed in storage parenchyma cells of maturing stalk internodes of sugarcane. These proteins are hypothesized to localize to the tonoplast where they are proposed to function in loading sucrose into the vacuole (Casu et al., 2015).

The functions of TSTs in sugar storage organs extends to tissues beyond the three main sugar crops. Cheng et al. (2018) showed that *CmTST2* was most highly expressed during melon (*Cucumis melo*) fruit development based on expression analysis using qRT-PCR and tissue-specific expression studies using promoter: *GUS* activity. The functionality of *CmTST2* was tested in strawberry (*Fragaria* × *ananassa*) and cucumber (*Cucumis sativus*) fruits, where overexpressing the gene by transient expression and stable transformation, respectively, increased sucrose, fructose, and glucose accumulation in the fruits (Cheng et al., 2018). Likewise, in watermelon (*Citrullus lanatus*), Ren et al. (2021) demonstrated that *ClTST2* plays a major role in storing sugar in vacuoles of fruit parenchyma cells, and they further validated this function by generating CRISPR knockouts of *ClTST2*, which showed reduced sugar levels compared to wild type (Ren et

al., 2021). Hence, the function of TSTs is conserved in many plant species and plays a major role in storing sugars in vacuoles of storage organs.

Regulation of TSTs in storage tissues

Many studies identified TSTs to be involved in sugar accumulation in vacuoles of storage tissues in multiple monocots and eudicots, yet very little is known about how TSTs are regulated. A recent study demonstrated that TST expression is surprisingly modulated by ERDL6, a tonoplast-localized H^+: glucose symporter, in apple (*Malus domestica*) and tomato fruits (Zhu et al., 2021). *MdTST1* and *MdTST2* genes, and the orthologous genes in tomato, are proposed to be upregulated by glucose exported from vacuoles by the MdERDL6-1 protein, which in turn leads to the import of sugars from the cytosol into the vacuoles by these TST proteins in fruits. However, it is not known how the cytoplasmic glucose signals are sensed and transmitted for modulating *MdTST1* and *MdTST2* gene expression. Additional studies are necessary to characterize the sugar signaling pathway regulating MdTST1 and MdTST2 activities and to determine whether ERDL6-like proteins function to control TST activity in other species.

In another study, the heterologous expression of the sugar beet vacuolar sucrose loader *BvTST2.1* in Arabidopsis mutants lacking TST1 and TST2 led to an increase in glucose and fructose levels but constant sucrose levels (Vu et al., 2020). To explore the cause of these surprising results, BvTST2.1 was transiently expressed in *Nicotiana benthamiana* leaves either with or without the Vacuolar Invertase (VI) inhibitor NbVIF along with *Agrobacterium* stains carrying the viral silencing suppressor gene P19. In leaves infiltrated with P19 and *BvTST2.1*, increased sucrose, glucose, and fructose levels were observed compared to the leaf infiltrated with just the P19 control plasmid.

However, in leaves infiltrated with P19, *BvTST2.1*, and *NbVIF*, sucrose levels were greatly increased, whereas glucose and fructose levels were similar to the control P19 levels. Furthermore, in Arabidopsis artificial microRNA lines with decreased levels of mRNAs for VI1 and VI2, the monosaccharide levels were decreased whereas sucrose levels were increased. Additionally, when BvSUC4, the sucrose transporter localized to the vacuole that functions to export sucrose to the cytosol was expressed in Arabidopsis plants, the monosaccharide levels were decreased significantly, but not sucrose. Summing up, the sucrose storage in the vacuoles of leaf mesophyll cells is dependent on the net activity of VIs that break down sucrose into monosaccharides and TSTs that move sucrose and monosaccharides into the vacuole. Follow up studies are required to investigate if these genes function similarly to maintain vacuolar sucrose levels in sink tissues. If similar processes are the primary regulators of sucrose levels in leaves and sink tissues of other plant species, it is plausible that the monosaccharide levels maintained by the activity of VIs, ERDL6, and SWEET transporters regulate the activity of TSTs and sucrose levels in vacuoles in a very dynamic manner.

Post-translational modification was also recently shown to regulate the activity of TSTs. Deng et al. (2020) demonstrated a calcium sensor and protein kinase function to control sugar homeostasis through modulating the activity of GhTST2 in cotton (*Gossypium hirsutum*). Calcium is well known to play important roles in plant development and environmental stress as a secondary messenger, however, little is known about its involvement in regulation of sugar levels (Hunter, 2020). Calcineurin B-like protein (GhCBL2) and a CBL-interacting protein kinase (GhCIPK6) were found to control vacuolar glucose transport. GhCIPK6 is proposed to be recruited to the tonoplast

by GhCBL2, where it promotes the phosphorylation of GhTST2 and enhances its glucose transport activity (Deng et al., 2020). Similarly, in Arabidopsis, it was observed that a mitogen-activated triple kinase-like protein kinase (VIK) interacts with and phosphorylates AtTMT1, which stimulates import of glucose into the vacuole (Wingenter et al., 2011).

Based on the observations from phosphorylation studies, we investigated whether phosphorylation sites are present in TSTs of maize, rice, barley, and Arabidopsis (Durek et al., 2010; Endler et al., 2009; Heazlewood et al., 2008; Walley et al., 2016; Wang et al., 2017). We found multiple phosphorylation sites (based on MS and experimental data) in ZmTST1 and ZmTST2 in maize, OsTMT1 and OsTMT2 in rice, and AtTMT1, AtTMT2, and AtTMT3 in Arabidopsis (Supplementary Table S1.3). Excitingly, in AtTMT2, HvTST2 (also known as HvSTP previously) (Endler et al., 2009; Hunter, 2020), and OsTMT2, we found that the exact serine phosphorylation site that regulates the glucose transport activity of GhTST2 is phosphorylated *in planta*. Hence, it is possible that a similar phosphorylation cascade, potentially mediated by a conserved CBL-CIPK signaling pathway, controls TST transport activities for AtTMT2, OsTMT2 and possibly other TSTs in additional plant species. Characterizing mutants of the orthologous genes and their molecular and biochemical functions will help test this hypothesis.

Conclusions and future directions

Our knowledge of how sugars, the primary sources of energy for plant growth and development, are partitioned at the molecular level in source and sink tissues is still rudimentary. Currently, we have limited understanding of how sugar transporters are

regulated and more research is necessary to fill in this gap, especially in cereals like rice and maize that serve as primary staples for much of the world. Knowledge of how various sugar transporters are regulated and how they function in source and storage tissues will be foundational to enable us to efficiently modify their functions to improve crop yield and nutrition to feed the growing human population.

Several exciting breakthroughs have been made in recent years regarding sugar transporters and their regulation. With the widespread adoption of new technologies, e.g., single cell RNA-sequencing, we anticipate the identification of new genes specifically expressed within distinct cell-types in the phloem (Bezrutczyk et al., 2021; Kim et al., 2021a). This expected trove of new data will enable a deeper understanding of how these various genes and any other key players function in concert to regulate sugar transport in a cell-specific manner. Furthermore, with the advent of advanced microscopy techniques and instruments, we are able to better visualize the structure of cells and observe subcellular processes in finer detail (Braybrook, 2015; Knoblauch and Oparka, 2012; Komis et al., 2018; Ovečka et al., 2018). The integration of functional, structural, and genetic analyses of sugar transporters will be key to improving sugar delivery and storage in crops.

As recent studies begin to uncover the mechanisms of regulation of various sugar transporters by transcriptional, post-transcriptional and post-translational changes, identification of more interacting proteins, e.g., protein kinases and phosphatases that control phosphorylation of SUTs and TSTs, will help us better understand how their activities are regulated. Additionally, by altering promoter elements or recognition motifs identified by transcription factors such as OsDOF11, OsNF-YB1, and OsNF-YC12, or

RNA binding proteins such as OsRRM, we might be able to fine tune expression of various SUTs and SWEETs to develop more productive crops. Importantly, combining this approach with the modification of *SWEET* promotor regions to confer resistance to pathogens could help develop disease resistant and high yielding crops. As we discover more mechanisms regulating these sugar transporters, we gain new avenues towards crop improvement.

Acknowledgements

We thank two anonymous reviewers for valuable comments that improved the manuscript, and the members of the Braun lab for helpful discussions of papers cited. We also thank the editors for organizing this special issue and inviting us to contribute. No conflict of interest is declared.

Funding

This work was supported by a US National Science Foundation Plant Genome Research Program grant (IOS-1444448) and a US Department of Energy, Office of Science, Office of Biological and Environmental Research grant (DE-SC0018072) to DMB.

References

Aoki, N., Hirose, T., Takahashi, S., Ono, K., Ishimaru, K., Ohsugi, R., 1999. Molecular cloning and expression analysis of a gene for a sucrose transporter in Maize (*Zea mays* L.). Plant Cell Physiol. 40(10), 1072-1078.

Bai, A.-N., Lu, X.-D., Li, D.-Q., Liu, J.-X., Liu, C.-M., 2016. NF-YB1-regulated expression of sucrose transporters in aleurone facilitates sugar loading to rice endosperm. Cell Res. 26(3), 384-388.

Baker, R.F., Leach, K.A., Boyer, N.R., Swyers, M.J., Benitez-Alfonso, Y., Skopelitis, T., Luo, A., Sylvester, A., Jackson, D., Braun, D.M., 2016. Sucrose transporter *ZmSut1* expression and localization uncover new insights into sucrose phloem loading. Plant Physiol. 172(3), 1876-1898.

Bezrutczyk, M., Hartwig, T., Horschman, M., Char, S.N., Yang, J., Yang, B., Frommer, W.B., Sosso, D., 2018. Impaired phloem loading in *zmsweet13a,b,c* sucrose transporter triple knock-out mutants in *Zea mays*. New Phytol. 218(2), 594-603.

Bezrutczyk, M., Zöllner, N.R., Kruse, C.P.S., Hartwig, T., Lautwein, T., Köhrer, K., Frommer, W.B., Kim, J.-Y., 2021. Evidence for phloem loading via the abaxial bundle sheath cells in maize leaves. Plant Cell 33(3), 531-547.

Bihmidine, S., Baker, R.F., Hoffner, C., Braun, D.M., 2015. Sucrose accumulation in sweet sorghum stems occurs by apoplasmic phloem unloading and does not involve differential *Sucrose transporter* expression. BMC Plant Biol. 15(1), 186.

Bihmidine, S., Hunter, C.T., 3rd, Johns, C.E., Koch, K.E., Braun, D.M., 2013. Regulation of assimilate import into sink organs: update on molecular drivers of sink strength. Front. Plant Sci. 4, 177-177.

Bihmidine, S., Julius, B.T., Dweikat, I., Braun, D.M., 2016. *Tonoplast Sugar Transporters (SbTSTs)* putatively control sucrose accumulation in sweet sorghum stems. Plant Signal. Behav. 11(1), e1117721-e1117721.

Blanvillain-Baufumé, S., Reschke, M., Solé, M., Auguy, F., Doucoure, H., Szurek, B., Meynard, D., Portefaix, M., Cunnac, S., Guiderdoni, E., Boch, J., Koebnik, R., 2017. Targeted promoter editing for rice resistance to *Xanthomonas oryzae* pv. *oryzae* reveals differential activities for *SWEET14*-inducing TAL effectors. Plant Biotechnol. J. 15(3), 306-317.

Braun, D.M., Slewinski, T.L., 2009. Genetic control of carbon partitioning in grasses: roles of *Sucrose Transporters* and *Tie-dyed* loci in phloem loading. Plant Physiol. 149(1), 71-81.

Braun, D.M., Wang, L., Ruan, Y.-L., 2014. Understanding and manipulating sucrose phloem loading, unloading, metabolism, and signalling to enhance crop yield and food security. J. Exp. Bot. 65(7), 1713-1735.

Braybrook, S.A., 2015. Chapter 13 - Measuring the elasticity of plant cells with atomic force microscopy, in: Paluch, E.K. (Ed.) Methods Cell Biol. Academic Press, pp. 237-254.

Casu, R.E., Rae, A.L., Nielsen, J.M., Perroux, J.M., Bonnett, G.D., Manners, J.M., 2015. Tissue-specific transcriptome analysis within the maturing sugarcane stalk reveals spatial regulation in the expression of cellulose synthase and sucrose transporter gene families. Plant Mol. Biol. 89(6), 607-628.

Chandran, D., Reinders, A., Ward, J.M., 2003. Substrate specificity of the *Arabidopsis thaliana* sucrose transporter AtSUC2. J. Biol. Chem 278(45), 44320-44325.

Chen, L.-Q., Qu, X.-Q., Hou, B.-H., Sosso, D., Osorio, S., Fernie, A.R., Frommer, W.B., 2012. Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. Science 335(6065), 207-211.

Cheng, J., Wen, S., Xiao, S., Lu, B., Ma, M., Bie, Z., 2018. Overexpression of the tonoplast sugar transporter CmTST2 in melon fruit increases sugar accumulation. J. Exp. Bot. 69(3), 511-523.

Chincinska, I., Gier, K., Krügel, U., Liesche, J., He, H., Grimm, B., Harren, F.J.M., Cristescu, S.M., Kühn, C., 2013. Photoperiodic regulation of the sucrose transporter StSUT4 affects the expression of circadian-regulated genes and ethylene production. Front. Plant Sci. 4, 26-26.

Cho, J.I., Burla, B., Lee, D.W., Ryoo, N., Hong, S.K., Kim, H.B., Eom, J.S., Choi, S.B., Cho, M.H., Bhoo, S.H., 2010. Expression analysis and functional characterization of the monosaccharide transporters, *OsTMTs*, involving vacuolar sugar transport in rice (*Oryza sativa*). New Phytol. 186(3), 657-668.

Deng, J., Yang, X., Sun, W., Miao, Y., He, L., Zhang, X., 2020. The calcium sensor CBL2 and its interacting kinase CIPK6 are involved in plant sugar homeostasis via interacting with tonoplast sugar transporter TST2. Plant Physiol. 183(1), 236-249.

Durek, P., Schmidt, R., Heazlewood, J.L., Jones, A., MacLean, D., Nagel, A., Kersten, B., Schulze, W.X., 2010. PhosPhAt: the *Arabidopsis thaliana* phosphorylation site database. An update. Nucleic Acids Res. 38(Database issue), D828-834.

Endler, A., Meyer, S., Schelbert, S., Schneider, T., Weschke, W., Peters, S.W., Keller, F., Baginsky, S., Martinoia, E., Schmidt, U.G., 2006. Identification of a vacuolar

sucrose transporter in barley and Arabidopsis mesophyll cells by a tonoplast proteomic approach. Plant Physiol. 141(1), 196-207.

Endler, A., Reiland, S., Gerrits, B., Schmidt, U.G., Baginsky, S., Martinoia, E., 2009. *In vivo* phosphorylation sites of barley tonoplast proteins identified by a phosphoproteomic approach. Proteomics 9(2), 310-321.

Eom, J.-S., Chen, L.-Q., Sosso, D., Julius, B., Lin, W., Qu, X.-Q., Braun, D., Frommer, W., 2015. SWEETs, transporters for intracellular and intercellular sugar translocation. Curr. Opin. Plant Biol. 25.

Eom, J.-S., Cho, J.-I., Reinders, A., Lee, S.-W., Yoo, Y., Tuan, P.Q., Choi, S.-B., Bang, G., Park, Y.-I., Cho, M.-H., Bhoo, S.H., An, G., Hahn, T.-R., Ward, J.M., Jeon, J.-S., 2011. Impaired function of the tonoplast-localized sucrose transporter in rice, *OsSUT2*, limits the transport of vacuolar reserve sucrose and affects plant growth. Plant Physiol. 157(1), 109-119.

Eom, J.-S., Choi, S.-B., Ward, J.M., Jeon, J.-S., 2012. The mechanism of phloem loading in rice (*Oryza sativa*). Mol. Cells 33(5), 431-438.

Eom, J.-S., Nguyen, C.D., Lee, D.-W., Lee, S.-K., Jeon, J.-S., 2016. Genetic complementation analysis of rice sucrose transporter genes in Arabidopsis *SUC2* mutant *atsuc2*. J. Plant Biol. 59(3), 231-237.

Esau, K., 1977. Anatomy of seed plants, 2nd ed. John Wiley and Sons, New York.

Evert, R., Eschrich, W., Heyser, W., 1978. Leaf structure in relation to solute transport and phloem loading in *Zea mays* L. Planta 138(3), 279-294.

Evert, R.F., Russin, W.A., Botha, C., 1996. Distribution and frequency of plasmodesmata in relation to photoassimilate pathways and phloem loading in the barley leaf. Planta 198(4), 572-579.

Fukumorita, T., Chino, M., 1982. Sugar, amino acid and inorganic contents in rice phloem sap. Plant Cell Physiol. 23(2), 273-283.

Gamalei, Y., 1989. Structure and function of leaf minor veins in trees and herbs. Trees 3(2), 96-110.

Geiger, D.R., Giaquinta, R.T., Sovonick, S.A., Fellows, R.J., 1973. Solute distribution in sugar beet leaves in relation to phloem loading and translocation. Plant Physiol. 52(6), 585-589.

Gottwald, J.R., Krysan, P.J., Young, J.C., Evert, R.F., Sussman, M.R., 2000. Genetic evidence for the *in planta* role of phloem-specific plasma membrane sucrose transporters. Proc. Natl. Acad. Sci. USA 97(25), 13979-13984.

Haritatos, E., Medville, R., Turgeon, R., 2000. Minor vein structure and sugar transport in *Arabidopsis thaliana*. Planta 211(1), 105-111.

Hayashi, H., Chino, M., 1990. Chemical composition of phloem sap from the uppermost internode of the rice plant. Plant Cell Physiol. 31(2), 247-251.

Heazlewood, J.L., Durek, P., Hummel, J., Selbig, J., Weckwerth, W., Walther, D., Schulze, W.X., 2008. PhosPhAt: a database of phosphorylation sites in *Arabidopsis thaliana* and a plant-specific phosphorylation site predictor. Nucleic Acids Res. 36(Database issue), D1015-1021.

Hirose, T., Imaizumi, N., Scofield, G.N., Furbank, R.T., Ohsugi, R., 1997. cDNA cloning and tissue specific expression of a gene for sucrose transporter from Rice (*Oryza sativa* L.). Plant Cell Physiol. 38(12), 1389-1396.

Hirose, T., Zhang, Z., Miyao, A., Hirochika, H., Ohsugi, R., Terao, T., 2010. Disruption of a gene for rice sucrose transporter, *OsSUT1*, impairs pollen function but pollen maturation is unaffected. J. Exp. Bot. 61(13), 3639-3646.

Hunter, K., 2020. CBL2-CIPK6-TST2-mediated regulation of sugar homeostasis. Plant Physiol. 183(1), 21-22.

Ishimaru, K., Hirose, T., Aoki, N., Takahashi, S., Ono, K., Yamamoto, S., Wu, J., Saji, S., Baba, T., Ugaki, M., Matsumoto, T., Ohsugi, R., 2001. Antisense expression of a rice sucrose transporter OsSUT1 in Rice (*Oryza sativa* L.). Plant Cell Physiol. 42(10), 1181-1185.

Jones, A.M., Xuan, Y., Xu, M., Wang, R.-S., Ho, C.-H., Lalonde, S., You, C.H., Sardi, M.I., Parsa, S.A., Smith-Valle, E., Su, T., Frazer, K.A., Pilot, G., Pratelli, R., Grossmann, G., Acharya, B.R., Hu, H.-C., Engineer, C., Villiers, F., Ju, C., Takeda, K., Su, Z., Dong, Q., Assmann, S.M., Chen, J., Kwak, J.M., Schroeder, J.I., Albert, R., Rhee, S.Y., Frommer, W.B., 2014. Border control—a membrane-linked interactome of *Arabidopsis*. Science 344(6185), 711-716.

Julius, B.T., Leach, K.A., Tran, T.M., Mertz, R.A., Braun, D.M., 2017. Sugar transporters in plants: New insights and discoveries. Plant Cell Physiol. 58(9), 1442-1460.

Jung, B., Ludewig, F., Schulz, A., Meissner, G., Woestefeld, N., Fluegge, U.-I., Pommerrenig, B., Wirsching, P., Sauer, N., Koch, W., 2015. Identification of the

transporter responsible for sucrose accumulation in sugar beet taproots. Nature Plants 1(1), 1-6.

Kim, J.-Y., Symeonidi, E., Pang, T.Y., Denyer, T., Weidauer, D., Bezrutczyk, M., Miras, M., Zöllner, N., Hartwig, T., Wudick, M.M., Lercher, M., Chen, L.-Q., Timmermans, M.C.P., Frommer, W.B., 2021a. Distinct identities of leaf phloem cells revealed by single cell transcriptomics. Plant Cell 33(3), 511-530.

Kim, P., Xue, C.Y., Song, H.D., Gao, Y., Feng, L., Li, Y., Xuan, Y.H., 2021b. Tissue-specific activation of *DOF11* promotes rice resistance to sheath blight disease and increases grain weight via activation of *SWEET14*. Plant Biotechnol. J. 19(3), 409-411.

Knoblauch, M., Knoblauch, J., Mullendore, D.L., Savage, J.A., Babst, B.A., Beecher, S.D., Dodgen, A.C., Jensen, K.H., Holbrook, N.M., 2016. Testing the Münch hypothesis of long distance phloem transport in plants. eLife 5, e15341.

Knoblauch, M., Oparka, K., 2012. The structure of the phloem – still more questions than answers. Plant J. 70(1), 147-156.

Komis, G., Novák, D., Ovečka, M., Šamajová, O., Šamaj, J., 2018. Advances in imaging plant cell dynamics. Plant Physiol. 176(1), 80-93.

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35(6), 1547-1549.

Lalonde, S., Wipf, D., Frommer, W.B., 2004. Transport mechanisms for organic forms of carbon and nitrogen between source and sink. Annu. Rev. Plant Biol. 55, 341-372.

Lasin, P., Weise, A., Reinders, A., Ward, J.M., 2020. Arabidopsis Sucrose Transporter AtSuc1 introns act as strong enhancers of expression. Plant Cell Physiol. 61(6), 1054-1063.

Leach, K.A., Tran, T.M., Slewinski, T.L., Meeley, R.B., Braun, D.M., 2017. *Sucrose transporter2* contributes to maize growth, development, and crop yield. J. Integr. Plant Biol. 59(6), 390-408.

Lemoine, R., La Camera, S., Atanassova, R., Dédaldéchamp, F., Allario, T., Pourtau, N., Bonnemain, J.-L., Laloi, M., Coutos-Thévenot, P., Maurousset, L., Faucher, M., Girousse, C., Lemonnier, P., Parrilla, J., Durand, M., 2013. Source-to-sink transport of sugar and regulation by environmental factors. Front. Plant Sci. 4(272).

Letunic, I., Bork, P., 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 49(W1), W293-W296.

Li, J., Blue, R., Zeitler, B., Strange, T.L., Pearl, J.R., Huizinga, D.H., Evans, S., Gregory, P.D., Urnov, F.D., Petolino, J.F., 2013. Activation domains for controlling plant gene expression using designed transcription factors. Plant Biotechnol. J. 11(6), 671-680.

Liu, D., Xu, L., Wang, W., Jia, S., Jin, S., Gao, J., 2020. OsRRM, an RNA-binding protein, modulates sugar transport in rice (*Oryza sativa* L.). Front. Plant Sci. 11(1889).

Milne, R.J., Perroux, J.M., Rae, A.L., Reinders, A., Ward, J.M., Offler, C.E., Patrick, J.W., Grof, C.P.L., 2016. Sucrose transporter localization and function in phloem unloading in developing stems. Plant Physiol. 173(2), 1330-1341.

Münch, E., 1930. Die Stoffbewegungen in der Pflanze. Jena: Gustav Fischer.

Oliva, R., Ji, C., Atienza-Grande, G., Huguet-Tapia, J.C., Perez-Quintero, A., Li, T., Eom, J.-S., Li, C., Nguyen, H., Liu, B., Auguy, F., Sciallano, C., Luu, V.T., Dossa, G.S., Cunnac, S., Schmidt, S.M., Slamet-Loedin, I.H., Vera Cruz, C., Szurek, B., Frommer, W.B., White, F.F., Yang, B., 2019. Broad-spectrum resistance to bacterial blight in rice using genome editing. Nat. Biotechnol. 37(11), 1344-1350.

Ovečka, M., von Wangenheim, D., Tomančák, P., Šamajová, O., Komis, G., Šamaj, J., 2018. Multiscale imaging of plant development by light-sheet fluorescence microscopy. Nature Plants 4(9), 639-650.

Pfister, B., Zeeman, S.C., 2016. Formation of starch in plant cells. Cellular and molecular life sciences : CMLS 73(14), 2781-2807.

Ren, Y., Li, M., Guo, S., Sun, H., Zhao, J., Zhang, J., Liu, G., He, H., Tian, S., Yu, Y., Gong, G., Zhang, H., Zhang, X., Alseekh, S., Fernie, A.R., Scheller, H.V., Xu, Y., 2021. Evolutionary gain of oligosaccharide hydrolysis and sugar transport enhanced carbohydrate partitioning in sweet watermelon fruits. Plant Cell.

Robinson-Beers, K., Evert, R.F., 1991. Ultrastructure of and plasmodesmatal frequency in mature leaves of sugarcane. Planta 184(3), 291-306.

Ruan, Y.-L., 2014. Sucrose metabolism: gateway to diverse carbon use and sugar signaling. Annu. Rev. Plant Biol. 65(1), 33-67.

Sauer, N., 2007. Molecular physiology of higher plant sucrose transporters. FEBS Lett. 581(12), 2309-2317.

Scofield, G.N., Hirose, T., Aoki, N., Furbank, R.T., 2007. Involvement of the sucrose transporter, OsSUT1, in the long-distance pathway for assimilate transport in rice. J. Exp. Bot. 58(12), 3155-3169.

Scofield, G.N., Hirose, T., Gaudron, J.A., Furbank, R.T., Upadhyaya, N.M., Ohsugi, R., 2002. Antisense suppression of the rice transporter gene, *OsSUT1*, leads to impaired grain filling and germination but does not affect photosynthesis. Funct. Plant Biol. 29(7), 815-826.

Sieburth, L.E., Meyerowitz, E.M., 1997. Molecular dissection of the *AGAMOUS* control region shows that cis elements for spatial regulation are located intragenically. Plant Cell 9(3), 355-365.

Singh, J., James, D., Murali Achary, V.M., Patel, M.K., Thakur, J.K., Reddy, M.K., Tripathy, B.C., 2021. Coordinated overexpression of *OsSUT1*, *OsSWEET11*, and *OsSWEET14* in rice impairs carbohydrate metabolism that has implications in plant growth, yield and susceptibility to *Xanthomonas oryzae pv oryzae (Xoo)*. bioRxiv, 2021.2001.2007.425507.

Slewinski, T.L., Braun, D.M., 2010. Current perspectives on the regulation of whole-plant carbohydrate partitioning. Plant Sci. 178(4), 341-349.

Slewinski, T.L., Garg, A., Johal, G.S., Braun, D.M., 2010. Maize *SUT1* functions in phloem loading. Plant Signal. Behav. 5(6), 687-690.

Slewinski, T.L., Meeley, R., Braun, D.M., 2009. *Sucrose transporter1* functions in phloem loading in maize leaves. J. Exp. Bot. 60(3), 881-892.

Smith, A.M., Zeeman, S.C., 2020. Starch: a flexible, adaptable carbon store coupled to plant growth. Annu. Rev. Plant Biol. 71(1), 217-245.

Srivastava, A.C., Dasgupta, K., Ajieren, E., Costilla, G., McGarry, R.C., Ayre, B.G., 2009. *Arabidopsis* plants harbouring a mutation in *AtSUC2*, encoding the predominant sucrose/proton symporter necessary for efficient phloem transport, are able to complete their life cycle and produce viable seed. Ann. Bot. 104(6), 1121-1128.

Srivastava, A.C., Ganesan, S., Ismail, I.O., Ayre, B.G., 2008. Functional characterization of the Arabidopsis AtSUC2 sucrose/H+ symporter by tissue-specific complementation reveals an essential role in phloem loading but not in long-distance transport. Plant Physiol. 148(1), 200-211.

Stein, O., Granot, D., 2019. An overview of sucrose synthases in plants. Front. Plant Sci. 10(95).

Thompson, R., Dale, J., 1981. Export of 14C-and 11C-labelled assimilate from wheat and maize leaves: effects of parachloromercurobenzylsulphonic acid and fusicoccin and of potassium deficiency. Can. J. Bot./Rev. Can. Bot. 59(12), 2439-2444.

Vu, D.P., Martins Rodrigues, C., Jung, B., Meissner, G., Klemens, P.A.W., Holtgräwe, D., Fürtauer, L., Nägele, T., Nieberl, P., Pommerrenig, B., Neuhaus, H.E., 2020. Vacuolar sucrose homeostasis is critical for plant development, seed properties, and night-time survival in Arabidopsis. J. Exp. Bot. 71(16), 4930-4943.

Walley, J.W., Sartor, R.C., Shen, Z., Schmitz, R.J., Wu, K.J., Urich, M.A., Nery, J.R., Smith, L.G., Schnable, J.C., Ecker, J.R., Briggs, S.P., 2016. Integration of omic networks in a developmental atlas of maize. Science 353(6301), 814-818.

Wang, E., Wang, J., Zhu, X., Hao, W., Wang, L., Li, Q., Zhang, L., He, W., Lu, B., Lin, H., Ma, H., Zhang, G., He, Z., 2008. Control of rice grain-filling and yield by a gene with a potential signature of domestication. Nat. Genet. 40(11), 1370-1374.

Wang, G., Wu, Y., Ma, L., Lin, Y., Hu, Y., Li, M., Li, W., Ding, Y., Chen, L., 2021. Phloem loading in rice leaves depends strongly on the apoplastic pathway. J. Exp. Bot. 72(10), 3723-3738.

Wang, L., Lu, Q., Wen, X., Lu, C., 2015. Enhanced sucrose loading improves rice yield by increasing grain size Plant Physiol. 169(4), 2848-2862.

Wang, Y., Tong, X., Qiu, J., Li, Z., Zhao, J., Hou, Y., Tang, L., Zhang, J., 2017. A phosphoproteomic landscape of rice (*Oryza sativa*) tissues. Physiol. Plant. 160(4), 458-475.

Weise, A., Lalonde, S., Kühn, C., Frommer, W.B., Ward, J.M., 2008. Introns control expression of sucrose transporter LeSUT1 in trichomes, companion cells and in guard cells. Plant Mol. Biol. 68(3), 251-262.

Williams, M., Farrar, J., Pollock, C., 1989. Cell specialization within the parenchymatous bundle sheath of barley. Plant, Cell Environ. 12(9), 909-918.

Wingenter, K., Schulz, A., Wormit, A., Wic, S., Trentmann, O., Hoermiller, I.I., Heyer, A.G., Marten, I., Hedrich, R., Neuhaus, H.E., 2010. Increased activity of the vacuolar monosaccharide transporter TMT1 alters cellular sugar partitioning, sugar signaling, and seed yield in Arabidopsis. Plant Physiol. 154(2), 665-677.

Wingenter, K., Trentmann, O., Winschuh, I., Hörmiller, I.I., Heyer, A.G., Reinders, J., Schulz, A., Geiger, D., Hedrich, R., Neuhaus, H.E., 2011. A member of the

mitogen-activated protein 3-kinase family is involved in the regulation of plant vacuolar glucose uptake. Plant J. 68(5), 890-900.

Wormit, A., Trentmann, O., Feifer, I., Lohr, C., Tjaden, J., Meyer, S., Schmidt, U., Martinoia, E., Neuhaus, H.E., 2006. Molecular identification and physiological characterization of a novel monosaccharide transporter from Arabidopsis involved in vacuolar sugar transport. Plant Cell 18(12), 3476-3490.

Wu, Y., Lee, S.-K., Yoo, Y., Wei, J., Kwon, S.-Y., Lee, S.-W., Jeon, J.-S., An, G., 2018. Rice transcription factor OsDOF11 modulates sugar transport by promoting expression of *Sucrose Transporter* and *SWEET* Genes. Mol. Plant 11(6), 833-845.

Xiong, Y., Ren, Y., Li, W., Wu, F., Yang, W., Huang, X., Yao, J., 2019. NF-YC12 is a key multi-functional regulator of accumulation of seed storage substances in rice. J. Exp. Bot. 70(15), 3765-3780.

Xu, Q., Liesche, J., 2021. Sugar export from Arabidopsis leaves: actors and regulatory strategies. J. Exp. Bot. 72(15), 5275-5284.

Xu, Q., Yin, S., Ma, Y., Song, M., Song, Y., Mu, S., Li, Y., Liu, X., Ren, Y., Gao, C., Chen, S., Liesche, J., 2020. Carbon export from leaves is controlled via ubiquitination and phosphorylation of sucrose transporter SUC2. Proc. Natl. Acad. Sci. USA 117(11), 6223-6230.

Zhu, L., Li, B., Wu, L., Li, H., Wang, Z., Wei, X., Ma, B., Zhang, Y., Ma, F., Ruan, Y.-L., Li, M., 2021. MdERDL6-mediated glucose efflux to the cytosol promotes sugar accumulation in the vacuole through up-regulating TSTs in apple and tomato. Proc. Natl. Acad. Sci. USA 118(1), e2022788118.

Zimmermann, M., Ziegler, H., 1975. List of sugars and sugar alcohols in sieve-tube exudates, in: Zimmermann, M., Milburn, H. (Eds.), Encyclopedia of Plant Physiology. Springer, New York, pp. 480–503.

TABLES

Table 1.1

Summary of *TST* genes in different sugar accumulating species and their functions

Species	**Gene Name**	**Function**	**References**
Beet roots	*BvTST2*	Sucrose accumulation in beet taproots	Jung et al. 2015
Sugarcane	*ShPST2a* and *ShPST2b*	Sucrose accumulation in maturing sugarcane stalks	Casu et al. 2015
Sorghum	*SbTST1* and *SbTST2*	Sugar accumulation in stems of sweet sorghum	Bihmidine et al. 2016
Melon*	*CmTST2*	Sucrose, glucose, and fructose accumulation in fruits	Chen et al. 2018
Cotton	*GhTST2*	Glucose accumulation in leaves of cotton	Deng et al. 2020
Watermelon	*ClTST2*	Sucrose, glucose, and fructose accumulation in watermelon fruits	Ren et al. 2021
Apple and Tomato	*MdTST1* and *MdTST2*	Sucrose, glucose, and fructose accumulation in fruits of tomatoes and apple leaves	Zhu et al. 2021

***** The *CmTST2* gene from melon was transiently expressed in strawberry fruit and cucumber as well.

Table S1. 1

List of SUTs and TSTs sequences used in the phylogeny

Species	Gene Name	Gene ID/ Genbank ID	Source
Arabidopsis thaliana	AtSUC1	AT1G71880.1	Arabidopsis TAIR
	AtSUC2	AT1G22710.1	
	AtSUC3	AT2G02860.1	
	AtSUC4	AT1G09960.1	
	AtSUC5	AT1G71890.1	
	AtSUC6	AT5G43610.1	
	AtSUC7	AT1G66570.1	
	AtSUC8	AT2G14670.1	
	AtSUC9	AT5G06170.1	
Galdieria sulphuraria	Gasu_08920	Gasu_08920	NCBI
Hordeum vulgare	HvSUT1	HORVU4Hr1G075200.3	Ensembl Plants (IBSC_v2)
	HvSUT2	HORVU5Hr1G000010.1	
	HvSUT3	HORVU1Hr1G035760.3	
	HvSUT4	HORVU6Hr1G093600.9	
	HvSUT5	HORVU2Hr1G112080.1	
Oryza sativa	OsSUT1	LOC_Os03g07480.2	MSU RGAP
	OsSUT2	LOC_Os12g44380.1	
	OsSUT3	LOC_Os10g26470.1	
	OsSUT4	LOC_Os02g58080.1	
	OsSUT5	LOC_Os02g36700.1	
Saccharum hybrid cultivar Q117	ShSUT1	AAV41028	NCBI
Sorghum bicolor	SbSUT1	Sobic.001G488700.1	Phytozome
	SbSUT2	Sobic.004G353600.1	
	SbSUT3	Sobic.001G254000.1	
	SbSUT4	Sobic.008G193300.1	
	SbSUT5	Sobic.004G190500.1	
	SbSUT6	Sobic.007G214500.1	
Zea mays	ZmSUT1	GRMZM2G034302_T01	MaizeGDB
	ZmSUT2	GRMZM2G307561_T03	
	ZmSUT3	GRMZM2G083248_T01	
	ZmSUT4	GRMZM2G145107_T01	
	ZmSUT5	GRMZM2G081589_T01	
	ZmSUT6	GRMZM2G106741_T01	
	ZmSUT7	GRMZM2G087901_T01	

Table S1. 2

Percentage identity matrices of TSTs in the phylogeny

	Gasu	OsTMT4	OsTMT3	MdTST5	OsTMT6	SbTST3	ZmTST3	OsTMT5	HvTST1	AtTMT3	AtTMT1	BvTST3	MdTST4	SlTMT1	SlTMT3	CmTST3	ClTST3	BvTST1	BvTST2.1	BvTST2.2	HvTST2	ZmTST1	OsTMT1	SbTST1	ShPST2a	OsTMT2	ZmTST2	SbTST2	AtTMT2	CmTST2	ClTST2	MdTST3	MdTST2	SlTMT2	CmTST1	ClTST1	MdTST1	GhTST2
Gasu	100.00	28.60	26.14	30.24	31.96	31.17	32.40	32.82	31.00	30.18	30.50	30.13	29.63	30.55	30.91	31.80	31.73	29.69	31.50	32.53	31.07	31.53	31.87	31.22	31.14	31.95	31.80	31.80	32.22	32.10	31.48	31.00	31.44	30.94	30.79	30.28	30.57	31.82
OsTMT4	28.60	100.00	40.55	42.57	46.21	45.20	44.41	46.27	45.40	42.65	43.23	45.00	44.31	44.43	43.97	43.52	43.98	44.92	42.68	43.69	46.92	45.69	46.07	46.07	45.09	46.17	46.01	46.32	45.79	45.23	45.30	44.22	43.76	46.06	45.81	45.96	46.07	45.69
OsTMT3	26.14	40.55	100.00	47.64	48.42	48.11	47.43	48.78	47.26	49.15	49.00	50.69	48.06	52.38	51.17	52.29	51.90	51.67	50.00	50.75	55.89	54.15	56.45	56.19	53.32	55.81	54.83	55.74	54.35	49.79	50.87	51.96	51.96	52.79	51.45	51.98	51.96	50.52
MdTST5	30.24	42.57	47.64	100.00	55.93	55.62	54.67	55.31	53.99	61.31	57.57	62.02	65.46	65.22	61.58	63.03	63.79	58.52	59.78	60.28	62.92	63.59	64.26	64.53	60.17	62.98	61.69	62.79	61.57	64.20	63.99	63.16	62.88	63.31	64.15	64.62	63.90	64.64
OsTMT6	31.96	46.21	48.42	55.93	100.00	78.26	74.28	72.59	70.83	60.43	60.31	66.19	62.53	64.65	63.36	62.18	62.94	63.33	60.14	61.61	66.11	65.98	66.07	66.76	62.86	67.31	64.65	66.57	63.62	64.76	64.95	63.80	63.11	62.02	63.31	63.22	62.60	62.93
SbTST3	31.17	45.20	48.11	55.62	78.26	100.00	82.28	70.09	69.58	61.52	59.44	65.00	63.57	64.07	63.33	64.30	65.20	62.31	62.43	63.75	65.88	65.15	66.39	66.94	63.40	68.00	65.19	67.40	64.50	65.17	65.18	63.79	63.23	62.76	64.08	64.28	62.78	63.14
ZmTST3	32.40	44.41	47.43	54.67	74.28	82.28	100.00	69.27	66.04	60.64	58.70	63.57	60.59	62.39	61.48	61.63	63.11	61.29	59.63	60.20	64.02	62.71	64.25	64.48	61.39	64.91	62.92	64.03	62.25	62.17	62.10	62.61	61.48	62.22	61.56	62.04	61.57	61.48
OsTMT5	32.82	46.27	48.78	55.31	72.59	70.09	69.27	100.00	73.87	59.90	59.37	62.92	61.87	63.28	62.34	61.39	62.40	60.85	59.59	60.28	62.97	63.51	64.15	64.79	61.54	65.05	63.42	64.99	61.05	62.95	62.66	61.44	60.82	60.60	61.87	62.24	60.22	62.26
HvTST1	31.00	45.40	47.26	53.99	70.83	69.58	66.04	73.87	100.00	58.27	58.63	60.82	60.03	61.21	61.84	59.68	60.32	60.64	57.45	57.96	61.32	61.41	61.24	61.69	58.62	60.98	58.48	60.34	60.15	60.93	61.01	60.46	59.60	59.19	61.11	60.96	59.54	59.97
AtTMT3	30.18	42.65	49.15	61.31	60.43	61.52	60.64	59.90	58.27	100.00	58.79	63.73	65.59	66.71	65.15	63.99	64.61	63.71	61.43	62.36	65.39	63.61	64.35	65.62	61.50	65.49	64.03	65.87	64.71	65.81	65.82	64.96	64.54	65.35	65.43	66.33	63.37	65.25
AtTMT1	30.50	43.23	49.00	57.57	60.31	59.44	58.70	59.37	58.63	58.79	100.00	63.69	63.06	64.60	63.33	61.61	62.66	63.83	61.62	61.34	63.74	62.26	62.57	62.85	61.58	64.68	63.25	64.77	66.34	65.31	65.73	64.62	63.51	65.37	68.86	68.63	65.42	67.82
BvTST3	30.13	45.00	50.69	62.02	66.19	65.00	63.57	62.92	60.82	63.73	63.69	100.00	69.15	71.17	70.91	68.75	69.22	65.44	64.51	65.27	67.83	66.76	66.81	67.65	65.08	69.60	67.32	69.41	67.99	69.14	68.96	67.09	65.97	68.44	68.21	68.64	67.51	68.58
MdTST4	29.63	44.31	48.06	65.46	62.53	63.57	60.59	61.87	60.03	65.59	63.06	69.15	100.00	71.43	70.69	68.55	70.41	65.37	63.84	66.39	68.83	68.55	68.88	69.85	64.15	69.00	66.07	67.99	67.46	65.79	67.65	68.51	68.37	67.58	67.87	68.47	67.77	67.75
SlTMT1	30.55	44.43	52.38	65.22	64.65	64.07	62.39	63.28	61.21	66.71	64.60	71.17	71.43	100.00	83.90	69.26	70.25	67.97	65.69	68.38	70.65	70.33	69.86	70.99	66.11	70.82	68.46	70.26	69.72	70.95	71.19	70.51	69.12	70.56	69.71	69.00	68.38	69.94
SlTMT3	30.91	43.97	51.17	61.58	63.36	63.33	61.48	62.34	61.84	65.15	63.33	70.91	70.69	83.90	100.00	69.33	70.35	66.95	64.33	68.34	70.00	69.53	69.61	70.63	66.48	69.89	67.33	69.32	68.98	71.02	71.84	70.33	68.33	69.84	69.74	69.45	67.29	69.09
CmTST3	31.80	43.52	52.29	63.03	62.18	64.30	61.63	61.39	59.68	63.99	61.61	68.75	68.55	69.26	69.33	100.00	94.34	64.01	62.32	65.23	66.67	67.45	65.79	67.83	64.73	69.71	66.80	69.16	66.48	69.92	69.74	67.97	67.55	67.92	65.73	66.57	65.69	68.01
ClTST3	31.73	43.98	51.90	63.79	62.94	65.20	63.11	62.40	60.32	64.61	62.66	69.22	70.41	70.25	70.35	94.34	100.00	64.34	63.87	66.25	67.41	67.50	67.04	68.57	65.98	70.17	67.36	69.57	68.08	71.07	70.89	69.12	68.43	68.94	66.76	67.74	66.57	69.03
BvTST1	29.69	44.92	51.67	58.52	63.33	62.31	61.29	60.85	60.64	63.71	63.83	65.44	65.37	67.97	66.95	64.01	64.34	100.00	62.79	64.65	67.63	67.77	66.90	68.69	63.10	68.36	65.57	67.22	69.74	70.07	69.48	67.54	66.85	71.43	70.99	71.27	72.02	72.23
BvTST2.1	31.50	42.68	50.00	59.78	60.14	62.43	59.63	59.59	57.45	61.43	61.62	64.51	63.84	65.69	64.33	62.32	63.87	62.79	100.00	84.56	67.91	67.31	68.23	68.51	63.75	68.13	65.66	67.03	70.29	70.30	70.26	69.09	68.13	68.73	67.73	68.15	66.85	66.67
BvTST2.2	32.53	43.69	50.75	60.28	61.61	63.75	60.20	60.28	57.96	62.36	61.34	65.27	66.39	68.38	68.34	65.23	66.25	64.65	84.56	100.00	69.18	68.36	69.64	69.41	64.94	70.98	68.49	69.85	71.39	72.78	72.19	71.04	70.49	69.32	70.26	70.26	67.99	68.21
HvTST2	31.07	46.92	55.89	62.92	66.11	65.88	64.02	62.97	61.32	65.39	63.74	67.83	68.83	70.65	70.00	66.67	67.41	67.63	67.91	69.18	100.00	87.43	90.53	90.26	72.53	77.57	73.48	74.97	73.16	72.78	73.29	72.74	71.23	71.19	70.54	70.82	71.86	71.35
ZmTST1	31.53	45.69	54.15	63.59	65.98	65.15	62.71	63.51	61.41	63.61	62.26	66.76	68.55	70.33	69.53	67.45	67.50	67.77	67.31	68.36	87.43	100.00	91.19	94.73	71.95	77.13	73.44	75.07	73.44	73.29	73.30	73.70	71.37	73.01	71.23	70.99	73.46	73.22
OsTMT1	31.87	46.07	56.45	64.26	66.07	66.39	64.25	64.15	61.24	64.35	62.57	66.81	68.88	69.86	69.61	65.79	67.04	66.90	68.23	69.64	90.53	91.19	100.00	95.12	72.26	77.67	74.97	76.32	73.82	74.18	74.42	74.49	72.15	73.35	71.47	71.61	73.56	72.50
SbTST1	31.22	46.07	56.19	64.53	66.76	66.94	64.48	64.79	61.69	65.62	62.85	67.65	69.85	70.99	70.63	67.83	68.57	68.69	68.51	69.41	90.26	94.73	95.12	100.00	73.00	78.46	75.58	76.93	74.79	74.35	74.86	74.21	71.74	73.76	72.71	72.85	73.56	73.32
ShPST2a	31.14	45.09	53.32	60.17	62.86	63.40	61.39	61.54	58.62	61.50	61.58	65.08	64.15	66.11	66.48	64.73	65.98	63.10	63.75	64.94	72.53	71.95	72.26	73.00	100.00	86.44	80.94	83.62	68.19	67.44	67.40	67.85	66.89	66.98	66.53	66.39	68.84	67.53
OsTMT2	31.95	46.17	55.81	62.98	67.31	68.00	64.91	65.05	60.98	65.49	64.68	69.60	69.00	70.82	69.89	69.71	70.17	68.36	68.13	70.98	77.57	77.13	77.67	78.46	86.44	100.00	88.19	91.68	73.48	73.09	73.19	72.52	72.11	72.75	71.94	71.66	74.46	73.68
ZmTST2	31.80	46.01	54.83	61.69	64.65	65.19	62.92	63.42	58.48	64.03	63.25	67.32	66.07	68.46	67.33	66.80	67.36	65.57	65.66	68.49	73.48	73.44	74.97	75.58	80.94	88.19	100.00	93.56	70.82	71.00	70.82	70.30	69.48	70.53	70.11	70.52	72.38	71.33
SbTST2	31.80	46.32	55.74	62.79	66.57	67.40	64.03	64.99	60.34	65.87	64.77	69.41	67.99	70.26	69.32	69.16	69.57	67.22	67.03	69.85	74.97	75.07	76.32	76.93	83.62	91.68	93.56	100.00	72.75	73.19	73.01	72.62	72.07	71.62	71.49	71.76	73.33	72.55
AtTMT2	32.22	45.79	54.35	61.57	63.62	64.50	62.25	61.05	60.15	64.71	66.34	67.99	67.46	69.72	68.98	66.48	68.08	69.74	70.29	71.39	73.16	73.44	73.82	74.79	68.19	73.48	70.82	72.75	100.00	78.16	77.86	75.14	74.72	72.75	71.75	72.17	72.99	73.20
CmTST2	32.10	45.23	49.79	64.20	64.76	65.17	62.17	62.95	60.93	65.81	65.31	69.14	65.79	70.95	71.02	69.92	71.07	70.07	70.30	72.78	72.78	73.29	74.18	74.35	67.44	73.09	71.00	73.19	78.16	100.00	93.64	80.57	78.93	75.00	73.31	74.27	75.07	75.20
ClTST2	31.48	45.30	50.87	63.99	64.95	65.18	62.10	62.66	61.01	65.82	65.73	68.96	67.65	71.19	71.84	69.74	70.89	69.48	70.26	72.19	73.29	73.30	74.42	74.86	67.40	73.19	70.82	73.01	77.86	93.64	100.00	80.27	78.49	74.59	73.55	74.27	74.21	75.44
MdTST3	31.00	44.22	51.96	63.16	63.80	63.79	62.61	61.44	60.46	64.96	64.62	67.09	68.51	70.51	70.33	67.97	69.12	67.54	69.09	71.04	72.74	73.70	74.49	74.21	67.85	72.52	70.30	72.62	75.14	80.57	80.27	100.00	90.35	75.10	72.14	72.55	75.00	74.39
MdTST2	31.44	43.76	51.96	62.88	63.11	63.23	61.48	60.82	59.60	64.54	63.51	65.97	68.37	69.12	68.33	67.55	68.43	66.85	68.13	70.49	71.23	71.37	72.15	71.74	66.89	72.11	69.48	72.07	74.72	78.93	78.49	90.35	100.00	73.19	70.62	70.62	73.22	72.89
SlTMT2	30.94	46.06	52.79	63.31	62.02	62.76	62.22	60.60	59.19	65.35	65.37	68.44	67.58	70.56	69.84	67.92	68.94	71.43	68.73	69.32	71.19	73.01	73.35	73.76	66.98	72.75	70.53	71.62	72.75	75.00	74.59	75.10	73.19	100.00	76.86	77.44	77.66	77.31
CmTST1	30.79	45.81	51.45	64.15	63.31	64.08	61.56	61.87	61.11	65.43	68.86	68.21	67.87	69.71	69.74	65.73	66.76	70.99	67.73	70.26	70.54	71.23	71.47	72.71	66.53	71.94	70.11	71.49	71.75	73.31	73.55	72.14	70.62	76.86	100.00	96.16	76.20	80.52
ClTST1	30.28	45.96	51.98	64.62	63.22	64.28	62.04	62.24	60.96	66.33	68.63	68.64	68.47	69.00	69.45	66.57	67.74	71.27	68.15	70.26	70.82	70.99	71.61	72.85	66.39	71.66	70.52	71.76	72.17	74.27	74.27	72.55	70.62	77.44	96.16	100.00	77.58	81.21
MdTST1	30.57	46.07	51.96	63.90	62.60	62.78	61.57	60.22	59.54	63.37	65.42	67.51	67.77	68.38	67.29	65.69	66.57	72.02	66.85	67.99	71.86	73.46	73.56	73.56	68.84	74.46	72.38	73.33	72.99	75.07	74.21	75.00	73.22	77.66	76.20	77.58	100.00	82.66
GhTST2	31.82	45.69	50.52	64.64	62.93	63.14	61.48	62.26	59.97	65.25	67.82	68.58	67.75	69.94	69.09	68.01	69.03	72.23	66.67	68.21	71.35	73.22	72.50	73.32	67.53	73.68	71.33	72.55	73.20	75.20	75.44	74.39	72.89	77.31	80.52	81.21	82.66	100.00

Table S1. 3

Summary of selected phosphorylated peptides identified in SUTs and TSTs of maize, Arabidopsis, rice, cotton, and barley

Species	Gene	Peptide Sequence	Reference
Maize	**ZmSUT1**	GDGELELsVGVR	Walley et al. 2016 data via MaizeGDB (https://maizegdb.org/)
		ANGEVETEPsGPLAVLK	
	ZMTST1	LYGPEEGLsWVARPVR	
		GQsALGsALGLIsR	
		HGsMAASQGKPLVDPMVTLFGsVHEK	
		MPEIMGsMR	
		stLFPNFGsMFsVADQQQVK	
		ADWDAEsQR	
		EGEDYAsDHGGDDIEDNLQsPLIsR	
		QAtsVEGK	
		EIAAPHGsILGAVGR	
	ZmTST2	EQITLYGPEEGQsWIARPSK	
		GPsMLGsVLsLAsR	
		stLFPNFGsMFsVTDQHAK	
		DDEEyAsDGAGGDyEDNLHsPLLsR	
		QAtGAEGKDIVHHGHR	
		RQsLLGEGGDGVSSTDIGGGWQLAWK	
		MsDAAMVHPsEVAAK	
	ZmTST3	DQVTLYGPEQGLsWVAQQVQGAR	
		SSVLGsAVGLAsR	
		QGsMYEQMKDPVVTLLGSVHDK	
		MPDSGAsAR	
		AsTLFPNLGsMLsVtER	
		HGGDWDEENVPPNDDLDDDEDEEEYLsDDEDAGAGAAAR	
		GGGGGGGALHAPLLsR	
		DGsHPPEsSPMQR	
		YSSItSGEAASTMGIGGGWQLAWK	
		MYLHEEGGGDGDsSDPAGGYVHAAALVSPSILYTK	
		DVLIGQSPtPAFDsPPPEtVANK	
Arabidopsis	**AtSUC1**	DAAALETQsPEDFDQPSPLRK	PhosPhAt (http://phosphat.uni-hohenheim.de)
		KtAGDLAGPSASVK	
	AtSUC5	AANNATALETQssPEDLGQPSPLR	
	AtSUC9	EVyGGDSAGDDKMK	
	AtTMT1	LyGtHENQsyLARPVPEQNssLGLR	
		HGsLANQsMILK	
		DPLVNLFGsLHEK	
		sGIFPHFGSM	
		DIESHYNKDNDDYAtDDGAGDDDDsDNDLR	
		QTtSMDKDMIPHPtsGstLsMR	
		YYLKEDGAEsR	
		GSIISIPGGPDGGGsYIHASALVSR	
		sVHGSAMVPPEK	
	AtTMT2	GGsTMSVLsRHGstMSRR	
		HEDWDEENLVGEGEDYPsDHGDDsEDDLHsPLIsR	
		IYLHQEGFPG**s**R	
Rice	**OsTMT1**	EGEDYGsDHGGDDIEDSLQsPLIsR	Wang et al, 2017

		EIAAPHGsIMGAVGR	
		MPEIMGsMR	
	OsTMT2	DDEEyAsDGAGGDYEDNVHsPLLSR	
		GsALsMR	
		IYLHQEEVPG**s**R	
Cotton	**GhTST2**	IYLHEEGIPG**s**R	Deng et al, 2020
Barley	**HvTST2**	IYLHEEGV**s**GDR	Endler et al, 2009

Amino acids in lower case denotes that the amino acid is phosphorylated. The underlined and bold serine residues in AtTMT2, HvTST2, and OsTMT2 peptide sequences denote similar phosphorylated peptides as identified in GhTST2.

FIGURES

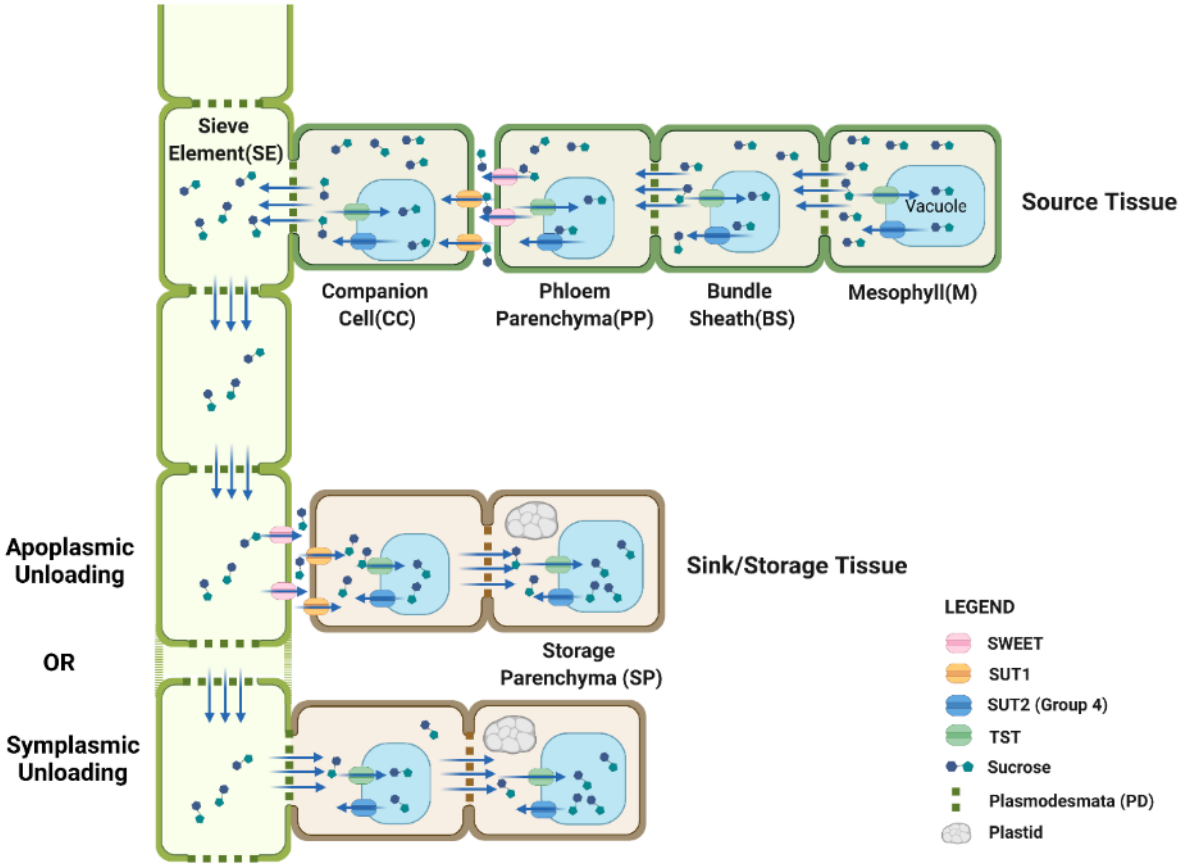


Figure 1.1

Path of sucrose transport in source and sink/storage tissues. (Top) In source tissues, apoplasmic phloem loading of sucrose is illustrated, where sucrose synthesized in Mesophyll cells (M), is mobilized to Bundle Sheath (BS) cells and to Phloem Parenchyma (PP) cells through plasmodesmata. Sucrose is then exported to the apoplasm by SWEETs and imported by SUT1 into the Companion Cell (CC) and ultimately moves into the sieve tube for delivery to other tissues. (Bottom) In sink/storage tissues, sucrose is unloaded from the sieve tube into the parenchyma cells either apoplasmically or symplasmically, depending on the location, age, and type of tissue. In symplasmic

unloading, sucrose flows through plasmodesmata down its concentration gradient into adjacent cells until it reaches the storage parenchyma cells (SP), where it is imported into vacuoles by TSTs. In apoplasmic phloem unloading, sucrose is likely exported from the sieve tube by SWEETs and taken up into adjacent cells by SUTs. Sucrose then moves symplasmically through plasmodesmata into nearby parenchyma cells and is transported into vacuoles by TSTs. Additionally, in many grain crops, sucrose can also be converted into hexoses and stored as starch in plastids. In many cell types, SUT2, a sugar transporter localized to the tonoplast functions to export sucrose from the vacuole.

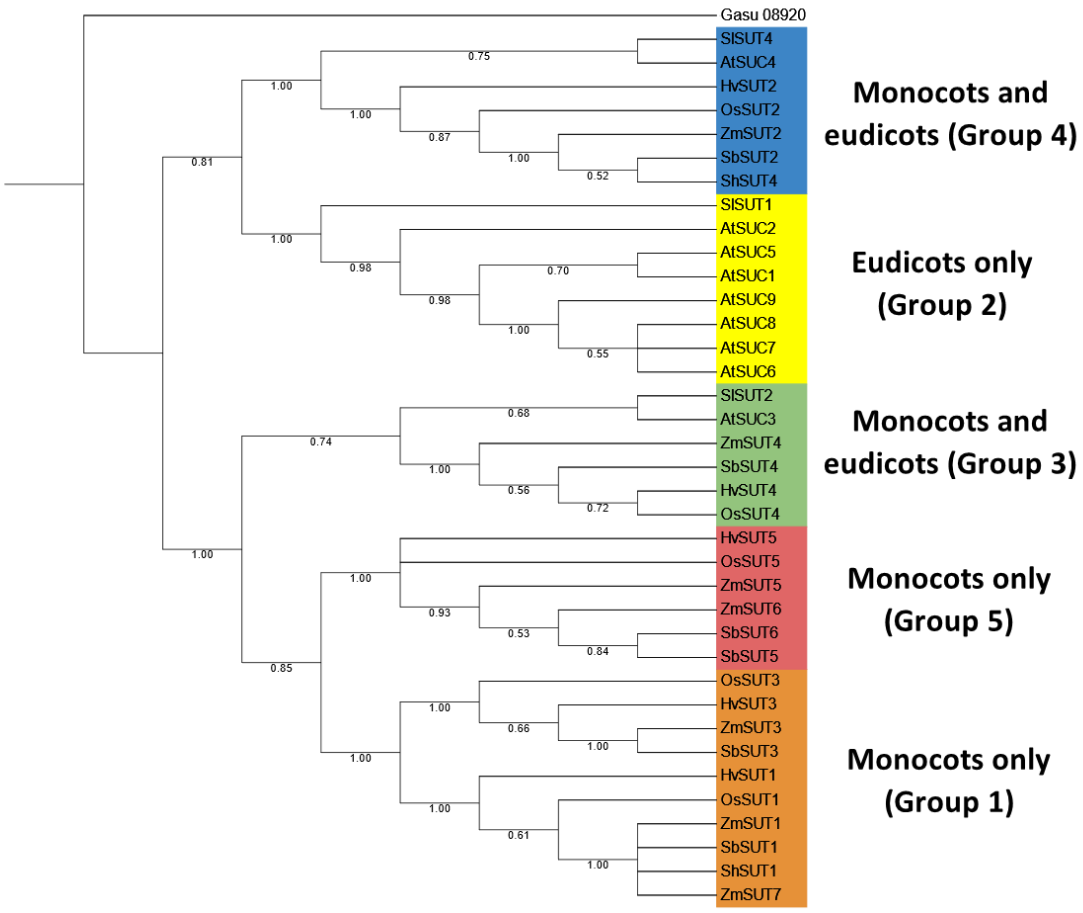


Figure 1.2

Phylogenetic tree of SUTs in select monocot and eudicot plants. Species shown are *Arabidopsis thaliana*, *Hordeum vulgare*, *Oryza sativa*, *Saccharum* hybrid, *Solanum lycopersicum*, *Sorghum bicolor*, *Zea mays*, and a red alga *Galdieria sulphuraria* (as the outgroup). SUTs can be divided into five groups (clades), with two groups containing SUT protein sequences from monocots only (salmon and orange), one group containing SUT sequences from eudicots only (yellow), and two groups containing SUT sequences from both monocots and eudicots (blue and green). The phylogenetic tree was

constructed in MEGA X (Kumar et al., 2018) using the Maximum Likelihood method. The branch support values denote the reliability for each internal branch based on 1000 bootstraps. The tree was visualized in interactive Tree of Life (Letunic and Bork, 2021).

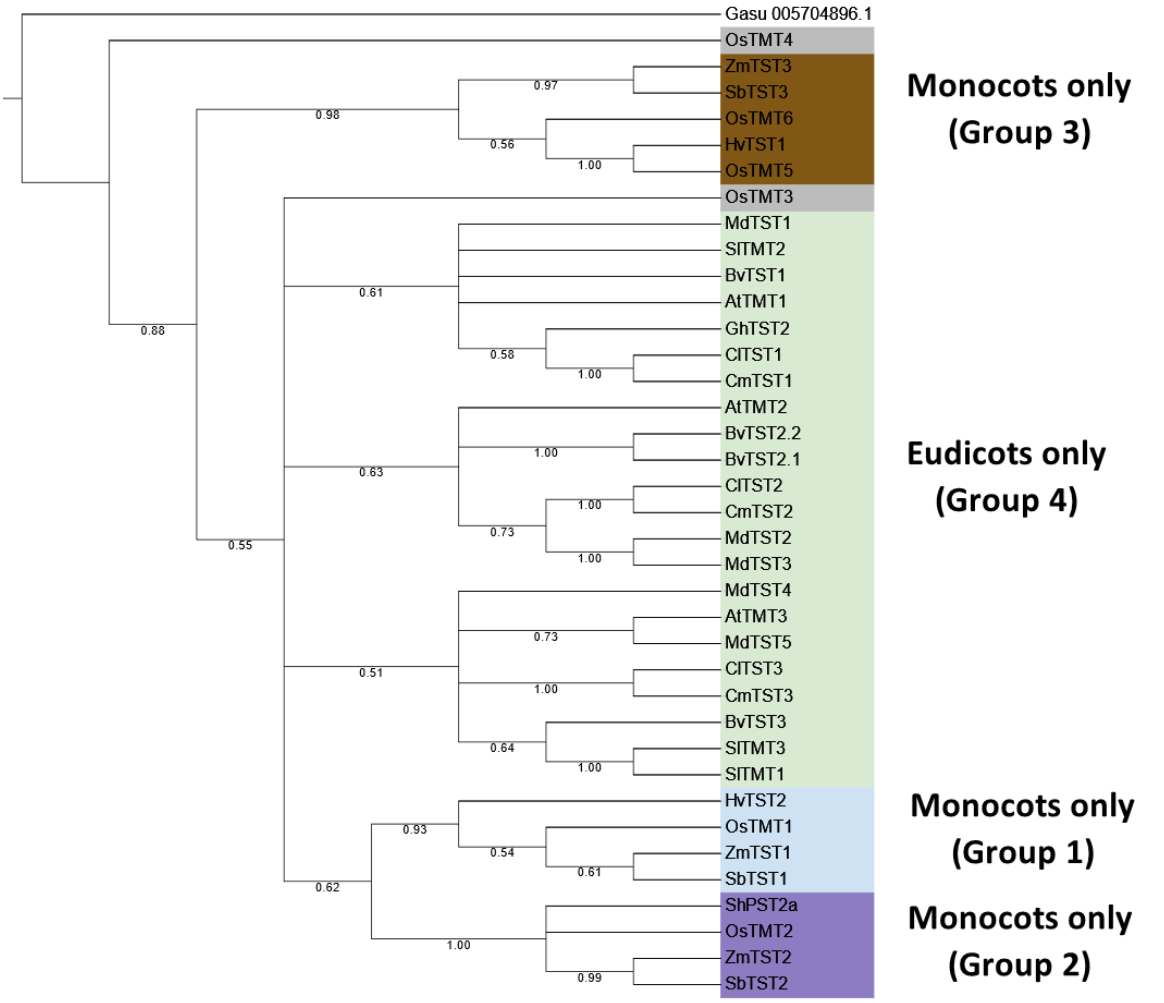


Figure 1.3

Phylogenetic tree of TSTs in select monocot and eudicot plants. Species shown are *Arabidopsis thaliana, Beta vulgaris, Citrullus lanatus, Cucumis melo, Gossypium hirsutum, Hordeum vulgare, Malus domestica, Oryza sativa, Saccharum* hybrid, *Solanum lycopersicum, Sorghum bicolor, Zea mays,* and a red alga *Galdieria sulphuraria* (as the outgroup). TSTs can be divided into three monocot-specific groups (light blue, purple, and brown) and one eudicot-specific group (light green). Sequences that are similar to TSTs but share less than 60% amino acid identity with other rice, maize, or sorghum TST

sequences are indicated as TST-like sequences (grey). The phylogenetic tree was constructed in MEGA X (Kumar et al., 2018) using the Maximum Likelihood method. The branch support values denote the reliability for each internal branch based on 1000 bootstraps. The tree was visualized in interactive Tree of Life (Letunic and Bork, 2021).

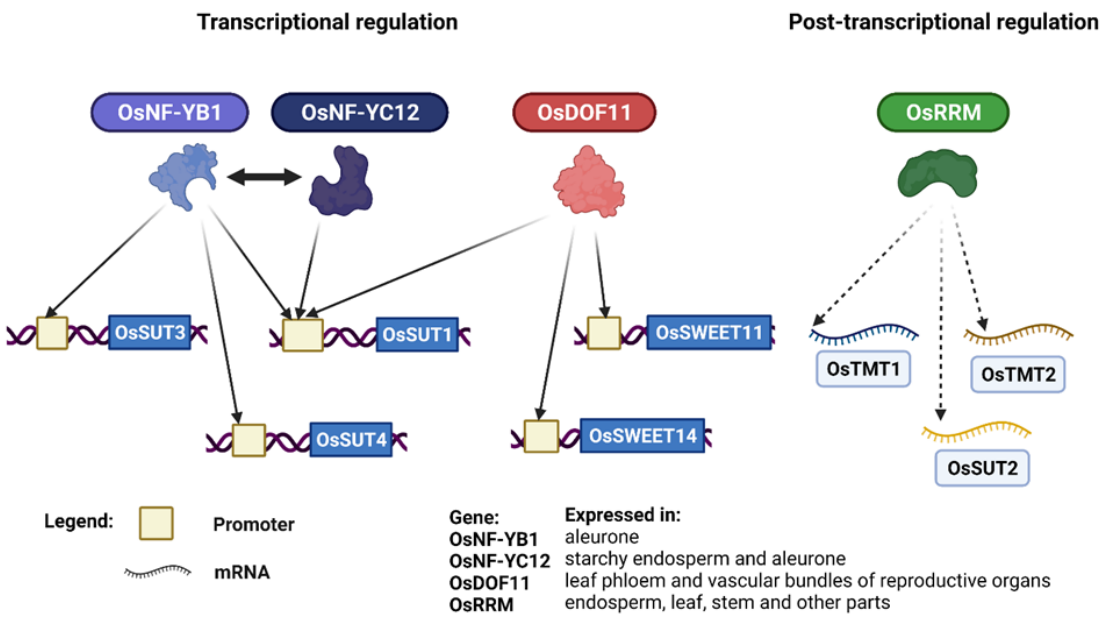


Figure 1.4

Schematic of the regulation of sugar transporters gene expression in rice. Various transcription factors and proteins regulating the expression of genes related to sugar transport in rice are shown. Dotted lines represented post-transcriptional regulation whereas solid line represents transcription factors regulating gene expression at the DNA level. OsNF-YB1, OsNF-YC12, and OsDOF11 are transcription factors known to bind directly to promoters of OsSUT and OsSWEET genes, whereas OsRRM acts on messenger RNAs of OsSUT2, OsTMT1 and OsTMT2. OsNF-YB1 and OsNF-YC12 are predicted to dimerize and collectively control the activity of OsSUT1 in the aleurone layer during grain-filling. Genes and the tissues that they are expressed in are listed.

Chapter 2 : Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L.

Note: The information in this chapter was published under the title:

Zhang, J., Zhang, X., Tang, H., Zhang, Q., Hua, X., Ma, X., Zhu, F., Jones, T., Zhu, X., Bowers, J., Wai, C.M., Zheng, C., Shi, Y., Chen, S., Xu, X., Yue, J., Nelson, D.R., Huang, L., Li, Z., Xu, H., Zhou, D., Wang, Y., Hu, W., Lin, J., Deng, Y., Pandey, N., Mancini, M., Zerpa, D., Nguyen, J.K., Wang, L., Yu, L., Xin, Y., Ge, L., Arro, J., Han, J.O., Chakrabarty, S., Pushko, M., Zhang, W., Ma, Y., Ma, P., Lv, M., Chen, F., Zheng, G., Xu, J., Yang, Z., Deng, F., Chen, X., Liao, Z., Zhang, X., Lin, Z., Lin, H., Yan, H., Kuang, Z., Zhong, W., Liang, P., Wang, G., Yuan, Y., Shi, J., Hou, J., Lin, J., Jin, J., Cao, P., Shen, Q., Jiang, Q., Zhou, P., Ma, Y., Zhang, X., Xu, R., Liu, J., Zhou, Y., Jia, H., Ma, Q., Qi, R., Zhang, Z., Fang, J., Fang, H., Song, J., Wang, M., Dong, G., Wang, G., Chen, Z., Ma, T., Liu, H., **Dhungana, S.R**., Huss, S.E., Yang, X., Sharma, A., Trujillo, J.H., Martinez, M.C., Hudson, M., Riascos, J.J., Schuler, M., Chen, L.-Q., Braun, D.M., Li, L., Yu, Q., Wang, J., Wang, K., Schatz, M.C., Heckerman, D., Van Sluys, M.-A., Souza, G.M., Moore, P.H., Sankoff, D., VanBuren, R., Paterson, A.H., Nagai, C., Ming, R., 2018. Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. Nature Genetics 50(11), 1565-1573.

https://doi.org/10.1038/s41588-018-0237-2

Contributions: S.R.D. annotated and analyzed the sugar transporters in the *S. spontaneum* genome and wrote sections of the Sugar Transporters paragraph describing SUTs and TSTs along with D.M.B.

ABSTRACT

Modern sugarcanes are polyploid interspecific hybrids, combining high sugar content from *S. officinarum* with hardiness, disease resistance and ratooning of *S. spontaneum.* Sequencing of a haploid *S. spontaneum*, AP85-441, facilitates assembly of 32 pseudo-chromosomes comprising 8 homologous groups of four members each, bearing 35,525 genes with alleles defined. The reduction of basic chromosome number from 10 to 8 in *S. spontaneum* is caused by fissions of two ancestral chromosomes followed by translocations to four chromosomes. Surprisingly, 80% of NBS-encoding genes associated with disease resistance located in four rearranged chromosomes and 51% of those in rearranged regions. Re-sequencing of 64 *S. spontaneum* genomes revealed balancing selection in rearranged regions, maintaining their diversity. Introgressed *S. spontaneum* chromosomes in modern sugarcanes were randomly distributed in AP85-441 genome, indicating random recombination among homologs in different *S. spontaneum* accessions. The allele defined *Saccharum* genome offers new knowledge and resources to accelerate sugarcane improvement.

INTRODUCTION

Cultivated sugarcanes (*Saccharum* spp., Poaceae), are unusual among leading crops in that they are polyploid interspecific hybrids, with singularly complex genomes. Domesticated in New Guinea ~10,000 years ago, "reeds that produce honey without bees" were considered a luxury and an expensive spice since the 6th to 4th centuries BC. Introduced to the Old World around the 8th century (Watson, 2008), the spread of sugarcane to Caribbean, South American, Indian Ocean and Pacific island nations drove

large human migrations, including slave labors (Mintz, 1986). Now the world's number one crop by harvested tonnage and fifth most valuable crop (FAO, 2012), sugarcane is cultivated on ~26 million hectares of land in > 90 countries, and 1.83 billion metric tonnes are harvested annually with a gross production value approaching $57 billion, providing 80% of sugar and 40% of ethanol, as the primary sugar and biofuel feedstock crop.

While the high sugar content of modern sugarcane cultivars derives from cultivated 'noble' forms of *S. officinarum*, their hardiness, disease resistance and ratooning capacity were obtained during 'nobilization', specifically backcrossing into *S. officinarum* selected traits from a sugar-poor relative, *S. spontaneum* (Roach, 1972). 'Noble' *Saccharum officinarum* cultivars, typically 2n = 8x = 80, accumulate sucrose in the stem reaching up to 50% of the dry weight, but are vulnerable to biotic and abiotic stresses. Dutch breeders in Java made inter-specific crosses between *S. officinarum* and a wild relative, *S. spontaneum*, to obtain disease resistance and stress tolerance traits of *S. spontaneum* while backcrossing to *S. officinarum* to recover high biomass and high sugar content (Brandes and Sartoris, 1936). Consequently, modern sugarcane cultivars are interspecific hybrids with approximately 80% chromosomes from *S. officinarum*, 10-15% chromosomes from *S. spontaneum*, and 5-10% recombinant chromosomes (D'Hont et al., 1996).

The lowest chromosome number recorded for natural *Saccharum* accession is a 2n = 5x = 40 *S. spontaneum*, which no longer exists – however, exactly one haploid (1n = 4x = 32) *S. spontaneum*, AP85-441, generated from another culture of octoploid SES208 (Moore et al., 1989), provides a foundation for assembly of a prototypical version of the

sugarcane chromosome set. This study illuminates the hereditary blueprint and evolutionary history of one of our most important, and most complex, crop genomes.

RESULTS

Genome sequencing and assembly

The genome size of AP85-441 was estimated at 3.36 Gbp by flow cytometry (Zhang et al., 2012). From a bacterial artificial chromosome (BAC) library of AP85-441, 35,156 BAC clones were pooled into 712 libraries (mostly of 48 BACs: Supplementary Table 1) and individual BAC pools were sequenced independently by a Hiseq-2500 with PE250, yielding 267.5 Gbp data that were assembled using three different assemblers, ALLPATH-LG (Gnerre et al., 2011), SPAdes (Bankevich et al., 2012) and SOAPdenovo2 (Luo et al., 2012), yielding a 2.56 Gbp assembly with contig N50 of 7.4 kb (Supplementary Table 2 and Supplementary Table 3). To reduce fragmentation, 295 Gbp data from Pacific Biosciences (PacBio) RS II system (Supplementary Table 4), was used for self-correction and assembly by CANU (Koren et al., 2017), incorporating assembled BAC sequences, correcting and polishing with 90x Illumina paired end sequences, yielding 3.13 Gbp with contig N50 of 45 kb (Supplementary Table 5). The hybrid assembled contigs and BAC contigs correspond with ~ 99.72% accuracy (Supplementary Table 6).

High throughput chromatin conformation capture (Hi-C) is an extension of 3C technology, in which cross-linked chromatin is digested with an appropriate restriction enzyme and then ligated to obtain an interacting fragment (Dekker, 2006). This approach, which was pioneered by Lieberman-Aiden, Dekker and Burton et al. (Burton et al., 2013; Lieberman-Aiden et al., 2009), was used in grasses before in the assemblies of barley

(Mascher et al., 2017) and wild emmer wheat (Avni et al., 2017). To provide a scaffold for contig assembly, four Hi-C libraries were constructed from young leaves of AP85-441. Chimeric fragments representing the original cross-linked long-distance physical interactions were processed into paired-end sequencing libraries, then 1 billion of 150 bp paired-end Illumina reads were produced and uniquely mapped onto the draft assembly contigs. Due to polyploidy, existing Hi-C scaffolding programs such as LACHESIS (Burton et al., 2013) and SALSA (Jay Ghurye et al., 2017) link *S. spontaneum* allelic haplotypes together and are no longer suitable for this autopolyploid genome. We developed a Hi-C-based scaffolding algorithm (ALLHIC) that integrates four functions, pruning, partition, optimization and building, to select contigs specific for polyploid genome assembly (See methods and Supplementary Figures 1-3). The quality of Hi-C sequencing was evaluated using HiC-Pro (Servant et al., 2015) (Supplementary Table 7 and Supplementary Figure 4).

Hi-C-based physical map was used to assemble 32 pseudo-chromosomes that anchor 2.9 Gbp of genome including 97% of gene content. A high-density genetic map of 998,370 SNPs was used to verify the Hi-C assembly, supporting that the two methods are consistent in both chromosomal assignment and order for 89% contigs (Supplementary Table 8). The 32 pseudo-chromosomes comprise 8 homologous groups with four sets of monoploid chromosomes, A, B, C and D (Fig. 2.1). A total of 219 (88.3 %) complete gene models among 248 ultra-conserved core eukaryotic genes (CEGs) in CEGMA (Parra et al., 2007) and 1,397 (97.01 %) among 1,440 conserved genes in BUSCO (Simao et al., 2015) were recalled in our assembly (Supplementary Tables 9 and 10). Further, 1,624 million (97.01%) of 1,674 million Illumina short reads were alignable and covered

97.3 % of the assembly (Supplementary Table 11). The assembly allowed us to predict 28 potential centromeric regions along the 32 chromosomes, with length ranging from 0.25 Mb to 11.85 Mb (Supplementary Table 12).

Allele specific annotation

High level of homologous gene retention was detected from sequencing multiple haplotypes in sugarcane, despite extreme autopolyploid redundancy (Garsmeur et al., 2011). In autopolyploid genomes, homologous genes at the same locus on homologous chromosomes are defined as alleles (Osborn et al., 2003). Using two rounds of MAKER followed by manual annotation to separate genes and alleles, we annotated 35,525 genes with alleles defined, including 4,289 (12.7%) genes with four alleles, 9,792 (27.6%) with three, 14,797 (41.7%) with two, and 6,647 (18.7%) with one. The total number of alleles is 82,773 with an average 2.3 alleles per gene. In unanchored sequences, 3,130 gene/alleles were annotated. We annotated 1,256 tandemly duplicated genes and 3,375 dispersedly duplicated paralogs (Table 2.1). The cytochrome P450 gene families illustrated the importance of annotating alleles in polyploid genomes, with total of 1,465 manually annotated alleles in 387 genes (Supplementary Figure 5). Without allele-specific annotation, the number of P450 genes in this genome would be 1,465, not 387.

Among the predicted gene models, 90.0% could be found in the sorghum genome (Paterson et al., 2009), and 80% in collinear positions. Comparison with rice, sorghum, maize and Arabidopsis (Supplementary Figure 6) showed that among 21,661 gene families, 1,278 (6%) were unique to *S. spontaneum.* Gene Ontology (GO) enrichment analysis revealed that these *S. spontaneum* specific genes were enriched in a list of GO categories, including response to wounding/external stimulus, serine-type

endopeptidase/peptidase inhibitor activity and ribosomal subunit (*P* and *FDR* <0.01, Fisher test; Supplementary Table 13).

AP85-441 contains 1,842 Mb repetitive sequences, accounting for 58.65 % of the assembled genome (Supplementary Table 14). Long terminal repeat (LTR) retrotransposons account for 45.62% of the genome, including 14.19% Ty1/copia and 26.04% Ty3/gypsy. Kimura distances analysis indicated a more recent LTR burst (Supplementary Figure 7), including Ty1/*copia* and Ty3/*gypsy* superfamilies occurred between 0.72 to 2.9 MYA.

Basic chromosome number reduction

The AP85-441 genome assembly showed chromosome reduction from 10 to 8 in *S. spontaneum* to involve a paleo-duplicated chromosome pair that have experienced frequent recombinations. Alignment to sorghum revealed chromosome fissions in ancestral homologs of sorghum chromosomes 5 and 8, paleo duplicated chromosome pairs A5 and A11 in grasses (Fig 2). The ancestor of SbChr05 (A12) split into two major segments, C5S (A12S) and C5L (A12L) (Salse et al., 2008) , that translocated into ancestors of SbChr06 (A2) and SbChr07 (A5), respectively (Fig. 2.2c event ①). The ancestor of SbChr8 (A11) split into two major segments, C8S (A11S) and C8L (A11L), and translocated into ancestors of SbChr09 (A6) and SbChr02 (A7+A9), respectively. The short fragments that appear to be homologous between SbChr8 and SsChr5 and between SbChr5 and SsChr7 are remains of homeologous genes in sorghum stratum SSA formed 13.4 MYA, well before sorghum and *Saccharum* diverged (Wang et al., 2011). Strikingly, even the smaller SSA region in S5 and larger region of SSA in S8 were conserved in the rearranged AP85-441 genome, reflected in sparse alignment of SsChr5

to SbChr8 and dense alignment of SsChr7 to SbChr5 at the tip of the short arm (Fig. 2.1), validating the high quality and accuracy of AP85-441 genome assembly.

Polyploidization in *S. spontaneum*

We assessed whether the two rounds of whole genome duplication (WGD) affecting sugarcane were allopolyploidization followed by autopolyploidization as proposed (Kim et al., 2014), or just two rounds of autopolyploidization. Although the sequenced genome is a haploid, each gamete contains 4 sets of homologous or hom(e)ologous chromosomes, representing two WGDs (i.e., from one to two to four). Comparison among hom(e)ologous haplotypes A, B, C and D revealed 7.7 million SNPs, 1.03 million short indels and 3,637 SVs, accounting for 11.2 Mb sequences and indicating heterozygosity of 0.98% in the *S. spontaneum* AP85-441 genome (Fig. 2.3 and Supplementary Table 15). To exploit the fact that paralogs are often located on all four of a set of hom(e)ologous chromosomes, we developed a framework to carry out a more sensitive study of gene pair similarities (Supplementary Note). However, no clear partition reflecting two events could be inferred, with each of three complementary approaches suggesting random association among the four members of most homologous series.

It is clear, however, from comparisons of chromosomal rearrangements that there were two discrete WGDs, rather than a single event. Two fissions in ancestral homologs of sorghum chromosomes SbChr5 and SbChr8 that resulted in translocations to a set of 2 chromosomes each occurred before the two rounds of WGDs in *Saccharum*, and after the divergence of *Saccharum* and *Miscanthus* (Fig. 2.2c). However, inversions in ancestral SbChr8S and SbChr8L in two pairs, SsChr2AB and SsChr7AB, indicate that these

occurred after the first WGD but before the second (Fig. 2.2c event ②). Among three regions showing collapsing of homologous sequences (upper region of SsChr1C, middle region of SsChr3D, upper region of SsChr8C), SsChr3B and SsChr8A have about 2x greater depth of Illumina short reads, suggesting that they are the collapsed homologs. SsChr1C region showed equal distribution among 3 homologs, indicating a deletion in SsChr1C (Supplementary Table 16 and Supplementary Figure 8).

Two inversions involving single chromosomes, ancestral SbChr5L (A12L) (homologous chromosome C of SsChr6) (Fig 2.2c event ③) and bottom of SsChr5C (Fig 2.2c event ③), presumably occurred after the two rounds of WGD. Chromosome reduction in *Miscanthus* was caused by fusion of one set of chromosomes homologous to SbChr4 and SbChr7 (Swaminathan et al., 2012). Inversions involving all four homologous chromosomes between SsChr4ABCD appear to have occured before the two rounds of WGD, but it is actually an inversion occurred in SbChr4 after *Saccharum* and *Sorghum* diverged from a common ancestor (Fig. 2.2c event ⑤). These analyses indicate that the two rounds of WGD are autopolyploidization and that they occurred with a brief time apart.

Allelic expression dominance

The homologous genome expression levels of the 4 HGs were similar in examined tissues (Supplementary Figure 9), indicating no significant global homologous genome dominance in *S. spontaneum*. To mitigate differential gene content among the homologous genome, we further extracted 4,289 sets of genes with four alleles in high co-linearity from AP85-441 (Supplementary Figure 10), but overall gene expression level from each haplotype was similar for the 4 HGs (Supplementary Figure 9). These results

are not surprising -- even recent allopolyploids such as *B. napus* (Chalhoub et al., 2014), *G. hirsutum* (Said et al., 2015), *T. aestivum* (Consortium, 2014) and *B. juncea* (Yang et al., 2016) displayed no homologous chromosome dominance. *S. spontaneum* is a recent autopolyploid, and homologous genomes are fluid and changing after each meiosis.

For breeding polyploid crops such as sugarcane, the segregation of alleles with different expression levels may contribute to the segregation of traits in a breeding population. To simplify the analysis of the allelic expression, we sorted the expression levels of four alleles for 4,289 genes by descending order from I to IV. Three allele pairs (I/II, II/III and III/IV) were compared for analyzing the differentially expressed alleles (Supplementary Figure 11-12). We defined the expressional level in three pairs less than two folds as neutral and the other as non-neutral. Of 4,289 of gene, on average, 37.6% displayed neutral and 62.4% are non-neutral, revealing that the expression of alleles varied. We further analyzed the variations of gene number among tissues, the results revealed that the number of both the neutral and non-neutral genes were similar among the examined tissues. However, the genes of these two type expressional patterns vary among examined tissues. On average, 36.3% of neutral, 56.4% of non-neutral were conserved among the tissues (Supplementary Figure 13).

NADP-ME type C4 pathway

The C4 photosynthesis pathway was discovered in sugarcane (Hatch and Slack, 1966; Kortschak et al., 1965). We identified 24 genes for seven key enzymes related to the NADP-ME C4 pathway (Supplementary Figure 14). Increased expression of core C4 enzymes plays a major role in the evolution of C4 photosynthesis (Sage, 2004). Based on gene expression and phylogenetic analysis, eight genes, *SsCA1*, *SsCA2*, *SsPEPC1*,

SsPEPC-k1, *SsNADP-MDH2*, *SsNADP-ME2*, *SsPPDK1*, and *SsPPDK-RP2* were identified as C4-type genes (Supplementary Table 17). A tandem duplication of *SsNADP-ME2*, *SsNADP-ME1*, also displayed the C4 expression profile similar to that of *SsNADP-ME2*. But the ortholog of *SsNADP-ME1* in maize, *ZmNADP-ME (GRMZM2G122479*) displayed non-C4 type expression (Wang et al., 2014), suggesting that neofunctionalization of *SsNADP-ME1* for C4 in sugarcane occurred after the divergence of maize and *Trib. Andropogoneae*.

Sugar transporters

Sucrose transporters (SUTs) are hypothesized to load sucrose into the phloem of leaf minor veins, and also function to retrieve sucrose from the apoplasm during transport (Baker et al., 2016; Julius et al., 2017b; Slewinski and Braun, 2010). In the step prior to phloem loading, SWEETs (Sugars Will Eventually be Exported Transporters) are potentially responsible for sucrose efflux into cell wall space from phloem parenchyma cells/ bundle sheath (Chen et al., 2012; Williams et al., 2000). SWEETs play different important roles in the multiple physiological processes(Eom et al., 2015). In sugarcane and sweet sorghum, the stems are the principal sink tissues that store very high concentrations of sugars within the parenchyma cells (Bihmidine et al., 2015; Bihmidine et al., 2013; Patrick et al., 2013). Tonoplast Sugar Transporters (TSTs) have been characterized as sucrose transporters highly associated with vacuolar sucrose accumulation from sugar beet taproot (Jung et al., 2015), the stems of sugarcane (Casu et al., 2015) and sweet sorghum (Bihmidine et al., 2016), and watermelon fruit (Ren et al., 2018). Whereas there are three TST genes in the sorghum genome (Bihmidine et al., 2016), the family has expanded in the *S. spontaneum* genome, which has four genes,

consisting of 13 homeologs. Hence, it is reasonable for us to hypothesize that TSTs are the most promising players to sequester sucrose into the vacuoles of the sugarcane stem (Bihmidine et al., 2016; Casu et al., 2015; Julius et al., 2017a).

In the *S. spontaneum* genome, we identified 123 sugar transporters from 9 subfamilies, including 4 members in the TST family, 4 members in VGT (vacuolar glucose transporters family), 3 member in pGlcT (plastdic glucose transporters family), 4 members in INT (inositol transporters family), 31 members in PLT (polyol transporters family), 14 members in SFP (early response to dehydration 6-like family), 6 members in SUT (sucrose transporter family), 22 members in SWEET , and 35 members in STP (sugar transporters family or hexose transporters family) (Supplementary Table 18). Phylogenetic analysis of those sugar transporters show gene family expansion in STP and PLT families compared with sorghum (22 in STP and 17 in PLT), rice (21 in STP and 11 in PLT) and Arabidopsis (14 in STP and 9 in PLT) (Supplementary Figure 15). Tandem duplication analysis indicated that 19 and 23 genes of STP and PLT could be assigned to tandem duplication, compared to 11 and 9 genes in sorghum. The cause of STP and PLT family expansions in *S. spontaneum* is tandem duplication.

Disease resistance genes

S. spontaneum contributed disease resistance genes to modern sugarcane hybrid cultivar. We identified 361 sequences putatively encoding nucleotide binding site (NBS) proteins, including 22N-type, 169 NL-type, 68 CN-type, and 102 CNL-type. The NBS-encoding gene number is larger than that in sorghum (Cheng et al., 2010), caused by the species specific tandem duplication in *S. spontaneum*. Surprisingly, 80% of the NBS-encoding genes located in the four rearrangement chromosomes (Ss02, Ss05, Ss06 and

Ss07) and 51% of those were in the rearranged regions, including SsChr5 (Sb05S) 57.6~89.1 Mbp, SsChr6 (Sb05L) 54.6~90.6 Mbp, SsChr7 (Sb08S) 62.0~83.3 Mbp, SsChr2 (Sb08L) 98.5~125.9 Mbp (Supplementary Table 19). R genes are 7 times more likely to locate in the four rearranged regions than in other chromosomes or regions (Fisher's Exact Test, P-value < 2.2e-16; Supplementary Table 20).

S. spontaneum fraction in hybrid sugarcane cultivars

Modern sugarcane cultivars are the product of complex and repeated hybridization between *S. officinarum* and *S. spontaneum*, resulting in complex hybrids with different chromosome numbers and different morphology than their progenitors. Previous studies estimated that *S. spontaneum* genome contributed approximately 10-20% to modern hybrid sugarcane. In the modern hybrid sugarcane SP80-3280, approximately 12.25% of sequences are contributed by *S. spontaneum*. We mapped the sequences back to AP85-441 genome and they were randomly and evenly distributed, not comprising a set or sets of chromosomes as expected (Fig. 2.4). Analysis of integrated *S. spontaneum* fraction in 15 re-sequenced hybrid genomes also yielded random distribution throughout the genome (Supplementary Figure 16).

Origin and genetic diversity of *S. spontaneum*

Most of the genetic diversity found within *S. spontaneum* has not been introgressed into commercial sugarcane, and in principle this germplasm represents a rich source of desirable agronomic traits related to stress tolerance and biomass accumulation (Wang et al., 2008). *S. spontaneum* has a broad natural range extending throughout Asia, the Indian Subcontinent, the Mediterranean and Africa (Panje and Babu, 1960), and

natural populations displays a wide range of phenotypic, genetic and ploidy level diversity.

In practice, however, nucleotide diversity (π) across *S. spontaneum* was estimated to be 0.00021±0.000002 (Supplementary Tables 21 and 22 and Supplementary Figure 17), much lower than other clonally propagated crops such as potato (Hardigan et al., 2017), cassava (Bredeson et al., 2016), grape (Myles et al., 2011), and citrus (Wu et al., 2018). We re-sequenced 64 diverse *S. spontaneum* accessions from the world germplasm collection, identifying 4.48 million, high confidence variants that include 3,961,408 single nucleotide polymorphisms (SNPs), 201,854 insertions and 291,346 deletions, averaging 1.52 variants per kb. We identified 671,265 variants (15%) in genic regions, including 41,960 synonymous, 101,826 nonsynonymous and 491,493 intronic variants.

Both PCA and admixture-based analyses clustered the 64 *S. spontaneum* accessions into three distinct groups (Fig. 2.5a and c; for other K values, see Supplementary Figure 18a-b) that were also supported by phylogenetic relationships among the 64 accessions inferred by bootstrapping and geographic origins (Supplementary Table 23), with group 1 originating from China, Philippines, Indonesia, and Papua New Guinea; and group 2 and group 3 originating from India, Pakistan and Iran. The regions of Pan-Malaysia might be the possible ancient hybrid zones among three groups. Ploidy varies widely within the three groups, from 6x-16x. By mapping the ploidy levels on the BS tree (Fig. 2.5b), the topology shows the accessions of different ploidy levels (from hexaploid to hexadecaploid) diverged independently from ancestors in three groups, suggest the fluid ploidy levels may have independently evolved from ancestral progenitors.

Regions of *S. spontaneum* with large scale chromosomal rearrangements compared to sorghum have higher genetic diversity (higher π value) than non-rearranged regions, and may have undergone much stronger balancing selection (Supplementary Table 22 and Supplementary Figure 19). Although several individual chromosomes do not show significant differences, comparisons averaging values on all chromosomes show nucleotide diversity (π) in rearranged regions (0.00025±0.00003) to be much higher than in non-rearranged regions (0.00021±0.00001, P=0.000234). The Tajima's D in rearranged regions (-0.659±0.052) is much higher than in non-rearranged regions (-0.720±0.011, P=0.005013). SNP density is also higher in rearranged (360.27±48.41) than non-rearranged regions (297.46±12.65, P=0.001798). In addition, the GO (Gene Ontology) enrichments show the non-rearranged regions were enriched in GOs (both *FDR* and P <0.05, Fisher's exact test; Supplementary Table 23) related to basic life cycles, primarily in photosynthesis, respiration and ATP synthesis. Whereas, the rearranged regions were enriched in lots of GOs (both *FDR* and P <0.05, Fisher's exact test; Supplementary Table 24) related to secondary life processes, mainly in secondary cell biosynthesis and metabolism, transmembrane transport, and ion binding.

An intriguing hypothesis is whether genomic rearranged regions might have had a role in adaptation to different habitats. The rearranged regions in *S. spontaneum* with high level of genomic diversity (π value) might be the results of possessing more adaptive genes related to habitat or stress adaptation, such as responses to various abiotic stresses (drought, salinity, alkaline, metal ions etc.), which are controlled by genes of secondary cell biosynthesis and metabolism, transmembrane transport, and ion binding as detected in these regions. Conversely, the non-rearranged regions mainly controlling the basic life

cycles maintains the lower level of genomic diversity is also meaningful. The rearranged regions have undergone stronger balancing selection after the polyploidization events. Adding further support to this notion is that 80% of the NBS-encoding genes were located in the four rearrangement chromosomes.

DISCUSSION

The identification of 80% of disease resistance genes in rearranged chromosomes suggests that reduction of basic chromosome number might have contributed to retaining disease resistance genes. Following chromosome fissions and translocations in a diploid ancestor, translocated fragments may have had little recombination. Following WGD, additional chromosomal rearrangements in these translocated regions may have further suppressed recombination (Fig. 2.1). Population genomic analyses detected balancing selection in these rearranged regions, a mechanism to maintain genetic diversity. It is likely an unintended consequence that these rearranged chromosome arms are enriched with NBS-encoding genes, resulting in more disease resistance genes retained in *S. spontaneum*, leading to higher resistance to disease and abiotic stresses than other *Saccharum* species, making *S. spontaeum* the source of disease and stress tolerance in sugarcane breeding program.

Integration of *S. spontaneum* chromosome segments into modern sugarcane hybrid cultivars by 3-4 generations of backcrossing at random would result in about one set of monoploid *S. spontaneum* chromosomes. The *S. spontaneum* fraction of the sugarcane hybrid cultivar SP80-3280 and 15 re-sequenced hybrid genomes each appear randomly distributed in the reference AP85-441 genome, indicating random recombination of homologous chromosomes in different accessions that undergone many

rounds of meiosis after their separation. This is indirect evidence that *S. spontaneum* is autopolyploid, and it reinforced the importance of allele-specific annotation for mining effective alleles of R genes in hybrid cultivars.

Defining alleles in an autopolyploid genome clarifies gene or gene family analysis as demonstrated in P450 and other gene families. This reference genome offers substantial new knowledge of and unprecedented genomic resources for sugarcane breeders and researchers to mine disease resistance and other alleles in rearranged chromosomes from historic hybrid cultivars, and track them in breeding populations to shorten the 13-year breeding cycle.

ONLINE METHODS

Genome sequencing

Illumina short reads sequencing. DNA was extracted from leaf tissue of a single soil-grown plant using the Qiagen DNeasy Plant Mini Kit and applied to 280-bp and 500-bp paired-end (PE) libraries construction using the NEBNext Ultra DNA Library Prep Kit for Illumina sequencing. Sequencing was performed using Illumina HiSeq 2500 platform.

Construction of BAC libraries and sequencing. Nuclei were isolated from the young leaf tissues of AP85-441 following the method described by Ming et al.(Ming et al., 2001). The high molecular weight DNA embedded in agarose was partially digested using HindIII. Fractions at approximately 100 kb were recovered and cloned into pSMART BAC vector (Lucigen, LA). A total of 38,400 BAC clones were constructed and selected for sequencing. 48 BAC clones were pooled together and DNA libraries were prepared with PhasePrep^{TM} BAC DNA Kit (Sigma, USA) following the

manufacturer's protocols. BAC DNA libraries were sequenced using Illumina HiSeq 2500 with 250 bp Paired-End sequencing strategy.

PacBio Library construction and sequencing. More than 5 µg of sheared and concentrated DNA was applied to size-selection by BluePippin system. ~20 kb SMRTbell^{TM} libraries were prepared according to the released protocol from PACBIO Company. A total of 176 Single-Molecule Real-Time (SMRT) cells were run on Pacbio RSII system with P6-C4 chemistry.

Hi-C library construction and sequencing. Four Hi-C libraries were created from tender leaves of AP85-441 in BioMarker Technologies Company as described before (Xie et al., 2015). Briefly, the leaves were fixed with formaldehyde, lysed, and then the cross-linked DNA digested with *HindIII* overnight. Sticky ends were biotinylated and proximity-ligated to form chimeric junctions, that were enriched for and then physically sheared to a size of 500-700 bp. Chimeric fragments representing the original cross-linked long-distance physical interactions were then processed into paired-end sequencing libraries and 1,001 million of 150 bp paired-end reads were produced on illumina HiSeq X ten platform.

Genome assembly overview

The sugarcane AP85-441 contig-level assembly incorporated sequencing data from a mixture of sequencing technologies (Supplementary Figure 1), including BAC pools sequenced with Illumina HiSeq2500, whole-genome shotgun sequencing with PacBio RSII as well as Hi-C reads, followed by illumina short reads polishing. Each BAC pool was independently assembled using ALLPATH-LG (Gnerre et al., 2011), SPAdes (Bankevich et al., 2012) and SOAPdenovo2 (Luo et al., 2012) and best results

were retained. For PacBio assembly, CANU v1.5 (Koren et al., 2017) was used as it is capable of avoiding collapsed repetitive regions and haplotypes. Self-correction was performed with parameter corOutCoverage=100, which allows us to correct all of the input PacBio reads. The corrected reads along with BAC-assembled contigs were imported to the assembly step. Chromosomal assembly was constructed based on proximity-guided assembly (PGA) using our newly developed program, ALLHIC, which is designed for polyploid genome scaffolding (see Supplementary Notes for details).

Genome annotation

Repeat prediction. We first customized a *de novo* repeat library of the genome using RepeatModeler (see URLs), which can automatically execute two *de novo* repeat finding programs, including RECON (version 1.08) (Bao and Eddy, 2002) and RepeatScout (version 1.0.5) (Price et al., 2005). The consensus TE sequences generated above were imported to RepeatMasker (version 4.05) (Smit and Green, 2013-2015) to identify and cluster repetitive elements. Unknown TEs were further classified using TEclass (version 2.1.3) (Abrusan et al., 2009). To identify tandem repeats within the genome, the Tandem Repeat Finder (TRF) package (version 4.07) (Benson, 1999) was used with the modified parameters of "1 1 2 80 5 200 2000 –d -h" in order to find high order repeats. Telomeres and centromeres were identified based on the .dat output files above. Repeat sequences with more than 10 monomers 'AAACCT' were identified as telomeres. For centromeres identification, we used a similar method described in *Oropetium thomaeum* genome (VanBuren et al., 2015). The largest repeat arrays were identified and clustered as centromeres. To further investigate LTRs, we applied LTR_retriever pipeline(Ou and Jiang, 2017), which is able to integrate results from public programs, e.g. LTR_FINDER

(Xu and Wang, 2007) and LTRharvest (Ellinghaus et al., 2008), and efficiently remove false positives from the initial predictions. The predict LTRs were further classified into intact and non-intact LTRs and the insertion time was estimated as $T=K/2\mu$ (K is the divergence rate and μ is the neutral mutation rate, the default is 1.38×10^{-8} in LTR_retriever) using the scripts implemented in LTR_retriever package (Ou and Jiang, 2017).

Gene annotation

To get high-quality annotation of protein-coding genes, we carried out two rounds of MAKER running following extensive and careful manual inspections in JBrowse (Buels et al., 2016).

In the first round of MAKER running, 10 selected RNA-seq samples were imported into Trinity *de novo* assembly and genome-guided assembly pipelines with default parameters (Haas et al., 2013). RSEM was used to calculate transcript abundance (Li and Dewey, 2011). Transcripts with FPKM < 1 and iso-percentage < 3 % were removed from further analysis. The filtered transcripts were imported to PASA program for construction of comprehensive transcripts as PASA is able to take advantage of the high sensitivity of reference-based assembly while leveraging the ability of *de novo* assembly to detect novel transcripts (Haas et al., 2003). The PASA-assembled transcripts described above were used for training. The nearly “full-length” transcripts were evaluated by comparing with UniProt plant protein database (last accessed on 8 December 2016) and proteins that were covered at least 95 % were retained as candidates. Then *ab initio* gene predictors, including SNAP (Korf, 2004), GENEMARK (Lomsadze et al., 2005) and AUGUSTUS (Stanke et al., 2006), were each trained with those selected proteins. After that, MAKER

pipeline was used to integrate multiple tiers of coding evidence, including *ab initio* gene prediction, transcript evidence and protein evidence, and generate a comprehensive set of protein-coding genes.

In the second round of MAKER running, the predicted gene models with AED score equal to 0 were extracted for re-training using SNAP (Korf, 2004), GENEMARK (Lomsadze et al., 2005) as well as AUGUSTUS (Stanke et al., 2006). In addition, the RNA-seq reads were mapped to AP85-441 genome using HiSAT2 (Kim et al., 2015) version 2.10 and re-assembled using StringTie (Pertea et al., 2016) version 1.3.4, which is a reference-based RNA assembler. Meanwhile, published full-length transcripts based on IsoSeq in sugarcane were also recruited for annotation (Hoang et al., 2017). The four haplotypes (A, B, C, D) were split into four sub-genomes with each containing 8 pseudo-molecules.

Gene structures were visualized in JBrowse (Buels et al., 2016) along with RNA-seq assembled transcripts and homologs from Sorghum, maize and rice genome. We compared the two rounds of MAKER annotation and selected the better ones if their structures are better supported by homologous proteins or RNA-seq assembled transcripts. Genes in the first round of annotation were kept if their structures have not significant improvement in the second round.

Extensive manual inspection of the annotation identified that 28,306 gene models had a significant difference in protein length or sequence similarity compared to other reported protein sequences in NCBI NR database. The corresponding genomic DNA sequences of these genes were extracted for further careful annotation using FGENESH

online version (see URLs) with pre-trained parameters for Sorghum genes. Genes with significant improvement were replaced with the FGENESH annotation.

BUSCO (Simao et al., 2015) version 3 was used for evaluation of annotation completeness. 1,397 (97.1 %) out of 1,440 conserved genes were re-annotated in AP85-441 genome, among which 1,101 (76.5 %) were complete and duplicated BUSCO genes.

Allelic variation analysis

Construction of a monoploid genome. To compare the allelic variations among the four haplotypes, we first generated a monoploid genome. The concept of the monoploid genome is aimed to retain consensus sequences among four haplotypes and cover genetic materials as many as possible. The longest pseudo-molecule was used as reference for each set of haplotypes and the other three haplotypes were mapped against the reference for SNP/indels and structural variations (SVs) calling using nucmer(Kurtz et al., 2004) program. Mapping results were filtered and only best hits were retained. The program show-snps implemented in MUMmer package (Kurtz et al., 2004) was used to identify SNPs and indels with parameters –Clr, which means only SNPs/indels from ambiguous mapping were reported. Consensus sequences were extracted using a homemade PERL script. Insertions that are larger than 50 bp were identified on Assemblytics (Nattestad and Schatz, 2016), a web-based SV analytics tool, and further inserted into the reference genome. Finally, a monoploid genome, containing 8 representative pseudomolecules and 785 Mb sequences, is generated for further analysis (Figure 2.3).

Identification of alleles. Identification of alleles in AP85-441 genome was based on two strategies: 1) synteny-based and 2) coordinate-based approaches. Inter-haplotype syntenic blocks were identified by MCScanX (Wang et al., 2012) and organized into a four-

column table, containing allele A, B, C or D. In addition, genes that were not shown in the table above were mapped to the monoploid genome using GMAP (Wu and Watanabe, 2005) and the ones that have at least 50% overlaps on coordinates were considered as potential alleles. Sequences similarities were checked among alleles based on reciprocal blast and genes without significant similarities to any other allele were removed from the table.

Analysis of allelic variations. We use a reference-based strategy to identify SNPs/indels and structural variations. Similar to the section above, nucmer (Kurtz et al., 2004) program was used to map haplotypes A, B, C and D to the monoploid genome and SNPs were extracted from ambiguous best mapping. Short indels (1-10 bp) and large structural variations were recalled by Assemblytics (Nattestad and Schatz, 2016) based on the alignments above.

Identification of the sequences in hybrid sugarcane that originates from *S. spontaneum*

SP80-3280 genome was firstly masked using the customized TE library and then split into 1-kb fragments. Each of the fragments were blast against AP85-441 and LA-purple (unpublished) masked genomes respectively and the mapping score was calculated for each blast hit using the following formula:

$$S = N * I$$

where S indicates mapping score, N indicates the number of matched bases and I indicates identity in each blast hit.

Fragments were further classified into sequences from AP85-441 and sequences from LA-purple if they have best mapping score in the corresponding category. Sequences were classified into Fragments from both if they have similar mapping scores (less 5% difference) in LA-purple and AP85-441 genomes.

Allelic expression dominance

Tissues including leaves, stems and roots were collected from mature plants and RNA-seq analysis in this part is based on the three samples. RNA-seq reads were trimmed by Trimmomatic and then mapped to AP85-441 genome by HiSAT2 (Kim et al., 2015). FPKM was calculated based on unique mapping reads using StringTie package (Pertea et al., 2016). To analyze allelic expression dominance, we extracted 4,289 genes with full of four alleles from AP85-441 annotation files. To simplify the analysis of the allelic expression, the expression levels of the genes were sorted by descending order from I to IV. Three allele pairs (I vs II, II vs III and III vs IV) were compared for analyzing the deferentially expressed alleles. Allele pairs with differential expression less than twofold change were defined as neutral pair, and otherwise as non-neutral.

Re-sequencing and population analysis

Reads mapping and variants calling. The raw pair-end reads of 64 *Saccharum spontaneum* accessions were trimmed to remove the adaptors and low-quality bases using Trimmomatic (Bolger et al., 2014) after quality control by FastQC (Andrews, 2010). The reads were filtered with a sliding window of size 7, with average Phred score scale = 20 within the window. The trimmed reads were mapped to the *S. spontaneum* genome using Bowtie2(Langmead and Salzberg, 2012) with default parameters. The mapped reads were sorted, and duplicated reads were removed using SAMtools (Li et al., 2009). We

estimated the rate of uniquely mapped reads outputted from both BWA (Li and Durbin, 2009) and Bowtie2 (Langmead and Salzberg, 2012). Bowtie2 generated ten times as many uniquely mapped reads (UMR) than alignments generated from BWA (Supplementary Figure 21).

The Realigner Target Creator and Indel Realigner programs from the Genome Analysis Toolkit (GATK) package (McKenna et al., 2010) were used for global realignment of reads around indels from the sorted BAM files. The HaplotypeCaller was used to estimate the SNPs and Indels for putative diploids using the default parameters. The HaplotypeCaller outputted 42,585,337 unfiltered variants (SNPs and Indels). The distribution of calling depths (DP) of each raw variant were estimated as a criterion for variants filtering. Low depths and repetitive variants were removed from the raw VCF file if they had DP < 2 or DP > 45, minQ < 30. We allowed the variants sites with maximum-missing rate as 15%. These filtering strategies reduced the raw unfiltered set of 42.59 million variants (SNPs and Indels) to the working set of 4.48 million (4,476,608) variants. SnpEff (v3.6c) (Cingolani et al., 2012) was used to assign variants effects on the basis of gene models from *S. spontaneum* genome annotation. The variants sites were annotated as the SNPs and Indels, as well as intergenic and genic regions (including the synonymous, nonsynonymous, intronic, upstream and downstream variants).

Genome-wide genetic diversity estimation. Population genetic statistics of SNP density, π and Tajima's D were calculated directly from the filtered VCF file in 1000-kb window and 500-kb step for π, and non-overlapping intervals for SNPs density and Tajima's D in VCFtools (Danecek et al., 2011). The high confidence 4,476,608 variant set was used for statistical estimations.

PCA. Principal component analysis (PCA) was performed using the GCTA software on the filtered 4,476,608 variants. The input Plink binary files are transformed from the filtered VCFs file using VCFtools (Danecek et al., 2011) and PLINK (Purcell et al., 2007). The top three principal components were used for assigning the 64 accessions and downstream population structure analysis.

Phylogeny. Bi-allelic and polymorphic 3,969,408 SNPs were used for reconstructing the phylogenetic relationships among 64 accessions. Before tree construction, we filtered and pruned the SNPs (with MAF< 0.2, and missing rate > 0.15, LD threshold = 0.2). Finally, 37,617 SNPs are selected in total for the constructed tree using SNPhylo software. The multiple consensus sequences were aligned using MUSCLE (Edgar, 2004). ML trees were constructed using maximum likelihood method by running DNAML programs in the PHYLIP package (Plotree and Plotgram, 1989). In addition, BS tree was constructed by bootstrapping (bootstrap =10000) analysis using PHANGORN package (Schliep, 2011). Figtree v.1.4 (see URLs) was used for to visualize the trees.

Population genetic structure. Ancestral population stratification among 64 accessions was inferred using Admixture software. The optimal ancestral population structure was estimated from the same variants set with STRUCTURE (Evanno et al., 2005) using ancestral population sizes K = 1~20 and choosing the population with smallest cross-validation error. The parameter standard errors were estimated using bootstrapping (bootstrap =200) when doing the admixture analyses. DISTRUCT (Rosenberg, 2004) was used to plot the population stratification results for K = 1 through K = 20 (Supplementary Figure 18).

Differentiation of genomic diversity among four homologous haploid sets. The reads mapped to each of four homologous haploid sets (A, B, C and D) of AP85-441 genome were retrieved for each of 64 accessions using the SAMtools (Li et al., 2009) and Bedtools (Quinlan and Hall, 2010). The four sets of retrieved reads for each of 64 accessions were mapped to each of eight chromosomes in a consensus monoploid genome separately using Bowtie2 with default parameters. The variants were called from cohort of 256 BAM files generated from previous step for each of eight chromosomes. The HaplotypeCaller of GATK was used to estimate the SNPs and Indels for putative diploids using the default parameters. The HaplotypeCaller outputted 17,531,765 unfiltered variants (SNPs and Indels). The distribution of calling depths (DP) of each raw variant was estimated as a criterion for variants filtering. Low depths and repetitive variants were removed from the raw VCF file if they had DP < 1 or DP > 5, minQ < 20. We allowed the variants sites with max-missing rate as 50%. These filtering strategies reduced the raw unfiltered set of variants (SNPs and Indels) to the working set of 68,911 variants.

Genomic diversity of genomic rearranged regions. To test whether genomic rearranged regions have a genetic difference from that of non-rearrangement regions, we compared the population genetic statistics π, SNP density, and Tajima's D between rearrangement and non-rearrangement regions in each of four sets (A, B, C and D) of chromosome 2, 5, 6 and 7. The genomic rearrangement regions inferred by collinear dot plot and alleles phasing are showed Supplementary Table 21. We used the T-test and Mann–Whitney U test with the one-tailed hypothesis to compare the differences of statistics (π, SNP density, and Tajima's D) between RAR and Non-RAR. To find the difference in gene

functions between RAR and Non-RAR, we conducted go enrichment for the gene models in RAR and Non-RAR. We firstly blastX the *S. spontaneum* gene models in NCBI Nr database of *Oryza_sativa* (see URLs). Then, the functional annotation and go enrichments of gene models were conducted in Blast2Go v4.1 software(Conesa et al., 2005). We used gene models of RAR or Non-RAR as tested gene sets, the whole gene models as reference. The significance of enrichment was valued using the Fisher's exact test.

Genomic diversity among different polyploidy accessions. To test the effects of polyploidization on the genetic diversity, we therefore compare the population nucleotide diversity (π) among accessions with different ploidy levels. We used 1,000-kb sliding window and 500-kb step to calculate the values of each statistic. In addition, we divided the 64 accessions into four groups (ploidy 6, 8, 10 and 13-16) depend on their ploidy level. The four groups are used to calculate the pairwise Weir and Cockerham's *Fst* between the two of them using VCFtools (v0.1.12b)[46] with 1,000-k sliding window and 500-k step.

Data availability

The genome assembly and gene annotation have been deposited in the NCBI database under BioProject number SUB4330926 and BioSample number SUB4330926.

URLs

FGENESH online version, http://www.softberry.com/berry.phtml?topic=fgenesh&group=help&subgroup=gfind; FigTree, http://tree.bio.ed.ac.uk/software/figtree/; NCBI NR database for *Oryza sativa*,

ftp://ftp.ncbi.nih.gov/blast/db; RepeatModeler, http://www.repeatmasker.org/RepeatModeler/,

Acknowledgements

We thank Leah McHale for reviewing and commenting the section on disease resistance genes. This project was supported by startup fund from Fujian Agriculture and Forestry University to R.M., the International Consortium for Sugarcane Biotechnology project #35 to R.M., US DOE DE-SC0010686 to R.M., EBI BP2012OO2J17 to R.M., US NSF Plant Genome Research Program grant IOS-1025976 to D.M.B., the 863 program (2013AA102604) to J.Z., NSFC (31201260) to J.Z., and Program for New Century Excellent Talents in Fujian Province to J.Z. This work was funded in part by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC 18420) to M.H and L.Q.C.

Author Contributions:

R.M. conceived this genome project and coordinated research activities; R.M., C.N., J.Z., and Q.Y. designed the experiments; H.T. and XingtanZ. developed novel algorithms to assemble autopolyploid genomes; T.J., P.H.M., C.N. generated and maintained haploid plant materials and mapping populations; J.Z., Q.Z., X.H., ZhenL., Y.W., L.W., J.X. collected and maintained plant materials; R.M. and Q.Y. genertaed BAC libraries; C.M.W., J.A., J.O.H., S.C., M.P., A.S. isolated BAC DNA; S.C., L.H., W.Z., Y.M., Z.Y., F.D. and X.C. sequenced and processed the raw data; XingtanZ, H.T., J.Z., Q.Z. conducted comparative genomic analysis; H.T., XingtanZ., J.Z., Q.Z., X.X., H.X., P.C., Z.K. and L.L. assembled and annotated the genome; J.Z., X.Z., Q.Z., H. T.,Y.S., S.C.,

X.X. studied genome evolution; J.Z., X.H., XinguangZ., M.L., F.C., G.Z. conducted the C4 photosynthesis analysis; J.Z., X.Z., Q.Z., X.H.,Y.S., L.H., ZhenL., Y.W., W.H., JishanL. analysed gene families; R.M. XingtanZ., Z.F., J.Z., Q.Z., X.H., X.M., Y.S., S.C., X.X., J.Y., L.H., ZhenL., H.X., D.Z., Y.W., W.H., L.W., JishanL., Y.D., P.M., Z.Y., F.D., ZhicongL., H.L., H.Y., W.Z., P.L., G.W., Y.Y., J.S., J.H., JingxianL.,Q.S., Q.J., P.Z.,Y.M., XunxiaoZ., R.X., JuanL., Y.Z., H.J., Q.M., R.Q. Z.Z., Y.D., N.P., M.M., D.Z., J.K.N., L.W., L.Y., Y.X., L.G. manually checked the gene allele anotation; S.R.D., D.M.B., S.H.E., L.Q.C. analyzed sugar transporters, D.N., M.S. analyzed P450 genes, Z.Z., K.W. analyzed chromosome centromere; C.Z. and D.S. tested whole genome duplications; X.M., R.V.B., C.M.W., X.Y.J., J.W., J.H.T., M.C.M., J.J.R. analyzed re-sequenced populations, M.C.S., D.H., M.A.V., G.M.S. contributed SP80-3280 genome; R.M., A.H.P., J.Z., H.T., XingtanZ. wrote the manuscript.

Competing financial interests: The authors have no competing financial interests.

References

Abrusan, G., Grundmann, N., DeMester, L., Makalowski, W., 2009. TEclass--a tool for automated classification of unknown eukaryotic transposable elements. Bioinformatics 25(10), 1329-1330.

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data.

Avni, R., Nave, M., Barad, O., Baruch, K., Twardziok, S.O., Gundlach, H., Hale, I., Mascher, M., Spannagl, M., Wiebe, K., Jordan, K.W., Golan, G., Deek, J., Ben-Zvi, B., Ben-Zvi, G., Himmelbach, A., MacLachlan, R.P., Sharpe, A.G., Fritz, A., Ben-David, R., Budak, H., Fahima, T., Korol, A., Faris, J.D., Hernandez, A., Mikel, M.A., Levy, A.A., Steffenson, B., Maccaferri, M., Tuberosa, R., Cattivelli, L., Faccioli, P., Ceriotti, A., Kashkush, K., Pourkheirandish, M., Komatsuda, T., Eilam, T., Sela, H., Sharon, A., Ohad, N., Chamovitz, D.A., Mayer, K.F.X., Stein, N., Ronen, G., Peleg, Z., Pozniak, C.J., Akhunov, E.D., Distelfeld, A., 2017. Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. Science 357(6346), 93-97.

Baker, R.F., Leach, K.A., Boyer, N.R., Swyers, M.J., Benitez-Alfonso, Y., Skopelitis, T., Luo, A., Sylvester, A., Jackson, D., Braun, D.M., 2016. Sucrose transporter *ZmSut1* expression and localization uncover new insights into sucrose phloem loading. Plant physiology 172, 1876-1898.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing.

Journal of computational biology : a journal of computational molecular cell biology 19(5), 455-477.

Bao, Z., Eddy, S.R., 2002. Automated de novo identification of repeat sequence families in sequenced genomes. Genome Res 12(8), 1269-1276.

Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic acids research 27(2), 573-580.

Bihmidine, S., Baker, R.F., Hoffner, C., Braun, D.M., 2015. Sucrose accumulation in sweet sorghum stems occurs by apoplasmic phloem unloading and does not involve differential Sucrose transporter expression. BMC plant biology 15, 186.

Bihmidine, S., Hunter, C.T., Johns, C.E., Koch, K.E., Braun, D.M., 2013. Regulation of assimilate import into sink organs: update on molecular drivers of sink strength. Frontiers in Plant Science 4, 177.

Bihmidine, S., Julius, B.T., Dweikat, I., Braun, D.M., 2016. Tonoplast Sugar Transporters (SbTSTs) putatively control sucrose accumulation in sweet sorghum stems. Plant signaling & behavior 11(1), e1117721.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15), 2114-2120.

Brandes, E.W., Sartoris, G.B., 1936. Sugarcane: Its origin and improvement. Yearbook of the United States Department of Agriculture, 561-624.

Bredeson, J.V., Lyons, J.B., Prochnik, S.E., Wu, G.A., Ha, C.M., Edsinger-Gonzales, E., Grimwood, J., Schmutz, J., Rabbi, I.Y., Egesi, C., 2016. Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. Nature biotechnology 34(5), 562.

Buels, R., Yao, E., Diesh, C.M., Hayes, R.D., Munoz-Torres, M., Helt, G., Goodstein, D.M., Elsik, C.G., Lewis, S.E., Stein, L., Holmes, I.H., 2016. JBrowse: a dynamic web platform for genome visualization and analysis. Genome Biol 17, 66.

Burton, J.N., Adey, A., Patwardhan, R.P., Qiu, R., Kitzman, J.O., Shendure, J., 2013. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. Nat Biotechnol 31(12), 1119-1125.

Casu, R.E., Rae, A.L., Nielsen, J.M., Perroux, J.M., Bonnett, G.D., Manners, J.M., 2015. Tissue-specific transcriptome analysis within the maturing sugarcane stalk reveals spatial regulation in the expression of cellulose synthase and sucrose transporter gene families. Plant molecular biology 89(6), 607-628.

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I.A., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., Samans, B., 2014. Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. Science 345(6199), 950-953.

Chen, L.Q., Qu, X.Q., Hou, B.H., Sosso, D., Osorio, S., Fernie, A.R., Frommer, W.B., 2012. Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. Science 335(6065), 207-211.

Cheng, X., Jiang, H., Zhao, Y., Qian, Y., Zhu, S., Cheng, B., 2010. A genomic analysis of disease-resistance genes encoding nucleotide binding sites in Sorghum bicolor. Genet Mol Biol 33(2), 292-297.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 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 6(2), 80-92.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21(18), 3674-3676.

Consortium, I.W.G.S., 2014. A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. Science 345(6194), 1251788.

D'Hont, A., Grivet, L., Feldmann, P., Rao, S., Berding, N., Glaszmann, J.C., 1996. Characterisation of the double genome structure of modern sugarcane cultivars (Saccharum spp.) by molecular cytogenetics. Mol Gen Genet 250(4), 405-413.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., 2011. The variant call format and VCFtools. Bioinformatics 27(15), 2156-2158.

Dekker, J., 2006. The three 'C' s of chromosome conformation capture: controls, controls, controls. Nature methods 3(1), 17-21.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research 32(5), 1792-1797.

Ellinghaus, D., Kurtz, S., Willhoeft, U., 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC bioinformatics 9, 18.

Eom, J.S., Chen, L.Q., Sosso, D., Julius, B.T., Lin, I.W., Qu, X.Q., Braun, D.M., Frommer, W.B., 2015. SWEETs, transporters for intracellular and intercellular sugar translocation. Curr Opin Plant Biol 25, 53-62.

Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular ecology 14(8), 2611-2620.

Garsmeur, O., Charron, C., Bocs, S., Jouffe, V., Samain, S., Couloux, A., Droc, G., Zini, C., Glaszmann, J.C., Van Sluys, M.A., D'Hont, A., 2011. High homologous gene conservation despite extreme autopolyploid redundancy in sugarcane. The New phytologist 189(2), 629-642.

Gnerre, S., Maccallum, I., Przybylski, D., Ribeiro, F.J., Burton, J.N., Walker, B.J., Sharpe, T., Hall, G., Shea, T.P., Sykes, S., Berlin, A.M., Aird, D., Costello, M., Daza, R., Williams, L., Nicol, R., Gnirke, A., Nusbaum, C., Lander, E.S., Jaffe, D.B., 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proceedings of the National Academy of Sciences of the United States of America 108(4), 1513-1518.

Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith, R.K., Jr., Hannick, L.I., Maiti, R., Ronning, C.M., Rusch, D.B., Town, C.D., Salzberg, S.L., White, O., 2003. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 31(19), 5654-5666.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N., Regev, A., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature protocols 8(8), 1494-1512.

Hardigan, M.A., Laimbeer, F.P.E., Newton, L., Crisovan, E., Hamilton, J.P., Vaillancourt, B., Wiegert-Rininger, K., Wood, J.C., Douches, D.S., Farré, E.M., 2017. Genome diversity of tuber-bearing Solanum uncovers complex evolutionary history and targets of domestication in the cultivated potato. Proceedings of the National Academy of Sciences 114(46), E9999-E10008.

Hatch, M.D., Slack, C.R., 1966. Photosynthesis by sugar-cane leaves. A new carboxylation reaction and the pathway of sugar formation. The Biochemical journal 101(1), 103-111.

Hoang, N.V., Furtado, A., Mason, P.J., Marquardt, A., Kasirajan, L., Thirugnanasambandam, P.P., Botha, F.C., Henry, R.J., 2017. A survey of the complex transcriptome from the highly polyploid sugarcane genome using full-length isoform sequencing and de novo assembly from short read sequencing. BMC Genomics 18(1), 395.

Jay Ghurye, Mihai Pop, Sergey Koren, Chin, C.-S., 2017. Scaffolding of long read assemblies using long range contact information. BMC Genomics 18(527).

Julius, B.T., Leach, K.A., Tran, T.M., Mertz, R.A., Braun, D.M., 2017a. Sugar Transporters in Plants: New Insights and Discoveries. Plant Cell Physiol 58(9), 1442-1460.

Julius, B.T., Leach, K.A., Tran, T.M., Mertz, R.A., Braun, D.M., 2017b. Sugar transporters in plants: New insights and discoveries. Plant and Cell Physiology 58(9), 1442-1460.

Jung, B., Ludewig, F., Schulz, A., Meißner, G., Wöstefeld, N., Flügge, U.I., Pommerrenig, B., Wirsching, P., Sauer, N., Koch, W., Sommer, F., Mühlhaus, T.,

Schroda, M., Cuin, T.A., Graus, D., Marten, I., Hedrich, R., Neuhaus, H.E., 2015. Identification of the transporter responsible for sucrose accumulation in sugar beet taproots. Nat Plants 1, 14001.

Kim, C., Wang, X., Lee, T.H., Jakob, K., Lee, G.J., Paterson, A.H., 2014. Comparative Analysis of Miscanthus and Saccharum Reveals a Shared Whole-Genome Duplication but Different Evolutionary Fates. The Plant cell 26(6), 2420-2429.

Kim, D., Langmead, B., Salzberg, S.L., 2015. HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12(4), 357-360.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., Phillippy, A.M., 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27(5), 722-736.

Korf, I., 2004. Gene finding in novel genomes. BMC bioinformatics 5, 59.

Kortschak, H.P., Hartt, C.E., Burr, G.O., 1965. Carbon Dioxide Fixation in Sugarcane Leaves. Plant physiology 40(2), 209-213.

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., Salzberg, S.L., 2004. Versatile and open software for comparing large genomes. Genome Biol 5(2), R12.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nature methods 9(4), 357-359.

Li, B., Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC bioinformatics 12, 323.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25(14), 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25(16), 2078-2079.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., Sandstrom, R., Bernstein, B., Bender, M.A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L.A., Lander, E.S., Dekker, J., 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326(5950), 289-293.

Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y.O., Borodovsky, M., 2005. Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic acids research 33(20), 6494-6506.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D.W., Yiu, S.M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, J., Lam, T.W., Wang, J., 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience 1(1), 18.

Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S.O., Wicker, T., Radchuk, V., Dockter, C., Hedley, P.E., Russell, J., Bayer, M., Ramsay, L., Liu, H., Haberer, G., Zhang, X.Q., Zhang, Q., Barrero, R.A., Li, L., Taudien, S., Groth, M., Felder, M., Hastie, A., Simkova, H., Stankova, H., Vrana, J., Chan, S., Munoz-Amatriain, M., Ounit, R., Wanamaker, S., Bolser, D., Colmsee, C., Schmutzer, T., Aliyeva-Schnorr, L., Grasso, S., Tanskanen, J., Chailyan, A.,

Sampath, D., Heavens, D., Clissold, L., Cao, S., Chapman, B., Dai, F., Han, Y., Li, H., Li, X., Lin, C., McCooke, J.K., Tan, C., Wang, P., Wang, S., Yin, S., Zhou, G., Poland, J.A., Bellgard, M.I., Borisjuk, L., Houben, A., Dolezel, J., Ayling, S., Lonardi, S., Kersey, P., Langridge, P., Muehlbauer, G.J., Clark, M.D., Caccamo, M., Schulman, A.H., Mayer, K.F.X., Platzer, M., Close, T.J., Scholz, U., Hansson, M., Zhang, G., Braumann, I., Spannagl, M., Li, C., Waugh, R., Stein, N., 2017. A chromosome conformation capture ordered sequence of the barley genome. Nature 544(7651), 427-433.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research 20(9), 1297-1303.

Ming, R., Moore, P., Zee, F., Abbey, C., Ma, H., Paterson, A., 2001. Construction and characterization of a papaya BAC library as a foundation for molecular dissection of a tree-fruit genome. Theoretical and Applied Genetics 102(6-7), 892-899.

Mintz, S.W., 1986. Sweetness and power: The place of sugar in modern history. Penguin.

Moore, P.H., C. Nagai, Fitch, M.M.M., 1989. Production and evaluation of sugarcane haploids. Proc. Int. Soc. Sugar Cane Technol. 20, 599-607.

Myles, S., Boyko, A.R., Owens, C.L., Brown, P.J., Grassi, F., Aradhya, M.K., Prins, B., Reynolds, A., Chia, J.-M., Ware, D., 2011. Genetic structure and domestication history of the grape. Proceedings of the National Academy of Sciences 108(9), 3530-3535.

Nattestad, M., Schatz, M.C., 2016. Assemblytics: a web analytics tool for the detection of variants from an assembly. Bioinformatics 32(19), 3021-3023.

Osborn, T.C., Pires, J.C., Birchler, J.A., Auger, D.L., Chen, Z.J., Lee, H.-S., Comai, L., Madlung, A., Doerge, R., Colot, V., 2003. Understanding mechanisms of novel gene expression in polyploids. Trends in genetics 19(3), 141-147.

Ou, S., Jiang, N., 2017. LTR_retriever: A Highly Accurate And Sensitive Program For Identification Of LTR Retrotransposons. bioRxiv.

Panje, R., Babu, C., 1960. Studies in Saccharum spontaneum distribution and geographical association of chromosome numbers. Cytologia 25(2), 152-172.

Parra, G., Bradnam, K., Korf, I., 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23(9), 1061-1067.

Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., 2009. The Sorghum bicolor genome and the diversification of grasses. Nature 457(7229), 551.

Patrick, J.W., Botha, F.C., Birch, R.G., 2013. Metabolic engineering of sugars and simple sugar derivatives in plants. Plant Biotechnology Journal 11(2), 142-156.

Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., Salzberg, S.L., 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nature protocols 11(9), 1650-1667.

Plotree, D., Plotgram, D., 1989. PHYLIP-phylogeny inference package (version 3.2). cladistics 5(163), 6.

Price, A.L., Jones, N.C., Pevzner, P.A., 2005. De novo identification of repeat families in large genomes. Bioinformatics 21 Suppl 1, i351-358.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., Daly, M.J., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics 81(3), 559-575.

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26(6), 841-842.

Ren, Y., Guo, S., Zhang, J., He, H., Sun, H., Tian, S., Gong, G., Zhang, H., Levi, A., Tadmor, Y., Xu, Y., 2018. A Tonoplast Sugar Transporter Underlies a Sugar Accumulation QTL in Watermelon. Plant physiology 176(1), 836-850.

Roach, B.T., 1972. Nobilization of sugarcane. Proc Int Soc Sugar Cane Technol 14, 206-216.

Rosenberg, N.A., 2004. DISTRUCT: a program for the graphical display of population structure. Molecular Ecology Resources 4(1), 137-138.

Sage, R.F., 2004. The evolution of C4 photosynthesis. New phytologist 161(2), 341-370.

Said, J.I., Song, M., Wang, H., Lin, Z., Zhang, X., Fang, D.D., Zhang, J., 2015. A comparative meta-analysis of QTL between intraspecific Gossypium hirsutum and interspecific G. hirsutum× G. barbadense populations. Molecular Genetics and Genomics 290(3), 1003-1025.

Salse, J., Bolot, S., Throude, M., Jouffe, V., Piegu, B., Quraishi, U.M., Calcagno, T., Cooke, R., Delseny, M., Feuillet, C., 2008. Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. The Plant cell 20(1), 11-24.

Schliep, K.P., 2011. phangorn: phylogenetic analysis in R. Bioinformatics 27(4), 592.

Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.-J., Vert, J.-P., Heard, E., Dekker, J., Barillot, E., 2015. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome biology 16(1), 259.

Simao, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19), 3210-3212.

Slewinski, T.L., Braun, D.M., 2010. Current perspectives on the regulation of whole-plant carbohydrate partitioning. Plant Science 178, 341-349.

Smit, A., Hubley, R, Green, P., 2013-2015. RepeatMasker Open-4.0. http://www.repeatmasker.org.

Stanke, M., Schoffmann, O., Morgenstern, B., Waack, S., 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC bioinformatics 7, 62.

Swaminathan, K., Chae, W.B., Mitros, T., Varala, K., Xie, L., Barling, A., Glowacka, K., Hall, M., Jezowski, S., Ming, R., Hudson, M., Juvik, J.A., Rokhsar, D.S., Moose, S.P., 2012. A framework genetic map for Miscanthus sinensis from RNAseq-based markers shows recent tetraploidy. BMC Genomics 13, 142.

VanBuren, R., Bryant, D., Edger, P.P., Tang, H., Burgess, D., Challabathula, D., Spittle, K., Hall, R., Gu, J., Lyons, E., Freeling, M., Bartels, D., Ten Hallers, B., Hastie, A., Michael, T.P., Mockler, T.C., 2015. Single-molecule sequencing of the desiccation-tolerant grass Oropetium thomaeum. Nature 527(7579), 508-511.

Wang, L.-P., Jackson, P.A., Lu, X., Fan, Y.-H., Foreman, J.W., Chen, X.-K., Deng, H.-H., Fu, C., Ma, L., Aitken, K.S., 2008. Evaluation of sugarcane× Saccharum

spontaneum progeny for biomass composition and yield components. Crop Science 48(3), 951-961.

Wang, L., Czedik-Eysenberg, A., Mertz, R.A., Si, Y., Tohge, T., Nunes-Nesi, A., Arrivault, S., Dedow, L.K., Bryant, D.W., Zhou, W., 2014. Comparative analyses of C 4 and C 3 photosynthesis in developing leaves of maize and rice. Nature biotechnology 32(11), 1158.

Wang, X., Tang, H., Paterson, A.H., 2011. Seventy million years of concerted evolution of a homoeologous chromosome pair, in parallel, in major Poaceae lineages. The Plant cell 23(1), 27-37.

Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.H., Jin, H., Marler, B., Guo, H., Kissinger, J.C., Paterson, A.H., 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic acids research 40(7), e49.

Watson, A.M., 2008. Agricultural Innovation in the Early Islamic World: The Diffusion of Crops and Farming Techniques, 700-1100. Cambridge University Press.

Williams, L.E., Lemoine, R., Sauer, N., 2000. Sugar transporters in higher plants--a diversity of roles and complex regulation. Trends Plant Sci 5(7), 283-290.

Wu, G.A., Terol, J., Ibanez, V., Lopez-Garcia, A., Perez-Roman, E., Borreda, C., Domingo, C., Tadeo, F.R., Carbonell-Caballero, J., Alonso, R., Curk, F., Du, D., Ollitrault, P., Roose, M.L., Dopazo, J., Gmitter, F.G., Rokhsar, D.S., Talon, M., 2018. Genomics of the origin and evolution of Citrus. Nature 554(7692), 311-316.

Wu, T.D., Watanabe, C.K., 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21(9), 1859-1875.

Xie, T., Zheng, J.F., Liu, S., Peng, C., Zhou, Y.M., Yang, Q.Y., Zhang, H.Y., 2015. De novo plant genome assembly based on chromatin interactions: a case study of Arabidopsis thaliana. Molecular plant 8(3), 489-492.

Xu, Z., Wang, H., 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic acids research 35(Web Server issue), W265-268.

Yang, J., Liu, D., Wang, X., Ji, C., Cheng, F., Liu, B., Hu, Z., Chen, S., Pental, D., Ju, Y., 2016. The genome sequence of allopolyploid Brassica juncea and analysis of differential homoeolog gene expression influencing selection. Nature genetics 48(10), 1225.

Zhang, J., Nagai, C., Yu, Q., Pan, Y.-B., Ayala-Silva, T., Schnell, R.J., Comstock, J.C., Arumuganathan, A.K., Ming, R., 2012. Genome size variation in three Saccharum species. Euphytica 185(3), 511-519.

TABLES

Table 2.1

Allele annotation in AP85-441 genome

	Total No. of genes	No. of genes with 4 alleles	No. of genes with 3 alleles	No. of genes with 2 alleles	No. of genes with 1 allele	No. Of dispersely duplicated genes	No. of tandem duplicated genes
Chr1	6,677	682	1,663	2,903	1,429	654	211
Chr2	5,961	784	1,717	2,438	1,022	558	225
Chr3	5,097	525	1,419	2,158	995	443	180
Chr4	4,081	529	1,112	1,687	753	374	165
Chr5	4,325	476	1,077	1,852	920	391	145
Chr6	3,800	483	1,069	1,556	692	365	132
Chr7	4,013	516	1,135	1,643	719	427	139
Chr8	1,571	294	600	560	117	163	59
Gene with annotated alleles	35,525	4,289	9,792	14,797	6,647	/	/
Duplicated genes	4,631	/	/	/	/	3,375	1,256
Unanchored genes/alleles	3,130	/	/	/	/	/	/

FIGURES

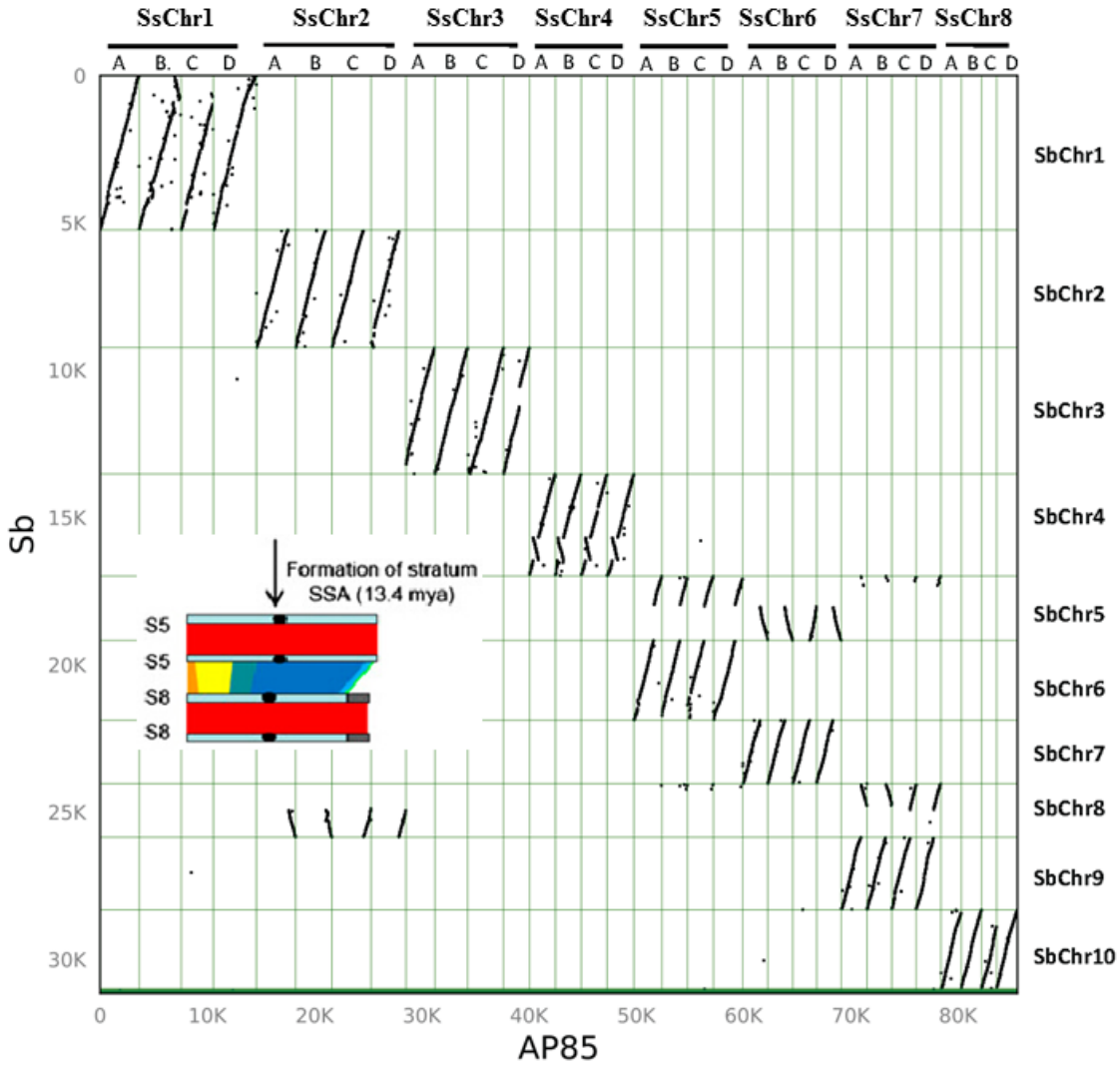


Figure 2.1

Alignment of *S. spontaneum* AP85-441 chromosomes with sorghum chromosomes.

A set of 4 homologous chromosomes aligned to a single sorghum chromosome. The reduction of basic chromosomes from 10 to 8 in *S. spontaneum* is caused by chromosome fissions followed by ranslocations of two ancestral chromosomes homologous to sorghum chromosomes 5 and 8. The inserted figure is part of Fig 3 in (Wang et al., 2011) showing the sorghum stratum SSA, which is mirrored in the alignment of SsChr5 to SbChr8 and SsChr7 to SbChr5 at the tip of the short arm.

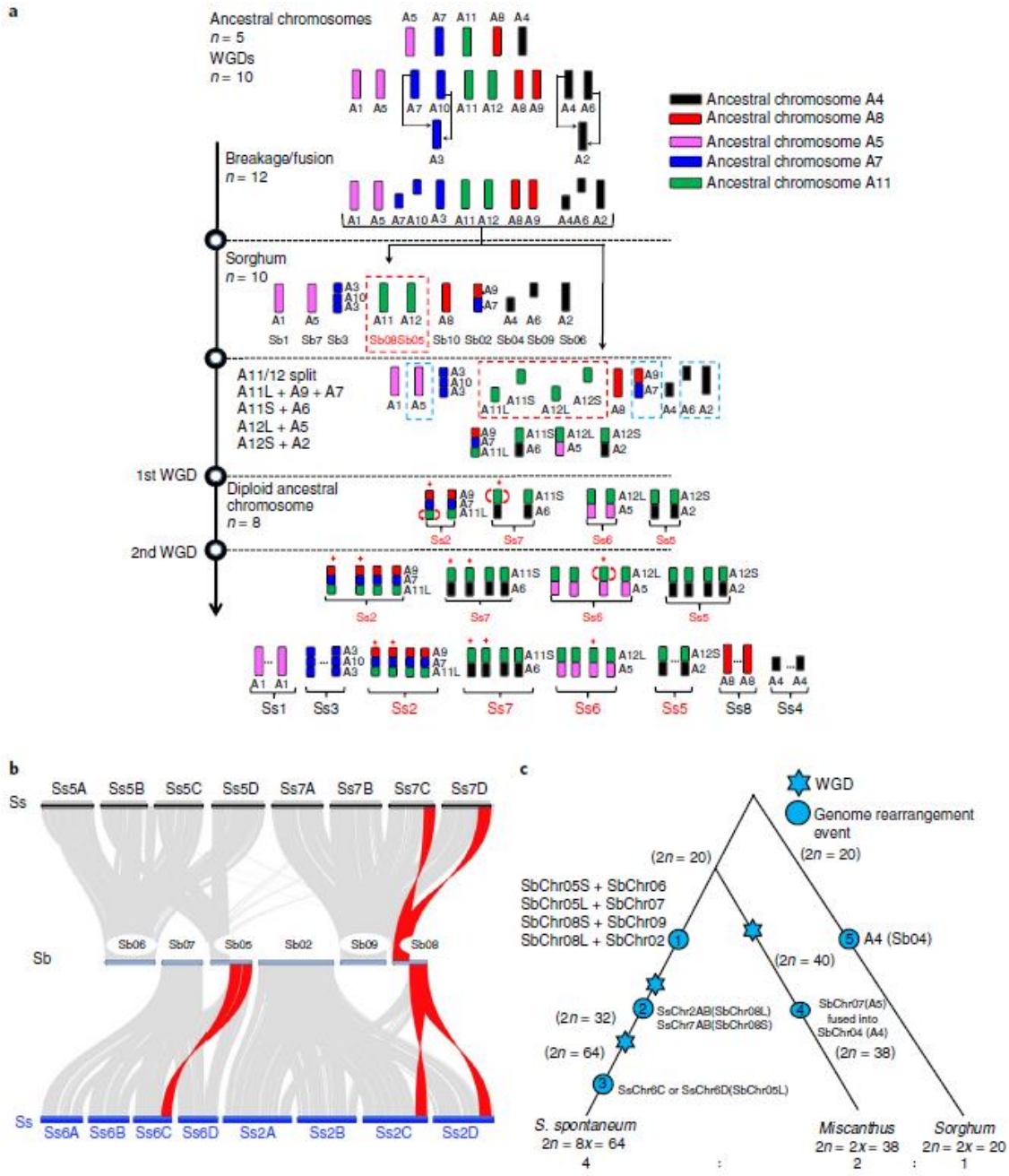


Figure 2.2

Evolutionary history of *S. spontaneum* chromosomes.

A. Evolutionary scenario of Sorghum and *S. spontaneum* from ancestral genome of Poaceae, showing how they evolved from n=10 of sorghum to n=8 of *S. spontaneum*,

Chromosomes are represented with color codes to illuminate the evolution of segments from a common ancestor with five chromosomes. Ancestral genomes were labeled with AX (X is a number from 1 to 12) according to Salse et al., 2008. The *S. spontaneum* re-arranged chromosomes were marked with Dotted box. The inversion events happened in chromosomes segments are shown with red arc, and the chromosomes recombined with inversed chromosome segment are marked with red asterisk. **B**. Genomic alignments between Ss2, Ss5, Ss6, Ss7 alleles and Sb02, Sb05, Sb06, Sb07, Sb08 and Sb09 are shown, the inverted regions are marked with red line. **C**. The genome duplication and rearrangements in *S. spontaneum*, *Miscanthus* (genetic map), and sorghum is shown. The circle and asterisk represent the genome rearrangement and whole genome duplication event, respectively. ① The basic chromosome number reduction from 10 to 8 *S. spontaneum* by two fissions followed by four translocations occurred after the divergence of *Saccharum* and *Miscanthus* and before the two rounds of WGD as described in the text. ② Two inversions occurred after the first round of WGD as shown by pairs of inversions in SsChr2AB (SbChr8L) and SsChr7AB (SbChr8S, Fig 1). ③ Three chromosomal fragments in SsChr6ABD (SbChr5L) are in inverted position and there are two possibilities: a. an inversion occurred before the two rounds of WGD, resulting in all four fragments in inverted position, followed by an inversion in SsChr6C after two rounds of WGD; b. an inversion occurred after the first round of inversion, resulting in SsChr6AB in inverted position, followed by an inversion in SsChr6D. ④ The ancestral chromosome SbChr07 (A5) fused into SbChr04 (A4) after an allopolyploidization event in *Miscanthus* (Swaminathan et al., 2012) resulting in the reduction of basic chromosome number from 20 = 19 . ⑤ Four chromosomal fragments in SsChr4ABCD are in inverted

position, giving the appearance that an inversion occurred before the two rounds of WGD. In reality, this is a sorghum specific inversion in SbChr04 (A4), because the orientation of this large chromosomal fragment is the same in rice, *Miscanthus*, and *S. spontaneum*.

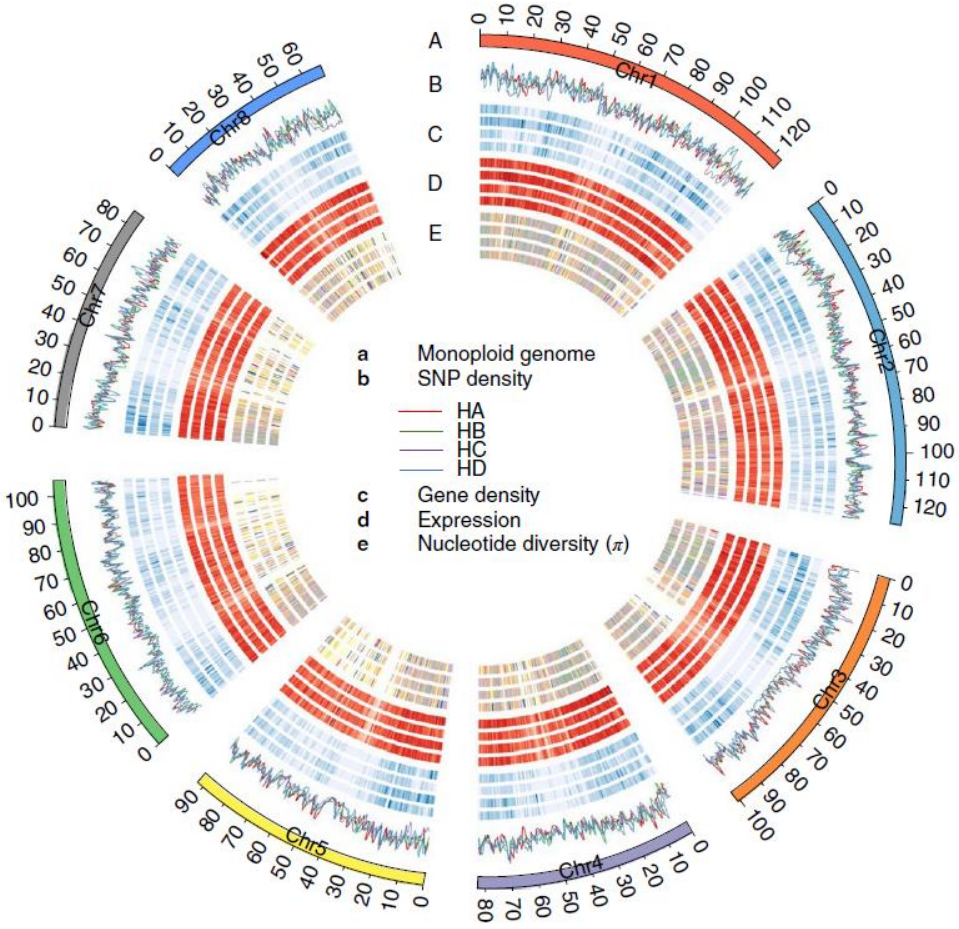


Figure 2.3

Distribution of genomic features along the sugarcane monoploid genome.

The rings indicate (from outermost to innermost) monoploid genome (A), SNP density among haplotypes (B), gene density (C), expression (D) and nucleotide diversity (E). HA, HB, HC and HD indicate four haplotypes in ring B, respectively, and ordered from outside to inside in ring C, D and E.

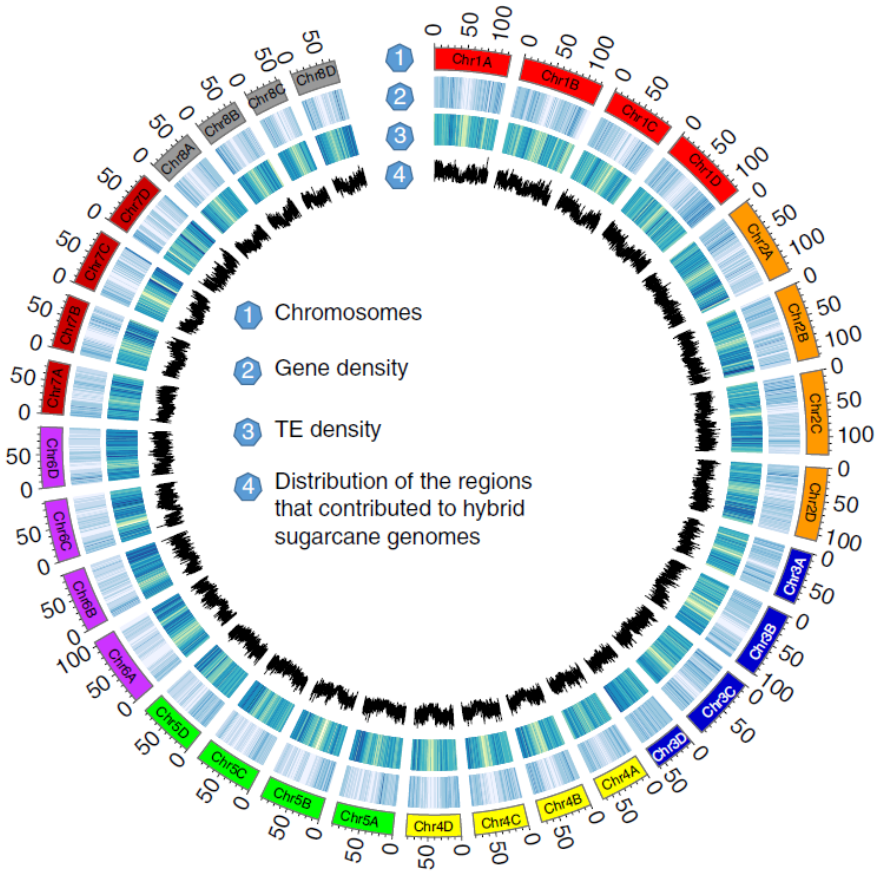


Figure 2.4

Distribution of the regions that contributed to hybrid sugarcane genome in AP85-441.

The rings indicate (from outermost to innermost) 32 pesudo-molecules (1), gene density (2), TE density (3) and distribution of regions that contributed to modern hybrid sugarcane genomes along AP85-441 chromosomes (4).

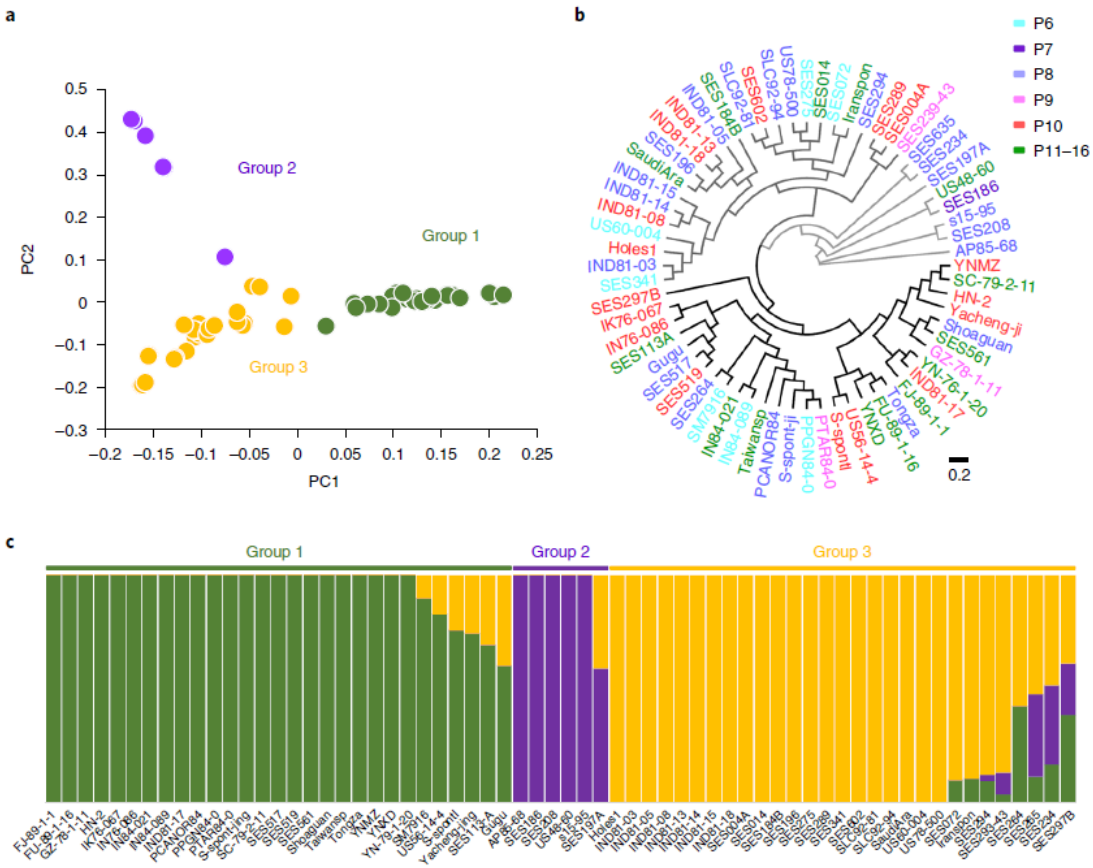


Figure 2.5

Population genetic structure and phylogenetic relationships among 64 *Saccharum spontaneum* accessions.

(a) Principal components of accessions variation. (b) Bootstrapped tree of 64 *S. spontaneum* accessions based on genetic distance. Color bars indicate accessions with different ploidy levels (p6, hexaploid; P7, heptaploid; p8, octoploid; p9, enneaploid; p10, decaploid; p11-p16, from hendecaploid to hexadecaploid). The scale bar shows substitutions per site. (c) ADMIXTURE plot for *S. spontaneum*, showing the distribution of K=3 genetic clusters with the smallest cross-validation error.

Supplemental data:

Supplemental data, figures and table are available at

https://doi.org/10.1038/s41588-018-0237-2

Chapter 3 : Genomic analyses of SUT and TST sugar transporter families in low and high sugar accumulating sugarcane species (*Saccharum spontaneum* and *Saccharum officinarum*)

Note: The information in this chapter was submitted to be published under the title:

Dhungana, S.R., and Braun, D.M. (2022) Genomic comparisons of SUT and TST sugar transporter families in low and high sugar accumulating sugarcane species (*Saccharum spontaneum* and *Saccharum officinarum*). Submitted to Tropical Plant Biology

ABSTRACT

Sugarcane (*Saccharum* spp.) is an economically vital crop that is the primary source of sugar in our food and is also used for ethanol production. Various studies have been performed to understand the molecular mechanisms underlying sugar accumulation in sugarcane, but the complex, polyploid genome has posed many challenges. Here, we analyzed the recently sequenced genomes of two sugarcane varieties: *Saccharum spontaneum*, a geographically widespread, stress tolerant, and low sugar-accumulating wild relative of domesticated sugarcane, and *Saccharum officinarum*, a high sugar-accumulating variety, to investigate the composition and roles of two sugar transporter protein families, Sucrose Transporters (SUTs) and Tonoplast Sugar Transporters (TSTs). We found an increase in the number of alleles for various *SUT* and *TST* genes in *S. officinarum* and *S. spontaneum* compared to sorghum and other grass species, and we identified new putative *TST* genes previously uncharacterized. We analyzed expression levels of these genes in various tissues at different stages and identified SUTs and TSTs likely involved in sugar transport and storage. We hypothesize that these sugar

transporter genes in *S. officinarum*, particularly *TSTs*, contribute to the ability of modern sugarcane hybrid varieties to accumulate large amounts of sugar in the stem.

Keywords:

Saccharum officinarum, *Saccharum spontaneum*, sugar transporters, SUTs, TSTs

Abbreviations:

Chr Chromosome

FPKM Fragments Per Kilobase Million

SUT Sucrose Transporter

SWEETs Sugars Will Eventually be Exported Transporters

TPM Transcripts Per Kilobase Million

TST Tonoplast Sugar Transporter

WGD Whole Genome Duplication

INTRODUCTION

Sugarcane (*Saccharum* spp.) is an economically important crop species and is the major source of sugar used in food, with additional uses in ethanol production (Patrick et al., 2013; Zhang et al., 2016). Many modern sugarcane hybrid cultivars are derived from crossing *S. officinarum* and *S. spontaneum* (Souza et al., 2011). *S. spontaneum* is a wild relative of domesticated sugarcane and grows widespread over a large geographical area throughout the world due to its resistance to diseases and ability to tolerate stressful

conditions (Bremer, 1961; Zhang et al., 2018). *S. officinarum*, on the other hand, is a high sugar-accumulating sugarcane species, which is more susceptible to disease and stress (Bremer, 1961; Zhang et al., 2018). Thus, as a result of hybrid vigor gained by crossing these species, modern varieties can withstand diseases, grow in diverse geographical areas, and accumulate high amounts of sugar in their stems.

In sugarcane leaves, as in other closely related grasses such as maize (*Zea mays*), sorghum (*Sorghum bicolor*), wheat (*Triticum aestivum*), and barley (*Hordeum vulgare*), the conducting phloem tissues of the minor veins are physically isolated from surrounding cells and have few plasmodesmata connecting their symplasts (Bihmidine et al., 2015; Evert et al., 1978; Evert et al., 1996; Robinson-Beers and Evert, 1991; Thompson and Dale, 1981). Hence, sucrose produced in the photosynthetic cells must be exported to the cell wall space (apoplast) by SWEETs (Sugars Will Eventually be Exported Transporters) prior to importing into the companion cell-sieve element complex for long-distance transport (Braun and Slewinski, 2009; Chen et al., 2012; Eom et al., 2015; Hu et al., 2018). Sucrose transporters (SUTs) are hypothesized to load sucrose into the phloem of leaf minor veins, and also function to retrieve sucrose from the apoplast during transport (Baker et al., 2016; Dhungana and Braun, 2021; Julius et al., 2017; Slewinski and Braun, 2010). In sugarcane and sweet sorghum, the stems are the principal sink tissues that store high concentrations of sugars within the parenchyma cells (Bihmidine et al., 2013; Lingle, 1987; Lingle, 1989; Patrick et al., 2013). Sugar transporters such as SUTs, Tonoplast Sugar Transporters (TSTs, previously known as Tonoplast Monosaccharide Transporters, TMTs), and SWEETs regulate the distribution of sugars in plants, and likely play critical roles in unloading sucrose from the transport

phloem and storing it in the stem (Babst et al., 2021; Bihmidine et al., 2015; Bihmidine et al., 2016; Braun, 2022; Casu et al., 2015; Chen, 2014; Dhungana and Braun, 2021; Eom et al., 2015; Julius et al., 2017; Jung et al., 2015).

With the recent sequencing of multiple sugarcane genomes and BAC libraries (Figueira et al., 2012; Garsmeur et al., 2018; Grativol et al., 2014; Zhang et al., 2018), we are gaining a better understanding regarding the complexity of the sugarcane pan-genome, its ancestry, and chromosome rearrangements. Limited studies have investigated the expression and functions of a number of sugar transporters in sugarcane hybrid species and progenitor species, but the functions of those involved in sugar storage in the stems remain to be well explored (Casu et al., 2003; Casu et al., 2015; Glassop et al., 2017; Rae et al., 2005; Reinders et al., 2006; Zhang et al., 2016). A recent study investigating a large set of sugar transporter genes in *S. spontaneum* and *S. officinarum* identified some *SUTs* and *TSTs* as likely candidates for sugar accumulation and transport (Zhang et al., 2021b). Similarly, another study found *TST* genes to be highly expressed in the stems of sugarcane and energy cane (Wang et al., 2021). Furthermore, Hu et al. (2018) identified a number of *SWEET* genes with different expression patterns in *S. spontaneu*m and *S. officinarum*, and established *SWEET13c* to be involved in the efflux of sugars in mature photosynthetic tissues in both sugarcane species. Here, we analyzed the genome sequences of *S. spontaneum* cultivar AP85-441 (1n=4x=32) (Zhang et al., 2018) and *S. officinarum* cultivar LA Purple (2n=8x=80) (R. Ming lab, manuscript in revision) to identify the composition of the *SUT* and *TST* gene families, which play diverse roles in sugar transport in many grass species (Braun, 2022; Dhungana and Braun, 2021). We performed phylogenetic analyses to compare gene families between

species and syntenic comparisons with a closely related species, sorghum, to identify *SUT* and *TST* genes in both sugarcane genomes. Additionally, we examined RNA expression data from various stem and leaf tissues across multiple growth stages in both sugarcane varieties to study the expression levels of these genes to infer their functions in relation to sugar partitioning.

RESULTS

Phylogenetic analysis of SUT proteins

A phylogenetic tree was constructed to investigate the evolutionary relationships between the identified sugarcane SUT protein sequences from both *Saccharum* species with SUTs from other closely related grass species and several eudicots (Figure 3.3.1). Based on the SUT phylogenetic analysis, five distinct SUT groups (clades) were identified. The sugarcane SUT proteins were all classified into groups previously found to contain grass SUT proteins, in agreement with earlier phylogenetic studies (Braun and Slewinski, 2009; Sauer, 2007; Zhang et al., 2016). Specifically, groups 1 and 5 were found only in monocot species, group 2 SUTs consisted only of eudicot species, and groups 3 and 4 consisted of both monocots and eudicot SUT proteins. Interestingly, for all sequences of a particular SUT group in both *S. spontaneum* and *S. officinarum*, we did not observe any significant species-specific clustering: all *S. spontaneum* and *S. officinarum* alleles for a particular clade are intermixed and did not show clear species-specific separation. Additionally, for all the sugarcane SUT sequences, as anticipated, the closest related sequences were from sorghum. We identified a base number of six genes in the sugarcane *SUT* gene family, which is the same number as in sorghum (Table 3.1).

However, we found large numbers of alleles for each of the six sugarcane *SUT* genes, whereas with sorghum and maize, there is only one allele for each of these (Supp. Table S3.1) (Babst et al., 2021; Bihmidine et al., 2015; Braun and Slewinski, 2009; Braun et al., 2014). Based on the phylogeny, and manual sequence analyses and curation, we identified a total of 19 alleles in *S. spontaneum* and 34 alleles in *S. officinarum* for the *SUT* gene family. Five of these *SUT* genes had larger number of alleles in *S. officinarum* compared to *S. spontaneum*. In *S. spontaneum*, we found two alleles each for *SsSUT1* and *SsSUT6*, three alleles for *SsSUT4*, and four alleles each for *SsSUT2*, *SsSUT3*, and *SsSUT5*. Similarly, in *S. officinarum*, we found four alleles for *SoSUT2* and *SoSUT6*, five alleles for *SoSUT3*, six alleles for *SoSUT1*, seven alleles for *SoSUT4*, and eight alleles for *SoSUT5*. Thus, *SUT5* had the greatest increase in the number of alleles, with *S. officinarum* having four more *SUT5* alleles than *S. spontaneum*. All SUT sequences were predicted to have 12 transmembrane spanning domains that are commonly found in sugar transporters belonging to the Major Facilitator Superfamily (Lalonde et al., 2004; Sauer, 2007).

Phylogeny of TST proteins

A phylogenetic analysis of the TST protein sequences was performed using orthologous TST sequences from the same species as with the SUT phylogeny (Figure 3.3.2). All SsTST and SoTST sequences were clustered into three monocot-specific TST clades. The monocot TST1, TST2, and TST3 clades were clearly resolved from eudicot TST members. Interestingly, we identified two different members each in the TST2 clade, TST2a and TST2b, and also in the TST3 clade, TST3a and TST3b, in both *Saccharum* species in agreement with a previous report (Wang et al., 2021). The number

of genes in the TST1 clade remained similar to the TST families in sorghum, maize, rice (*Oryza sativa*), *Setaria italica*, and *Brachypodium distachyon*. The TST2a sequences in both species were closely related to SbTST2 and ZmTST2, whereas both the TST2a and TST2b sequences for both sugarcane species were equally related to rice OsTMT2 and *Brachypodium* BdTST2 sequences (Cho et al., 2010). Similarly, the TST3a sequences from both sugarcane species were closest to the sorghum and maize TST3 sequences, whereas TST3b sequences from both sugarcane species were closer to BdTST3 sequences. We found a large increase in the number of alleles for each of these TST clades in sugarcane, while for the other grasses investigated there is only one allele per TST clade (Supp. Table S3.1) (Bihmidine et al., 2016; Cho et al., 2010).

Similarly, for the *TST* gene family, we identified 16 alleles in *S. spontaneum* and 25 in *S. officinarum* (Table 3.2). In *S. spontaneum*, we identified four alleles of *SsTST1*, two alleles each of *SsTST2a* and *SsTST2b*, five alleles of Ss*TST3a*, and three alleles of *SsTST3b*. Similarly, in *S. officinarum*, we identified seven alleles of *SoTST1*, eight alleles of *SoTST2a*, two alleles of *SoTST2b*, five alleles of *SoTST3a*, and three alleles of *SoTST3b*. Thus, we identified new *TST* genes in both sugarcane species. Similar to the SUTs, all the TST protein sequences were predicted to have 12 transmembrane domains.

Expression of *SUTs* and *TSTs* in various tissues at different growth stages

To gain further insights into potential functions of these sugar transporter genes, we examined RNA sequencing data from a collection of tissues across various developmental stages from both species. In *S. spontaneum*, we found *SsSUT1* (both *SsSUT1-B* and *SsSUT1-D* alleles) to be highly expressed in almost all tissues with high expression in seedling and mature leaves and in older internodes (internodes 6 and 9) at

both mature and premature stages (Figure 3.3). Most of the *SsSUT2* and *SsSUT4* alleles were moderately expressed in all tissues at all growth stages. *SsSUT5* alleles were expressed at low levels in seedling stems and young internodes (internode 3) at the mature stage. *SsSUT3* expression was not detected, and *SsSUT6* was expressed at extremely low levels. Similarly, *SsTST1* and *SsTST2a* were well expressed in almost all tissues with high expression levels in mature and premature stem internodes and moderate expression in seedling stems. *SsTST1* and *SsTST2a* levels peaked progressively as the internodes aged for samples collected at both premature and mature stages. Interestingly, *SsTST2b* had high expression in intermediate and older stem internodes (internodes 6 and 9) in both premature and mature stages. *SsTST3b* was expressed at very low levels in seedlings and premature and mature stems. *SsTST3a* expression was not detected.

In the case of *S. officinarum*, most of the alleles of *SoSUT1*, *SoSUT2* and *SoSUT4* were expressed at moderate levels in almost all tissues examined (Figure 3.4). The expression levels for *SoSUT1* alleles varied, but the majority had their highest expression in sclerenchyma cells of mature stem internodes (internode 13). Overall, the expression of *SoSUT1* alleles was high in leaf rolls and leaves at all stages (seedling, premature and mature) and moderate to low in stem internodes at all stages. Similarly, *SoSUT2* alleles had modest expression in all internodes at all stages and in premature and mature leaf rolls and low expression in premature and seedling leaves. For the majority of *SoSUT4* alleles, expression levels were highest in parenchyma and sclerenchyma cells of mature internodes and moderate in seedling stem and leaves. *SoSUT3* was expressed at very low or undetectable levels across all of the tissues analyzed. *SoSUT5* had low expression in

seedling stems and very low expression in stems at mature and premature stages, whereas *SoSUT6* was very lowly expressed. The majority of *SoTST1*, *SoTST2a* and *SoTST2b* alleles were expressed in almost all tissues at some levels (Figure 3.4). *SoTST1* alleles had their highest expression levels in parenchyma and sclerenchyma cells and displayed high levels in older internodes (internodes 9 and 15) at both mature and premature stages, with expression increasing progressively from the younger-to-older internodes. Moderate to low levels of *SoTST1* were found in seedling stems and leaf tissues across various stages. Similarly, *SoTST2a* and *SoTST2b* exhibited the highest expression in parenchyma cells of mature internodes, showed high expression in sclerenchyma cells of mature internodes, and had moderate expression in stems across seedling, premature and mature stages. For most of the *SoTST2a* alleles, RNA levels were comparatively higher in younger internodes (internode 3) and lower in older internodes (internode 15) in both mature and premature stages. Interestingly, *SoTST2b* had comparatively higher expression levels for intermediate internodes (internode 9) compared to other internodes at both mature and premature stages. *SoTST3b* showed very low expression levels in parenchyma and sclerenchyma cells of mature internodes and in stems. As in *S. spontaneum*, *SoTST3a* expression was not detected. The expression levels of major transcripts in various tissues are summarized in Figure 3.5.

Synteny of sugar transporter genes

In *S. spontaneum* the chromosome number has been reduced from the basic number of ten chromosomes commonly found in grasses to eight chromosomes, and there is evidence for chromosomal rearrangements (Garsmeur et al., 2018; Zhang et al., 2018). Based on the synteny maps between *S. spontaneum* and *S. bicolor* (Zhang et al., 2018),

the chromosomal locations of the *SUTs* and *TSTs* genes are conserved. The naming scheme for *SUTs* used here assigns names to sorghum (*SbSUT*) and maize (*ZmSUT*) sequences (Bihmidine et al., 2015; Braun and Slewinski, 2009) based on their orthology to rice (*OsSUT*) sequences (Aoki et al., 2003), as well as respecting precedence for sugarcane *SUTs* based on orthology to already named *S. spontaneum* sequences (Zhang et al., 2018; Zhang et al., 2016). All *SUT1, SUT3,* and *TST1* alleles are present in regions of chromosome (Chr) 1 for both species. Similarly, all *SUT5* alleles are located on Chr4 for both species. All *SsSUT4* alleles are in a region of Chr2, which has synteny to SbChr8, where *SbSUT2* (group 4) is located. Similarly, all *SsSUT2* alleles are located on Chr4 where *SbSUT4* (group 3) is located. The *SsSUT6* alleles are located on Chr6, which shares synteny with SbChr7, where *SbSUT6* is located. All *SsTST2a* and *SbTST2* alleles are located on Chr4, whereas *SsTST2b* alleles are located on Chr3A and an unanchored contig. The *SsTST3a* alleles reside on Chr8, which has synteny with SbChr10, where *SbTST3* is located, whereas *SsTST3b* alleles are located on Chr4. We found the presence of two *SsSUT2* sequences (*SsSUT2-D1* and *SsSUT2-D2*) on Chr4D in close proximity (Table 3.1), two *SsTST3a* sequences (*SsTST3a-C1* and *SsTST3a-C2*) near each other on Chr8C, and two *SsTST3b* sequences (*SsTST3b-D1* and *SsTST3b-D2*) in close proximity on Chr4D (Table 3.2), which suggest the recent tandem duplication of these genes.

However, in *S. officinarum,* the basic chromosome number is maintained similar to *S. bicolor*, containing 10 chromosomes with eight sets for each. The chromosomal locations of *SUTs* and *TSTs* are highly conserved between *S. officinarum* and *S. bicolor*. In both *S. bicolor* and *S. officinarum*, *TST1, SUT1,* and *SUT3* are located on Chr1. Similarly, *SoSUT5, SoSUT2,* and the majority of *SoTST2a* alleles are located on Chr4,

consistent with the location of the orthologous genes in *S. bicolor*. One copy of the *SoTST2a* gene (*SoTST2a-C2*) is located on Chr5C, which could possibly result from a gene duplication event. *SoTST2b* sequences are located on Chr10, where the majority of *SoTST3a* sequences are also located. *SoSUT6* and *SbSUT6* are located on Chr7 in both species. Similarly, *SbSUT2* and *SoSUT4* genes are located on Chr8 in both species. In the case of *S. officinarum*, we hypothesize that many gene duplication events occurred that led to the expansion of the *SUT* and *TST* gene families. For instance, Chr8B contains three copies of the *SoSUT4* gene (*SoSUT4-B1*, *SoSUT4-B2*, and *SoSUT4-B3*) located very close to one another. Similarly, Chr4A, contains three copies of the *SoTST2a* gene (*SoTST2a-A1*, *SoTST2a-A2*, *SoTST2a-A3*) and Chr4B contains two copies additional copies (*SoTST2a-B1* and *SoTST2a-B2*). Additionally, Chr1C and Chr1H contain two copies each of the *SoSUT1* gene (*SoSUT1-C1* and *SoSUT1-C2*, and *SoSUT1-H1* and *SoSUT1-H2* respectively) located close to one another, and on Chr1F, three copies of the *SoSUT3* gene (*SoSUT3-F1*, *SoSUT3-F2*, and *SoSUT3-F3*) are positioned close to each other. Similarly, Chr4B contains two copies of the *SoSUT5* gene (*SoSUT5-B1* and *SoSUT5-B2*). These multiple copies of closely positioned genes suggest that many gene duplication events have occurred that led to the expansion of these sugar transporter gene families.

DISCUSSION

Sugarcane species possess complex polyploid genomes, and thus, resolving their gene families can be challenging (Garsmeur et al., 2018; Grativol et al., 2014; Zhang et al., 2018). With recent developments in DNA sequencing and computation, however, whole genome sequencing approaches have provided opportunities to resolve the

composition of individual members of gene families. Here, we identified the *SUT* and *TST* sugar transporter gene families in wild sugarcane (*S. spontaneum*), which is widespread but low in sugar content, and in domesticated high sugar-accumulating *S. officinarum*. These investigations help clarify the genomic composition of these families, which may lead to increased understanding of their biological relevance in cultivated sugarcane as a high sugar-accumulating plant species.

Several studies have been performed to uncover the mechanisms of sucrose accumulation in sugarcane hybrid species, with ShSUT1 (which has more than 96% amino acid identity to SsSUT1 and SoSUT1) being the most well studied of the sugarcane sugar transporters (Casu et al., 2003; Casu et al., 2015; Glassop et al., 2017; Rae et al., 2005; Reinders et al., 2006; Zhang et al., 2016). Unlike other grass species, such as maize, rice, and sorghum, where SUT1 was shown to be expressed in the phloem (Baker et al., 2016; Milne et al., 2017), ShSUT1 was not expressed in the phloem, but in the mestome sheath and vascular parenchyma cells (Rae et al., 2005). SUT1 in maize and sorghum localized to the plasma membrane, and in maize was shown to be responsible for phloem-loading in the veins of source leaves (Babst et al., 2022; Baker et al., 2016; Milne et al., 2017; Rotsch et al., 2015; Slewinski et al., 2009; Tran et al., 2017). A recent study of the role of ShSUT1 in sucrose mobilization by using RNAi suppression found that ShSUT1 does not play a direct role in loading sugars in the phloem but is likely functional in retrieving sucrose from the apoplast of leaf and stem tissues (Glassop et al., 2017). Additionally, *ShSUT1* expression was high in sucrose-exporting mature leaves and in sucrose-accumulating internodes (Rae et al., 2005). Similarly, *SoSUT1* and *SsSUT1* were expressed moderately in almost all tissues with high expression in leaves and in leaf

rolls at all developmental stages and in intermediate and older internodes at premature and mature stages. *SoSUT1* had the highest expression in sclerenchyma cells of mature internodes, which denotes that it might function to retrieve leaked sucrose from the apoplast in these tissues. However, in parenchyma cells of mature internodes, only low levels of *SoSUT1* were detected. This difference in expression in parenchyma cells could be due to the difference in the internode sampled (internode 5 vs 13) or the difference in species (hybrid vs *S. officinarum*) in these two experiments. Collectively, these results suggest that *ShSUT1, SoSUT1* and *SsSUT1* likely function in retrieval of sucrose from the apoplast. Additionally, in contrast to *ShSUT1,* it is possible that another sugarcane *SUT1* homeolog may be expressed in the phloem and function to load sucrose into the companion cells in mature leaf veins as shown for maize and sorghum (Baker et al., 2016; Bihmidine et al., 2015; Slewinski et al., 2009). Future work investigating the cell-specific expression patterns of all of the sugarcane *SUT1* sequences is needed to clarify their functions.

ShSUT4 was reported to be upregulated in storage parenchyma cells in *Saccharum* hybrid species (Casu et al., 2015). ZmSUT2, which is orthologous to ShSUT4, SsSUT4, and SoSUT4, localized to the tonoplast and is hypothesized to efflux sucrose out of the vacuole (Leach et al., 2017), suggesting SsSUT4 and SoSUT4 likely play similar roles in sugarcane vacuoles. In a study comparing the levels of *SUT* gene expression in mature and seedling plants of sugarcane, *SsSUT1* and *SsSUT4* accounted for more than 70% of the transcripts observed for the entire *SUT* gene family (Zhang et al., 2016). Furthermore, *SsSUT4* expression remained at similar levels in seedling and mature tissues, suggesting that it contributes to sucrose accumulation throughout plant

development (Zhang et al., 2016). Similarly, in the RNA sequencing data analyzed here, *SoSUT4* levels were highest in parenchyma and sclerenchyma cells of mature internodes in *S. officinarum*. Moderate levels of *SoSUT4* and *SsSUT4* were found in stems and leaves at the premature, mature, and seedling stages in both *Saccharum* species, further suggesting that SUT4 contributes to sugar accumulation across various developmental stages. The biological functions of the remaining sugarcane *SUTs* remain to be determined. The above results suggest that various SUTs function in sucrose mobilization and storage in the stem of *Saccharum* species, with the expression of each *SUT* gene depending on the developmental phase.

Similarly, members of the *TST* gene family have been identified in various grass species, including sugarcane hybrid species. Both of the previously studied sugarcane hybrid ShPST2a and ShPST2b proteins, which are closely related to SsTST2b/SoTST2b (99% amino acid identity) and SsTST1/SoTST1 (97% amino acid identity), respectively, are strongly expressed in storage parenchyma cells and are hypothesized to localize to the tonoplast where they are proposed to function in loading sucrose into the vacuole (Casu et al., 2015). ShPST2a and ShPST2b may play separate roles during development as the expression of *ShPST2a* increased through stalk development, whereas *ShPST2b* was expressed uniformly throughout development (Casu et al., 2015). In *S. spontaneum*, we observed *SsTST1* and *SsTST2a* to be highly expressed in internodes with an increasing trend moving from younger to older internodes at both the premature and mature stages. We observed higher expression levels of *SsTST2b* in intermediate and older internodes compared to younger internodes at both premature and mature stages as well. In *S. officinarum*, we observed *SoTST1*, *SoTST2a*, and *SoTST2b* to be highly expressed in

parenchyma cells, with *SoTST1* showing progressively increased expression in younger to older internodes at both mature and premature stages. *SoTST2a* had relatively higher expression in younger internode (internode 3) at both mature and premature stages, whereas *SoTST2b* had relatively higher expression in intermediate internodes (internode 9) at both mature and premature stages compared to other internodes. Such changes in expression levels in specific tissues at specific stages could be a result of specialization of the TSTs, especially in relation to *TST2a* and *TST2b* genes, as their function might be redundant. Interestingly, in a recent study, Wang et al. (2021) identified *TST1* and *TST2b* to be strongly expressed in older stems of sugarcane and energy cane at the immature stage and mature stages. In addition, using a promoter reporter gene assay, the promoters of two *SsTST2b* alleles drove *GUS* expression in Arabidopsis stems, denoting their specificity for expression in stem tissues (Wang et al., 2021). Moreover, SsTMT3 (also known as SsTST1-A) and SsTMT4 (SsTST2a-A) were hypothesized to be involved in sugar accumulation in parenchyma cells in another recent study (Zhang et al., 2021b). Based on these observations, TST1 and TST2b likely function in storing sucrose and other sugars in stems in both sugarcane species. Thus, in S. *officinarum*, SoTST1, SoTST2a, and SoTST2b putatively play important roles in accumulation of sucrose and other sugars in parenchyma cells in the mature stems, contributing to their high sugar accumulation. Similarly, in sorghum, *SbTST1* and *SbTST2* have very strong expression in leaves and mature internodes throughout development (Bihmidine et al., 2016). Also, in a comparison of a sweet versus a grain sorghum, both of these *SbTST* genes were highly expressed in sweet sorghum stems (Bihmidine et al., 2016; Milne et al., 2017). Summing

up, we conclude that TST1, TST2a, and TST2b underpin sugar storage in the vacuoles of mature stems in sugarcane.

For both sugarcane species, we identified an increase in the number of *TST2* genes compared to sorghum and other grass species. The TST2b protein sequences are not as closely related to sorghum and maize TST amino acid sequences as the TST2a sequences are, however, both sugarcane TST2a and TST2b sequences are equally related to rice OsTMT2 and *Brachypodium* BdTST2 sequences in the phylogenetic tree. We hypothesize that both sugarcane species retained the two *TST2* genes, and that one each was lost in sorghum, maize, and *Setaria* (missing *TST2b*), and rice and *Brachypodium* (missing *TST2a*) at different points in time after their divergence from their common ancestors. Another possibility is that *TST2b* is the ancestral gene, which evolved into *TST2a* in sorghum, maize and *Setaria*. However, in this scenario, sugarcane retained both the ancient *TST2b* gene and the more recent *TST2a* gene. We also observed more alleles of *TST2a* genes and only a few alleles of *TST2b* genes in sugarcane, however the expression of *TST2b* genes was very high in certain tissues (i.e., in parenchyma cells in mature *S. officinarum* internodes and in premature and mature internodes in both *S. spontaneum and S. officinarum*). This differing expression pattern of *TST2a* and *TST2b* might represent the sub-functionalization of genes in the TST2 family in sugarcane. The increase in the total number of *TST* alleles identified in the *S. spontaneum* and *S. officinarum* genomes compared to other grasses likely underlies the increased storage of sucrose in the storage parenchyma of modern sugarcane hybrids. Further research to characterize the expression and functions of these genes is necessary to test this hypothesis.

Recent studies suggest that these two sugarcane species shared a common ancestor with sorghum (2n=20), and that once these species diverged, whole genome duplications (WGD) events and large chromosome rearrangements (in the case of *S. spontaneum*) occurred at separate intervals giving rise to different numbers of chromosomes and physical characteristics that we see in present day varieties (Garsmeur et al., 2018; Zhang et al., 2021a; Zhang et al., 2018). Specifically, in the *S. spontaneum* genome, it was hypothesized that genomic rearrangements and two rounds of WGD autopolyploidization occurred in a relatively short time frame after the divergence from sorghum, giving rise to an octoploid with eight basic sets of chromosomes (2n=8x=64) (Zhang et al., 2021a; Zhang et al., 2018). It is further hypothesized that two more rounds of WGD events also occurred in the *S. officinarum* lineage after its divergence from *S. spontaneum*, resulting in an octoploid with ten basic sets of chromosomes (2n=8x=80) (Sharma et al., 2018; Vilela et al., 2017; Zhang et al., 2021a). In support of this idea, a recent study which combined these two species to create a mosaic monoploid reference genome for *Saccharum* hybrid species reported that polyploidization in these species occurred after the divergence from sorghum (Garsmeur et al., 2018). These chromosomal rearrangements and whole genome duplication events likely account for the increased copy numbers of *SUT* and *TST* genes in *Saccharum* compared to sorghum and other analyzed grass genomes.

Finally, we note that the stringent selection criteria we used for manually analyzing the *SUT* and *TST* gene sequences resulted in the omission of 11 *SUT*-related sequences (5 in *S. spontaneum* and 6 in *S. officinarum*) and 4 *TST*-related sequences (all in *S. officinarum*) (Supp. Table S3.1). These were not included in the analyses above

because they were not predicted to be full length proteins or were merged with other gene models. However, it is possible that a subset of these, which are predicted to encode in-frame deletions of 25 or more contiguous amino acids, conceivably may be functional proteins, and therefore, our gene counts may be underestimates. As with all early version draft genomes, this assembly might contain errors, which could have caused the predicted protein sequences for certain genes to be incomplete. As newer versions of these genomes become available, the improved assembly might provide complete sequences for these genes. Future research will be needed to evaluate the functionality of such sequences.

In conclusion, sugar transporters, such as SUTs and TSTs, play major roles in the accumulation and distribution of sugars, primarily sucrose, from the photosynthetic leaf tissues to developing and storage tissues, such as stems. The present work defines the genomic composition of *SUTs* and *TSTs* in two *Saccharum* species, which will foster greater understanding of the molecular and genetic mechanisms regulating sucrose levels in cultivated sugarcane, other food, forage, and bioenergy grasses, and additional crop species. Such knowledge will enable us to devise methods to increase yield and plant biomass, which will contribute to enhanced food security and biofuel production.

METHODS

Genomic annotation of *S. spontaneum* and *S. officinarum* SUTs and TSTs:

The predicted peptide sequences for the recently published *S. spontaneum* (1n=4x=32) genome assembly (Zhang et al., 2018) were subjected to BLAST analysis against known sorghum SUT and TST peptide sequences with an e-value cut off of 10^{-6}. All BLAST hits for the SUTs were further analyzed (all were well above this threshold,

with no other sequences close to the cut-off), whereas only those hits that had at least 70% identity to SbTSTs were analyzed further. Those sequences with percentage identify scores below 70% were classified by subsequent BLAST analyses to be either members of other gene families (e.g., hexose transporters) or non-functional TSTs, which are likely pseudogenes because they were missing crucial domains and were substantially shorter in amino acid length. For the initial pass, all sugarcane sequences that satisfied the criteria above were aligned against the reference sorghum sequences using MUSCLE (Edgar, 2004). For *S. spontaneum* sequences that were either too long or too short when compared to the reference sorghum SUTs and TSTs, the corresponding DNA sequences were analyzed using GeneWise to check for frameshift or prediction errors, manually corrected, and then used for analysis (Birney et al., 2004). Closely grouped sequences were further analyzed for alignment using MUSCLE, and truncated sequences or sequences that did not align well with other proteins in the same group were discarded (Julius et al., 2021). We used stringent criteria and omitted predicted protein sequences that had a contiguous block of 25 or more amino acids missing when compared to the reference sorghum sequences, as we considered these likely pseudogenes (Supp. Table S3.1).

Similarly, for the *S. officinarum* (2n=8x=80) peptide sequences, parallel steps as described above were used. For identifying sequences with similarity to known SUTs and TSTs, the predicted protein sequences were subjected to BLAST analysis against selected monocot (sorghum, maize, rice and *S. spontaneum*) SUTs protein sequences using a cut-off e-value of 10^{-6}. All such sequences were aligned against reference sequences from various monocot species and compared. Sequence that had truncated or missing amino

acid sequences were manually checked by comparing the genomic DNA sequences against the expected SUT protein sequence that the gene would be expected to encode. For such sequences, the gene coordinates were obtained and DNA sequences from 5000 bp upstream of the predicted start site to 5000 bp downstream of the predicted stop site were taken and manually checked in Genewise. If sequences still did not show a matching sequence to known SUTs, the reverse complement of the DNA sequence was analyzed in Genewise to account for genes on the reverse strand. If no matching sequences were found with Genewise analysis, gene prediction was performed using FGENESH. In addition to truncated or rearranged sequences, we used stringent criteria and omitted predicted protein sequences that had a contiguous block of 25 or more amino acids missing when compared to the reference sorghum or *S. spontaneum* sequences, as we considered these likely to be pseudogenes.

Phylogeny of SUTs and TST gene families:

Phylogenetic trees were constructed with the curated SUTs and TSTs amino acid sequences using orthologous sequences from other plant species: sorghum (*Sorghum bicolor*), maize (*Zea mays*), rice (*Oryza sativa*), *Setaria italica, Arabidopsis thaliana, Solanum tuberosum, Brachypodium distachyon, Vitis vinifera, Glycine max, Medicago truncatula, Saccharum hybrid species, Beta vulgaris,* and a red algae *Galdieria sulphuraria* (as the outgroup)(Guindon et al., 2010). Alignment of sequences were performed using MUSCLE (Edgar, 2004). The phylogenetic tree was constructed using the maximum likelihood method implemented in the LG amino acid substitution model (Le and Gascuel, 2008) in MEGAX software (Kumar et al., 2018). Initial trees for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ

algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The reliability of the internal branches was inferred from 1000 bootstraps performed using the MEGAX software (Kumar et al., 2018) and the tree was visualized in Interactive Tree Of Life (http://itol.embl.de) (Letunic and Bork, 2021). Protein domains were analyzed using CCTOP Prediction Server (Dobson et al., 2015) which shows a prediction summary from 10 different programs. Each SUT and TST sequence were analyzed for 12 transmembrane domains as predicted by one or more of the programs. All accession IDs are available in Supplemental Table S3.1.

Analysis of RNA sequencing data:

Expression levels for *SUT* and *TST* genes were obtained from published RNA sequencing data for *S. spontaneum* (Zhang et al., 2016; Zhang et al., 2021b) at http://sugarcane.zhangjisenlab.cn/sgd/html/index.html. Genes that had zero expression values for all tissues were omitted from the heatmaps. To prevent a division by zero error, a pseudocount of +1 was added to the Fragments Per Kilobase Million (FPKM) values. The heatmap was constructed using the Bioconductor package pheatmap in R. The log2 of pseudocount-corrected FPKM was plotted and the columns were organized using hierarchical clustering based on distance mapping. Expression levels for *SUTs* and *TSTs* were obtained from RNA sequencing data for *S. officinarum* (R. Ming lab, manuscript in revision). The same settings as described for *S. spontaneum* were used for preparing the heatmap for *S. officinarum* except Transcripts Per Kilobase Million (TPM) values were used instead of FPKM values.

Acknowledgements

We thank Xiaodan Zhang and Ray Ming for kindly sharing the expression data for *S. officinarum SUTs* and *TSTs*. The research was supported by a US National Science Foundation Plant Genome Research Program grant (IOS-1025976) and a US Department of Energy, Office of Science, Office of Biological and Environmental Research grant (DE-SC0018072) to DMB.

Funding and/or Conflicts of interests/Competing interests

No conflict of interest is declared.

References:

Aoki, N., Hirose, T., Scofield, G.N., Whitfeld, P.R., Furbank, R.T., 2003. The sucrose transporter gene family in rice. Plant Cell Physiol 44(3), 223-232.

Babst, B.A., Braun, D.M., Karve, A.A., Baker, R.F., Tran, T.M., Kenny, D., Rohlhill, J., Knoblauch, J., Knoblauch, M., Lohaus, G., Tappero, R., Jensen, K.H., 2022. Sugar loading is not required for phloem sap flow in maize plants. Nat Plants, in press.

Babst, B.A., Karve, A., Sementilli, A., Dweikat, I., Braun, D.M., 2021. Physiology and whole-plant carbon partitioning during stem sugar accumulation in sweet dwarf sorghum. Planta 254(4), Article No. 80.

Baker, R.F., Leach, K.A., Boyer, N.R., Swyers, M.J., Benitez-Alfonso, Y., Skopelitis, T., Luo, A., Sylvester, A., Jackson, D., Braun, D.M., 2016. Sucrose transporter *ZmSut1* expression and localization uncover new insights into sucrose phloem loading. Plant Physiol 172, 1876-1898.

Bihmidine, S., Baker, R.F., Hoffner, C., Braun, D.M., 2015. Sucrose accumulation in sweet sorghum stems occurs by apoplasmic phloem unloading and does not involve differential *Sucrose transporter* expression. BMC Plant Biol 15, 186 DOI 110.1186/s12870-12015-10572-12878.

Bihmidine, S., Hunter III, C.T., Johns, C.E., Koch, K.E., Braun, D.M., 2013. Regulation of assimilate import into sink organs: Update on molecular drivers of sink strength. Front Plant Sci 4, 177. doi: 110.3389/fpls.2013.00177.

Bihmidine, S., Julius, B.T., Dweikat, I., Braun, D.M., 2016. *Tonoplast Sugar Transporters (SbTSTs)* putatively control sucrose accumulation in sweet sorghum stems. Plant Sig Behav 11(1), e1117721.

Birney, E., Clamp, M., Durbin, R., 2004. GeneWise and Genomewise. Genome Res 14(5), 988-995.

Braun, D.M., 2022. Phloem loading and unloading of sucrose: What a long, strange trip from source to sink. Annu Rev Plant Biol 73, in press.

Braun, D.M., Slewinski, T.L., 2009. Genetic control of carbon partitioning in grasses: Roles of *Sucrose Transporters* and *Tie-dyed* loci in phloem loading. Plant Physiol 149, 71-81.

Braun, D.M., Wang, L., Ruan, Y.-L., 2014. Understanding and manipulating sucrose phloem loading, unloading, metabolism, and signalling to enhance crop yield and food security. J Exp Bot 65, 1713-1735.

Bremer, G., 1961. Problems in breeding and cytology of sugar cane. Euphytica 10(1), 59-78.

Casu, R., Grof, C.L., Rae, A., McIntyre, C.L., Dimmock, C., Manners, J., 2003. Identification of a novel sugar transporter homologue strongly expressed in maturing stem vascular tissues of sugarcane by expressed sequence tag and microarray analysis. Plant Mol Biol 52(2), 371-386.

Casu, R.E., Rae, A.L., Nielsen, J.M., Perroux, J.M., Bonnett, G.D., Manners, J.M., 2015. Tissue-specific transcriptome analysis within the maturing sugarcane stalk reveals spatial regulation in the expression of cellulose synthase and sucrose transporter gene families. Plant Mol Biol 89(6), 607-628.

Chen, L.-Q., 2014. SWEET sugar transporters for phloem transport and pathogen nutrition. New Phytol 201(4), 1150-1155.

Chen, L.-Q., Qu, X.-Q., Hou, B.-H., Sosso, D., Osorio, S., Fernie, A.R., Frommer, W.B., 2012. Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. Science 335(6065), 207-211.

Cho, J.-I., Burla, B., Lee, D.-W., Ryoo, N., Hong, S.-K., Kim, H.-B., Eom, J.-S., Choi, S.-B., Cho, M.-H., Bhoo, S.H., Hahn, T.-R., Ekkehard Neuhaus, H., Martinoia, E., Jeon, J.-S., 2010. Expression analysis and functional characterization of the monosaccharide transporters, OsTMTs, involving vacuolar sugar transport in rice (*Oryza sativa*). New Phytol 186(3), 657-668.

Dhungana, S.R., Braun, D.M., 2021. Sugar transporters in grasses: Function and modulation in source and storage tissues. J Plant Phys 266, 153541.

Dobson, L., Reményi, I., Tusnády, G.E., 2015. CCTOP: a Consensus Constrained TOPology prediction web server. Nucleic Acids Res 43(W1), W408-W412.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32(5), 1792-1797.

Eom, J.-S., Chen, L.-Q., Sosso, D., Julius, B.T., Lin, I.W., Qu, X.-Q., Braun, D.M., Frommer, W.B., 2015. SWEETs, transporters for intracellular and intercellular sugar translocation. Current Opinion in Plant Biology 25(0), 53-62.

Evert, R.F., Eschrich, W., Heyser, W., 1978. Leaf structure in relation to solute transport and phloem loading in *Zea mays* L. Planta 138, 279-294.

Evert, R.F., Russin, W.A., Botha, C.E.J., 1996. Distribution and frequency of plasmodesmata in relation to photoassimilate pathways and phloem loading in the barley leaf. Planta 198(4), 572-579.

Figueira, T.R.e.S., Okura, V., Rodrigues da Silva, F., Jose da Silva, M., Kudrna, D., Ammiraju, J.S., Talag, J., Wing, R., Arruda, P., 2012. A BAC library of the SP80-3280 sugarcane variety (*Saccharum* sp.) and its inferred microsynteny with the sorghum genome. BMC Research Notes 5(1), 185.

Garsmeur, O., Droc, G., Antonise, R., Grimwood, J., Potier, B., Aitken, K., Jenkins, J., Martin, G., Charron, C., Hervouet, C., Costet, L., Yahiaoui, N., Healey, A., Sims, D., Cherukuri, Y., Sreedasyam, A., Kilian, A., Chan, A., Van Sluys, M.-A., Swaminathan, K., Town, C., Bergès, H., Simmons, B., Glaszmann, J.C., van der Vossen, E., Henry, R., Schmutz, J., D'Hont, A., 2018. A mosaic monoploid reference sequence for the highly complex genome of sugarcane. Nat Commun 9(1), 2638.

Glassop, D., Stiller, J., Bonnett, G.D., Grof, C.P., Rae, A.L., 2017. An analysis of the role of the ShSUT1 sucrose transporter in sugarcane using RNAi suppression. Funct Plant Biol 44(8), 795-808.

Grativol, C., Regulski, M., Bertalan, M., McCombie, W.R., Silva, F.R., Zerlotini Neto, A., Vicentini, R., Farinelli, L., Hemerly, A.S., Martienssen, R.A., Ferreira, P.C.G., 2014. Sugarcane genome sequencing by methylation filtration provides tools for genomic research in the genus *Saccharum*. Plant J 79(1), 162-172.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology 59(3), 307-321.

Hu, W., Hua, X., Zhang, Q., Wang, J., Shen, Q., Zhang, X., Wang, K., Yu, Q., Lin, Y.-R., Ming, R., Zhang, J., 2018. New insights into the evolution and functional divergence of the SWEET family in *Saccharum* based on comparative genomics. BMC Plant Biol 18(1), 270.

Julius, B.T., Leach, K.A., Tran, T.M., Mertz, R.A., Braun, D.M., 2017. Sugar transporters in plants: New insights and discoveries. Plant Cell Physiol 58(9), 1442-1460.

Julius, B.T., McCubbin, T.J., Mertz, R.A., Baert, N., Knoblauch, J., Grant, D.G., Conner, K., Bihmidine, S., Chomet, P., Wagner, R., Woessner, J., Grote, K., Peevers, J., Slewinski, T.L., McCann, M.C., Carpita, N.C., Knoblauch, M., Braun, D.M., 2021. Maize *Brittle Stalk2-Like3*, encoding a COBRA protein, functions in cell wall formation and carbohydrate partitioning. Plant Cell 33(10), 3348-3366.

Jung, B., Ludewig, F., Schulz, A., Meißner, G., Wöstefeld, N., Flügge, U.-I., Pommerrenig, B., Wirsching, P., Sauer, N., Koch, W., Sommer, F., Mühlhaus, T., Schroda, M., Cuin, T.A., Graus, D., Marten, I., Hedrich, R., 2015. Identification of the transporter responsible for sucrose accumulation in sugar beet taproots. Nat Plants 1(1), 14001.

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35(6), 1547-1549.

Lalonde, S., Wipf, D., Frommer, W.B., 2004. Transport mechanisms for organic forms of carbon and nitrogen between source and sink. Annu Rev Plant Biol 55(1), 341-372.

Le, S.Q., Gascuel, O., 2008. An improved general amino acid replacement matrix. Molecular Biology and Evolution 25(7), 1307-1320.

Leach, K.A., Tran, T.M., Slewinski, T.L., Meeley, R.B., Braun, D.M., 2017. *Sucrose transporter2* contributes to maize growth, development, and crop yield. J Int Plant Biol 59, 390-408.

Letunic, I., Bork, P., 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 49(W1), W293-W296.

Lingle, S.E., 1987. Aspects of sucrose transport in stem parenchyma of sweet sorghum. Plant Physiol 83, S-142.

Lingle, S.E., 1989. Evidence for the uptake of sucrose intact into sugarcane internodes. Plant Physiol 90(1), 6-8.

Milne, R.J., Perroux, J.M., Rae, A.L., Reinders, A., Ward, J.M., Offler, C.E., Patrick, J.W., Grof, C.P., 2017. Sucrose transporter localization and function in phloem loading and unloading. Plant Physiol 173, 1330-1341.

Patrick, J.W., Botha, F.C., Birch, R.G., 2013. Metabolic engineering of sugars and simple sugar derivatives in plants. Plant Biotech J 11(2), 142-156.

Rae, A.L., Perroux, J.M., Grof, C.P.L., 2005. Sucrose partitioning between vascular bundles and storage parenchyma in the sugarcane stem: a potential role for the ShSUT1 sucrose transporter. Planta 220(6), 817-825.

Reinders, A., Sivitz, A., Hsi, A., Grof, C., Perroux, J., Ward, J.M., 2006. Sugarcane ShSUT1: analysis of sucrose transport activity and inhibition by sucralose. Plant Cell Environ 29(10), 1871-1880.

Robinson-Beers, K., Evert, R.F., 1991. Ultrastructure of and plasmodesmatal frequency in mature leaves of sugarcane. Planta 184(3), 291-306.

Rotsch, D., Brossard, T., Bihmidine, S., Ying, W., Gaddam, V., Harmata, M., Robertson, J.D., Swyers, M., Jurisson, S.S., Braun, D.M., 2015. Radiosynthesis of 6'-deoxy-6'[^{18}F]fluorosucrose via automated synthesis and its utility to study *in vivo* sucrose transport in maize (*Zea mays*) leaves. PLoS One 10(5), e0128989.

Sauer, N., 2007. Molecular physiology of higher plant sucrose transporters. FEBS Lett. 581(12), 2309-2317.

Sharma, A., Song, J., Lin, Q., Singh, R., Ramos, N., Wang, K., Zhang, J., Ming, R., Yu, Q., 2018. Comparative analysis of homologous sequences of *Saccharum officinarum* and *Saccharum spontaneum* reveals independent polyploidization events. Front Plant Sci 9, 10.3389/fpls.2018.01414.

Slewinski, T.L., Braun, D.M., 2010. Current perspectives on the regulation of whole-plant carbohydrate partitioning. Plant Sci 178, 341-349.

Slewinski, T.L., Meeley, R., Braun, D.M., 2009. *Sucrose transporter1* functions in phloem loading in maize leaves. J Exp Bot 60, 881-892.

Souza, G.M., Berges, H., Bocs, S., Casu, R., D'Hont, A., Ferreira, J.E., Henry, R., Ming, R., Potier, B., Van Sluys, M.-A., Vincentz, M., Paterson, A.H., 2011. The sugarcane genome challenge: Strategies for sequencing a highly complex genome. Tropical Plant Biology 4(3), 145-156.

Thompson, R., Dale, J., 1981. Export of ^{14}C- and ^{11}C-labelled assimilate from wheat and maize leaves: effects of parachloromercurobenzylsulphonic acid and fusicoccin and of potassium deficiency. Can J Bot 59, 2439-2444.

Tran, T.M., Hampton, C.S., Brossard, T.W., Harmata, M., Robertson, J.D., Jurisson, S.S., Braun, D.M., 2017. In vivo transport of three radioactive [^{18}F]-fluorinated deoxysucrose analogs by the maize sucrose transporter ZmSUT1. Plant Physiol Biochem 115, 1-11.

Vilela, M.d.M., Del Bem, L.E., Van Sluys, M.-A., de Setta, N., Kitajima, J.P., Cruz, G.M.Q., Sforça, D.A., de Souza, A.P., Ferreira, P.C.G., Grativol, C., Cardoso-Silva, C.B., Vicentini, R., Vincentz, M., 2017. Analysis of three sugarcane homo/homeologous regions suggests independent polyploidization events of *Saccharum officinarum* and *Saccharum spontaneum*. Genome Biology and Evolution 9(2), 266-278.

Wang, J., Li, Y., Ching, M.W., Beuchat, G., Chen, L.Q., 2021. Identification and analysis of stem-specific promoters from sugarcane and energy cane for oil accumulation in their stems. Global Change Biology. Bioenergy 13(9), 1515-1527.

Wang, X., Tang, H., Paterson, A.H., 2011. Seventy million years of concerted evolution of a homoeologous chromosome pair, in parallel, in major Poaceae lineages. The Plant cell 23(1), 27-37.

Zhang, G., Ge, C., Xu, P., Wang, S., Cheng, S., Han, Y., Wang, Y., Zhuang, Y., Hou, X., Yu, T., 2021a. The reference genome of *Miscanthus floridulus* illuminates the evolution of Saccharinae. Nat Plants 7(5), 608-618.

Zhang, J., Zhang, X., Tang, H., Zhang, Q., Hua, X., Ma, X., Zhu, F., Jones, T., Zhu, X., Bowers, J., Wai, C.M., Zheng, C., Shi, Y., Chen, S., Xu, X., Yue, J., Nelson, D.R., Huang, L., Li, Z., Xu, H., Zhou, D., Wang, Y., Hu, W., Lin, J., Deng, Y., Pandey, N., Mancini, M., Zerpa, D., Nguyen, J.K., Wang, L., Yu, L., Xin, Y., Ge, L., Arro, J., Han, J.O., Chakrabarty, S., Pushko, M., Zhang, W., Ma, Y., Ma, P., Lv, M., Chen, F., Zheng, G., Xu, J., Yang, Z., Deng, F., Chen, X., Liao, Z., Zhang, X., Lin, Z., Lin, H., Yan, H., Kuang, Z., Zhong, W., Liang, P., Wang, G., Yuan, Y., Shi, J., Hou, J., Lin, J., Jin, J., Cao, P., Shen, Q., Jiang, Q., Zhou, P., Ma, Y., Zhang, X., Xu, R., Liu, J., Zhou, Y., Jia, H., Ma, Q., Qi, R., Zhang, Z., Fang, J., Fang, H., Song, J., Wang, M., Dong, G., Wang, G., Chen, Z., Ma, T., Liu, H., Dhungana, S.R., Huss, S.E., Yang, X., Sharma, A., Trujillo, J.H., Martinez, M.C., Hudson, M., Riascos, J.J., Schuler, M., Chen, L.-Q., Braun, D.M., Li, L., Yu, Q., Wang, J., Wang, K., Schatz, M.C., Heckerman, D., Van Sluys, M.-A., Souza, G.M., Moore, P.H., Sankoff, D., VanBuren, R., Paterson, A.H., Nagai, C., Ming, R., 2018. Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. Nat Genet 50, 1565-1573.

Zhang, Q., Hu, W., Zhu, F., Wang, L., Yu, Q., Ming, R., Zhang, J., 2016. Structure, phylogeny, allelic haplotypes and expression of sucrose transporter gene families in *Saccharum*. BMC Genom 17(1), 88.

Zhang, Q., Hua, X., Liu, H., Yuan, Y., Shi, Y., Wang, Z., Zhang, M., Ming, R., Zhang, J., 2021b. Evolutionary expansion and functional divergence of sugar transporters in *Saccharum* (*S. spontaneum* and *S. officinarum*). Plant J 105(4), 884-906.

TABLES

Table 3.1

***SUT* gene family in *S. spontaneum* and *S. officinarum*:** The table shows the *SUT* genes identified in the analyzed sugarcane species. A total of 19 alleles of six different *SUT* genes were identified in *S. spontaneum,* whereas 34 alleles were identified in *S. officinarum*. The *Sorghum bicolor* reference gene used is listed along with the corresponding *SUTs* in these two sugarcane species. The sugarcane *SUT2* and *SUT4* genes are named based on their homology to previously published sugarcane *SUT* sequences (Zhang et al., 2016), whereas the sorghum sequences are named based on their homology to rice *SUT* sequences. The start and end coordinates of genes are predicted (see methods).

Sorghum reference	Sugarcane Gene Name	Chr	Start	End	Gene ID
Sb01g045720-SbSUT1	SsSUT1-B	Chr1B	9671786	9677315	Sspon.001B0004212
	SsSUT1-D	Chr1D	9521443	9527484	Sspon.001D0004030
	SoSUT1-A	Chr01A	131430738	131436811	Soffic.01G0008450-1P
	SoSUT1-C1	Chr01C	134259518	134262994	Soffic.01G0008450-2C
	SoSUT1-C2	Chr01C	108512359	108518883	Soffic.01G0008450-2P
	SoSUT1-H1	Chr01H	84421484	84424983	Soffic.01G0008450-3H
	SoSUT1-H2	Chr01H	7379306	7391898	Soffic.01G0008450-3P
	SoSUT1-D	Chr01D	101354445	101360197	Soffic.01G0035780-2D
Sb04g038030-SbSUT4	SsSUT2-B	Chr4B	701130	705549	Sspon.004B0000310
	SsSUT2-C	Chr4C	1045466	1049591	Sspon.004C0000360
	SsSUT2-D1	Chr4D	823766	828264	Sspon.004D0000480
	SsSUT2-D2	Chr4D	893009	903924	Sspon.004D0000521
	SoSUT2-A	Chr04A	110941539	110946062	Soffic.04G0028100-1A
	SoSUT2-B	Chr04B	83469575	83474062	Soffic.04G0028100-1P

	SoSUT2-H	Chr04H	63564756	63569154	Soffic.04G0028100-2P
	SoSUT2-C	Chr04C	91073824	91078639	Soffic.04G0028100-3C
Sb01g022430-SbSUT3	SsSUT3-B	Chr1B	69508972	69514231	Sspon.001B0025961
	SsSUT3-C	Chr1C	66142497	66146366	Sspon.001C0021910
	SsSUT3-D	Chr1D	61813849	61817677	Sspon.001D0022020
	SsSUT3-A	Chr1A	65067966	65071810	Sspon.001A0021840
	SoSUT3-A	Chr01A	56962778	56966267	Soffic.01G0035780-1A
	SoSUT3-F1	Chr01F	41924929	41929708	Soffic.01G0035780-1T
	SoSUT3-F2	Chr01F	41974584	41976010	Soffic.01G0035780-2P
	SoSUT3-F3	Chr01F	45260229	45265006	Soffic.01G0035780-3F
	SoSUT3-G	Chr01G	66629288	66633986	Soffic.01G0035780-4G
Sb08g023310-SbSUT2	SsSUT4-A	Chr2A	119505644	119516555	Sspon.002A0043140
	SsSUT4-B	Chr2B	113717644	113723544	Sspon.002B0040235
	SsSUT4-hc	tig00047413	16521	23403	Sspon.ctg0462330
	SoSUT4-A	Chr08A	63347935	63356890	Soffic.08G0016150-1A
	SoSUT4-B1	Chr08B	60802156	60802784	Soffic.08G0016150-2B
	SoSUT4-B2	Chr08B	60694316	60699779	Soffic.08G0016150-2P
	SoSUT4-E	Chr08E	56004582	56011231	Soffic.08G0016150-3E
	SoSUT4-F	Chr08F	56903462	56910633	Soffic.08G0016150-4F
	SoSUT4-B3	Chr08B	60796462	60801133	Soffic.08G0016150-4P
	SoSUT4-H	Chr08H	53067433	53079664	Soffic.08G0016150-5H
Sb04g023860-SbSUT5	SsSUT5-A	Chr4A	29176716	29179997	Sspon.004A0012390
	SsSUT5-B	Chr4B	24680752	24683549	Sspon.004B0011210
	SsSUT5-C	Chr4C	28227745	28230693	Sspon.004C0012570
	SsSUT5-D	Chr4D	29175861	29178639	Sspon.004D0013960
	SoSUT5-B1	Chr04B	49245418	49248372	Soffic.04G0014730-1B
	SoSUT5-B2	Chr04B	49228483	49231186	Soffic.04G0014730-1P
	SoSUT5-D	Chr04D	58907273	58909974	Soffic.04G0014730-2D
	SoSUT5-C1	Chr04C	58967403	58970123	Soffic.04G0014730-2P
	SoSUT5-E	Chr04E	48337287	48340241	Soffic.04G0014730-3E
	SoSUT5-C2	Chr04C	59211679	59214619	Soffic.04G0014730-3P
	SoSUT5-G	Chr04G	49393211	49395946	Soffic.04G0014730-4G

	SoSUT5-H	Chr04H	25180148	25182869	Soffic.04G0014730-5H
Sb07g028120-SbSUT6	SsSUT6-A	Chr6A	4396234	4401330	Sspon.006A0002160
	SsSUT6-C	Chr6C	2657744	2662838	Sspon.006C0001250
	SoSUT6-B	Chr07B	76935062	76946300	Soffic.07G0017780-1B
	SoSUT6-D	Chr07D	62904804	62916466	Soffic.07G0017780-1P
	SoSUT6-F1	Chr07F	55746384	55754565	Soffic.07G0017780-2F
	SoSUT6-F2	Chr07F	54765249	54770625	Soffic.07G0017780-2P

Table 3.2

***TST* gene family in *S. spontaneum* and *S. officinarum*:** The table shows the *TST* genes identified in the analyzed sugarcane species. A total of 16 alleles of five different *TST* genes were identified in *S. spontaneum,* whereas 25 alleles of five different *TST* genes were identified in *S. officinarum*. The *Sorghum bicolor* reference gene used is listed along with the corresponding *TSTs* in these two sugarcane species. The start and end coordinates of genes are predicted (see methods).

Sorghum reference	Sugarcane Gene Name	Chr	Start	End	Gene ID
Sb01g030430-SbTST1	SsTST1-A	Chr1A	37679910	37684295	Sspon.001A0016560
	SsTST1-B	Chr1B	44537392	44541327	Sspon.001B0020210
	SsTST1-C	Chr1C	38138484	38143548	Sspon.001C0016151
	SsTST1-D	Chr1D	39904649	39909426	Sspon.001D0017272
	SoTST1-A	Chr01A	89112319	89117425	Soffic.01G0048790-1A
	SoTST1-F	Chr01F	92330118	92335206	Soffic.01G0048790-1P
	SoTST1-B	Chr01B	84770275	84773859	Soffic.01G0048790-2B
	SoTST1-C	Chr01C	66052679	66058094	Soffic.01G0048790-3C
	SoTST1-E	Chr01E	59884408	59889823	Soffic.01G0048790-4E
	SoTST1-G	Chr01G	88372338	88377444	Soffic.01G0048790-6G
	SoTST1-H	Chr01H	47762233	47767519	Soffic.01G0048790-7H
Sb04G008150-SbTST2	SsTST2a-A	Chr4A	55286300	55291355	Sspon.004A0018790
	SsTST2a-C	Chr4C	60229974	60234670	Sspon.004C0018690
	SoTST2a-A1	Chr04A	28216156	28221475	Soffic.04G0008740-1A
	SoTST2a-A2	Chr04A	27759002	27764315	Soffic.04G0008740-1P
	SoTST2a-B1	Chr04B	24188213	24193247	Soffic.04G0008740-2B
	SoTST2a-B2	Chr04B	24173384	24178415	Soffic.04G0008740-1T
	SoTST2a-C1	Chr04C	23342398	23347469	Soffic.04G0008740-3C
	SoTST2a-C2	Chr05C	20566698	20572018	Soffic.04G0008750-1C
	SoTST2a-E	Chr04E	23265727	23271070	Soffic.04G0008750-1P
	SoTST2a-A3	Chr04A	28162490	28167800	Soffic.04G0008750-1T

	SsTST2b-A	Chr3A	77414428	77431624	Sspon.003A0031750
	SsTST2b-hc	tig00072049	61974	65659	Sspon.ctg0727530
	SoTST2b-C	Chr10C	67564396	67569807	Soffic.10G0028380-1C
	SoTST2b-E	Chr10E	58570359	58575738	Soffic.10G0028380-2E
Sb10g031000-SbTST3	SsTST3a-A	Chr8A	1816887	1819757	Sspon.008A0000770
	SsTST3a-B	Chr8B	473538	476345	Sspon.008B0000222
	SsTST3a-C1	Chr8C	338203	341060	Sspon.008C0000240
	SsTST3a-C2	Chr8C	431585	435668	Sspon.008C0000320
	SsTST3a-D	Chr8D	291465	294330	Sspon.008D0000161
	SoTST3a-A	Chr10A	72513809	72516633	Soffic.10G0024430-1A
	SoTST3a-C	Chr05C	400852	403676	Soffic.10G0024430-1P
	SoTST3a-E	Chr10E	58102467	58105294	Soffic.10G0024430-2E
	SoTST3a-hc1	tig00117220	33754	36484	Soffic0038960
	SoTST3a-hc2	tig00167793	10875	13696	Soffic0086390
	SsTST3b-B	Chr4B	204673	206760	Sspon.004B0000020
	SsTST3b-D1	Chr4D	543230	545269	Sspon.004D0000240
	SsTST3b-D2	Chr4D	259858	261936	Sspon.004D0000100
	SoTST3b-C1	Chr04C	91727133	91728896	Soffic.04G0028480-3C
	SoTST3b-C2	Chr04C	55451238	55453322	Soffic.04G0028480-3P
	SoTST3b-F	Chr03F	83924080	83925514	Soffic.04G0028480-4F

Table S3. 1

Table with Accession IDs for sequences used for phylogenetic analyses of SUTs and TSTs along with truncated SUT-like and TST-like sequences that were removed from the analyses.

Species	Gene Name	Gene ID/ Accession ID
Arabidopsis thaliana	AtSUC1	AT1G71880.1
	AtSUC2	AT1G22710.1
	AtSUC3	AT2G02860.1
	AtSUC4	AT1G09960.1
	AtSUC5	AT1G71890.1
	AtSUC6	AT5G43610.1
	AtSUC7	AT1G66570.1
	AtSUC8	AT2G14670.1
	AtSUC9	AT5G06170.1
	AtTMT1	Q96290
	AtTMT2	Q8LPQ8
	AtTMT3	Q9SD00
Beta vulgaris	BvTST1	XP_010686712.1
	BvTST2.1	XP_010678631.1
	BvTST2.2	XP_010690557.1
	BvTST3	XP_010680636.1
Brachypodium distachyon	BdSUT1	BRADI1G73170.1
	BdSUT2	BRADI3G56740.1
	BdSUT3	BRADI3G25477.1
	BdSUT4	BRADI4G00320.1
	BdSUT5	BRADI3G46790.1
	BdTST3	BRADI1G29600.1
	BdTST2	BRADI3G08690.1
	BdTST1	BRADI3G32210.2
Galdieria sulphuraria	Gasu_08920	Gasu_08920
	Gasu_18190	Gasu_18190
	Gasu_29860	Gasu_29860
	Gasu_34550	Gasu_34550
	Gasu_56570	Gasu_56570
	Gasu_54152	ABM54152.1
	Gasu_005704896	XP_005704896.1
Glycine max	GmSUT1	CAD91334.1
	GmSUC3	XP_003530692.2
	GmSUC8	XP_003548077.1

	GmSUC8A	XP_006599447.1
	GmSUC8B	XP_006599448.1
	GmTST1a	GLYMA_06G015000
	GmTST1b	GLYMA_04G015000
	GmTST2a	GLYMA_04G000300
	GmTST2b	GLYMA_06G000200
	GmTST2c	GLYMA_02G311700
	GmTST2d	GLYMA_14G000900
	GmTST3a	GLYMA_16G112500
	GmTST3b	GLYMA_11G087700
	GmTST3c	GLYMA_01G157300
Medicago truncatula	MtSUT1-1	AFM28284.1
	MtSUT1-2	AFM28285.1
	MtSUT1-3	AFM28286.1
	MtSUT2	AFM28287.1
	MtSUT4-1	AFM28288.1
	MtSUT4-2	AFM28289.1
	MtTST1	MTR_3g116060
	MTTST2	MTR_3g118530
	MtTST3a	MTR_5g024740
	MtTST3b	MTR_5g044910
	MtTST3c	MTR_8g073100
Oryza sativa	OsSUT1	LOC_Os03g07480.2
	OsSUT2	LOC_Os12g44380.1
	OsSUT3	LOC_Os10g26470.1
	OsSUT4	LOC_Os02g58080.1
	OsSUT5	LOC_Os02g36700.1
	OsTMT1	LOC_Os10g39440.1
	OsTMT2	LOC_Os02g13560.4
	OsTMT3	LOC_Os03g03680.1
	OsTMT4	LOC_Os11g40540.1
	OsTMT5	LOC_Os02g58530.1
	OsTMT6	LOC_Os11g28610.1
Sorghum bicolor	SbSUT1	Sobic.001G488700.1
	SbSUT2	Sobic.008G193300.1
	SbSUT3	Sobic.001G254000.1
	SbSUT4	Sobic.004G353600.1
	SbSUT5	Sobic.004G190500.1
	SbSUT6	Sobic.007G214500.1
	SbTST1	Sobic.001G312900.1
	SbTST2	Sobic.004G099300.2
	SbTST3	Sobic.010G276100.1
Setaria italica	SiSUT1	Seita.9G523900.1

	SiSUT2	Seita.1G374900.1
	SiSUT3	Seita.9G258000.1
	SiSUT4	Seita.3G408800.1
	SiSUT5	Seita.1G207000.1
	SiTST1a	Si026035m
	SiTST1b	Si034411m
	SiTST2	Si016433m
	SiTST3	Si008056m
Saccharum officinarum	SoSUT1-A	Soffic.01G0008450-1P
	SoSUT1-C1	Soffic.01G0008450-2C
	SoSUT1-C2	Soffic.01G0008450-2P
	SoSUT1-H1	Soffic.01G0008450-3H
	SoSUT1-H2	Soffic.01G0008450-3P
	SoSUT1-D	Soffic.01G0035780-2D
	SoSUT2-A	Soffic.04G0028100-1A
	SoSUT2-B	Soffic.04G0028100-1P
	SoSUT2-H	Soffic.04G0028100-2P
	SoSUT2-C	Soffic.04G0028100-3C
	SoSUT3-A	Soffic.01G0035780-1A
	SoSUT3-F1	Soffic.01G0035780-1T
	SoSUT3-F2	Soffic.01G0035780-2P
	SoSUT3-F3	Soffic.01G0035780-3F
	SoSUT3-G	Soffic.01G0035780-4G
	SoSUT4-A	Soffic.08G0016150-1A
	SoSUT4-B1	Soffic.08G0016150-2B
	SoSUT4-B2	Soffic.08G0016150-2P
	SoSUT4-E	Soffic.08G0016150-3E
	SoSUT4-F	Soffic.08G0016150-4F
	SoSUT4-B3	Soffic.08G0016150-4P
	SoSUT4-H	Soffic.08G0016150-5H
	SoSUT5-B1	Soffic.04G0014730-1B
	SoSUT5-B2	Soffic.04G0014730-1P
	SoSUT5-D	Soffic.04G0014730-2D
	SoSUT5-C1	Soffic.04G0014730-2P
	SoSUT5-E	Soffic.04G0014730-3E
	SoSUT5-C2	Soffic.04G0014730-3P
	SoSUT5-G	Soffic.04G0014730-4G
	SoSUT5-H	Soffic.04G0014730-5H
	SoSUT6-B	Soffic.07G0017780-1B
	SoSUT6-D	Soffic.07G0017780-1P
	SoSUT6-F1	Soffic.07G0017780-2F
	SoSUT6-F2	Soffic.07G0017780-2P
	SoTST1-A	Soffic.01G0048790-1A

	SoTST1-F	Soffic.01G0048790-1P
	SoTST1-B	Soffic.01G0048790-2B
	SoTST1-C	Soffic.01G0048790-3C
	SoTST1-E	Soffic.01G0048790-4E
	SoTST1-G	Soffic.01G0048790-6G
	SoTST1-H	Soffic.01G0048790-7H
	SoTST2a-A1	Soffic.04G0008740-1A
	SoTST2a-A2	Soffic.04G0008740-1P
	SoTST2a-B2	Soffic.04G0008740-1T
	SoTST2a-B1	Soffic.04G0008740-2B
	SoTST2a-C1	Soffic.04G0008740-3C
	SoTST2a-C2	Soffic.04G0008750-1C
	SoTST2a-E	Soffic.04G0008750-1P
	SoTST2a-A3	Soffic.04G0008750-1T
	SoTST2b-C	Soffic.10G0028380-1C
	SoTST2b-E	Soffic.10G0028380-2E
	SoTST3a-A	Soffic.10G0024430-1A
	SoTST3a-C	Soffic.10G0024430-1P
	SoTST3a-E	Soffic.10G0024430-2E
	SoTST3a-hc1	Soffic0038960
	SoTST3a-hc2	Soffic0086390
	SoTST3b-C1	Soffic.04G0028480-3C
	SoTST3b-C2	Soffic.04G0028480-3P
	SoTST3b-F	Soffic.04G0028480-4F
Saccharum spontaneum	SsSUT1-B	Sspon.001B0004212
	SsSUT1-D	Sspon.001D0004030
	SsSUT2-B	Sspon.004B0000310
	SsSUT2-C	Sspon.004C0000360
	SsSUT2-D1	Sspon.004D0000480
	SsSUT2-D2	Sspon.004D0000521
	SsSUT3-B	Sspon.001B0025961
	SsSUT3-C	Sspon.001C0021910
	SsSUT3-D	Sspon.001D0022020
	SsSUT3-A	Sspon.001A0021840
	SsSUT4-A	Sspon.002A0043140
	SsSUT4-B	Sspon.002B0040235
	SsSUT4-hc	Sspon.ctg0462330
	SsSUT5-A	Sspon.004A0012390
	SsSUT5-B	Sspon.004B0011210
	SsSUT5-C	Sspon.004C0012570
	SsSUT5-D	Sspon.004D0013960
	SsSUT6-A	Sspon.006A0002160
	SsSUT6-C	Sspon.006C0001250

	SsTST1-A	Sspon.001A0016560
	SsTST1-B	Sspon.001B0020210
	SsTST1-C	Sspon.001C0016151
	SsTST1-D	Sspon.001D0017272
	SsTST2a-A	Sspon.004A0018790
	SsTST2a-C	Sspon.004C0018690
	SsTST2b-A	Sspon.003A0031750
	SsTST2b-hc	Sspon.ctg0727530
	SsTST3a-A	Sspon.008A0000770
	SsTST3a-B	Sspon.008B0000222
	SsTST3a-C1	Sspon.008C0000240
	SsTST3a-C2	Sspon.008C0000320
	SsTST3a-D	Sspon.008D0000161
	SsTST3b-B	Sspon.004B0000020
	SsTST3b-D1	Sspon.004D0000240
	SsTST3b-D2	Sspon.004D0000100
***Saccharum* hybrid cultivar Q117**	ShSUT1	AAV41028
	ShPST2a	AA037640.1
Saccharum hybrid cultivar ROC22	ShSUT4	ACV95498
Solanum tubersosum	StSUT1	XP_006351170.1
	StSUT2	NP_001275438.1
	StSUT3B	XP_006363242.1
	StSUT4	NP_001275070.1
	StTST1	PGSC0003DMG400009994
	StTST2	PGSC0003DMG400035539
	StTST3a	PGSC0003DMG400013648
	StTST3b	PGSC0003DMG400018666
Vitis vinifera	VvSUT1	ADP37121.1
	VvSUT2	ADP37124.1
	VvSUT11	NP_001268066.1
	VvSUT12	ADP37122.1
	VvSUT27	ADP37123.1
	VvTST1	VIT_18s0122g00850.t01
	VvTST2	VIT_03s0038g03940.t01
	VvTST3	VIT_07s0031g02270.t01
Zea mays	ZmSUT1	GRMZM2G034302_T01
	ZmSUT2	GRMZM2G307561_T03
	ZmSUT3	GRMZM2G083248_T01
	ZmSUT4	GRMZM2G145107_T01
	ZmSUT5	GRMZM2G081589_T01
	ZmSUT6	GRMZM2G106741_T01
	ZmSUT7	GRMZM2G087901_T01
	ZmTST1	Zm00001d029762_T004

	ZmTST2	Zm00001d016274_T011
	ZmTST3	Zm00001d014872_T001
Saccharum officinarum	SUT-like sequences	Soffic.01G0008450-1A
	SUT-like sequences	Soffic.01G0035780-1P
	SUT-like sequences	Soffic.01G0035780-3P
	SUT-like sequences	Soffic.04G0028100-2B
	SUT-like sequences	Soffic.08G0016150-1P
	SUT-like sequences	Soffic.08G0016160-1A
Saccharum spontaneum	SUT-like sequences	Sspon.001B0004220
	SUT-like sequences	Sspon.001B0004240
	SUT-like sequences	Sspon.002B0040270
	SUT-like sequences	Sspon.004D0000530
	SUT-like sequences	Sspon.004D0000910
Saccharum officinarum	TST-like sequences	Soffic.04G0008740-2P
	TST-like sequences	Soffic.04G0008740-3P
	TST-like sequences	Soffic.04G0008740-4G
	TST-like sequences	Soffic0054360

FIGURES

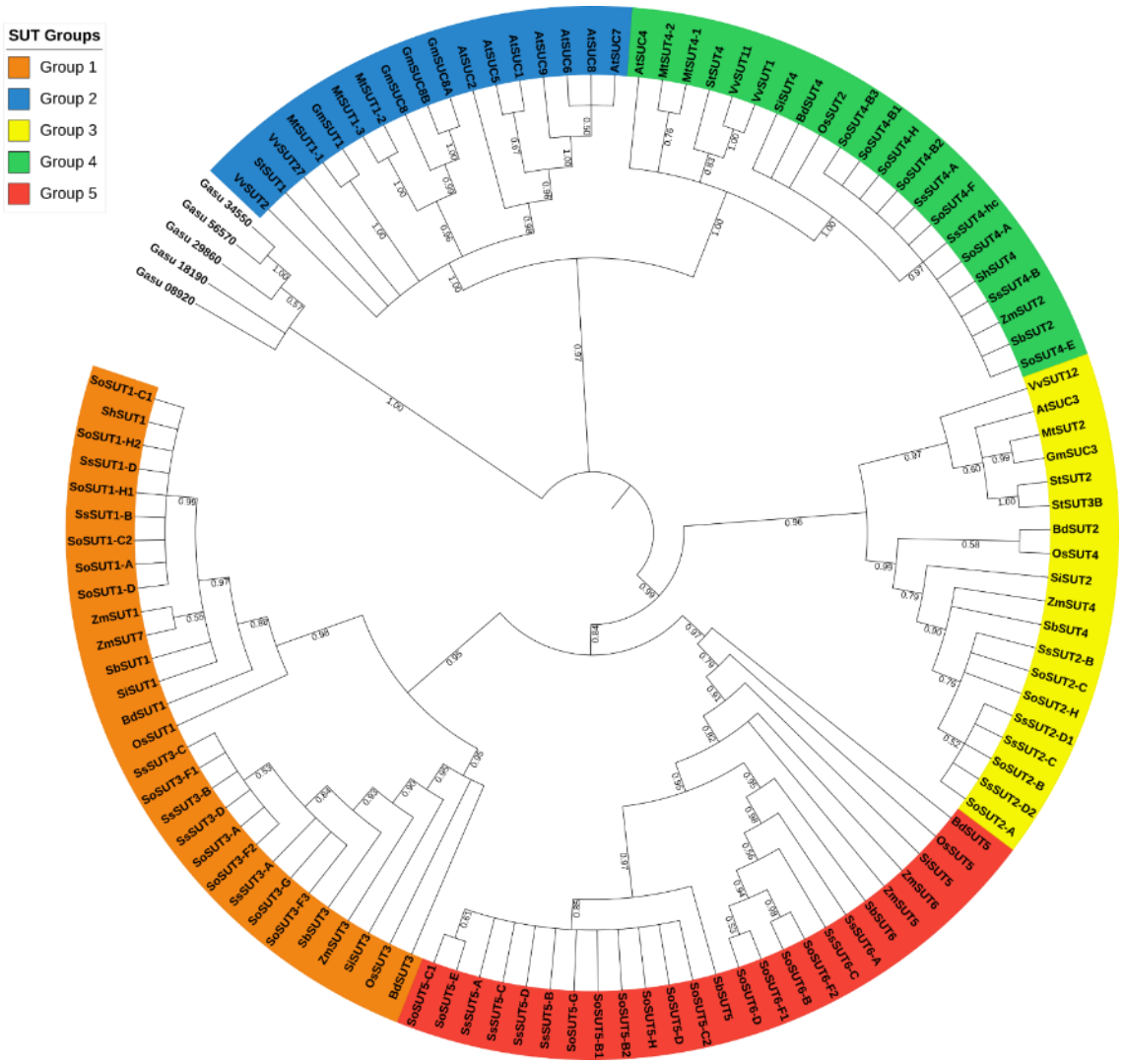


Figure 3.1

Phylogenetic tree of the *S. spontaneum* and *S. officinarum* Sucrose Transporter (SUT) protein family with other species. The phylogenetic tree was constructed using the maximum likelihood method with SUT sequences from various eudicot and monocot species: *Arabidopsis thaliana (At), Brachypodium distachyon (Bd), Galdieria sulphuraria (Gasu), Glycine max (Gm), Medicago truncatula (Mt), Oryza sativa (Os), Sorghum bicolor (Sb), Setaria italica (Si), Saccharum officinarum (So), Saccharum spontaneum*

(Ss), Saccharum hybrid cultivar Q117 (Sh), Solanum tuberosum (St), and *Zea mays (Zm).* Different SUT clades are highlighted: Group 1 (orange) and group 5 (red) are found in monocots only, group 3 (yellow) and group 4 (green) are found in both eudicots and monocots, and group 2 (blue) is found in eudicots only. The branch support values denote the reliability for each internal branch based on 1000 bootstraps. All branches with a bootstrap value below 50 are collapsed. The tree was rooted using SUT sequences from a red alga *(Galdieria sulphuraria)* as the outgroup.

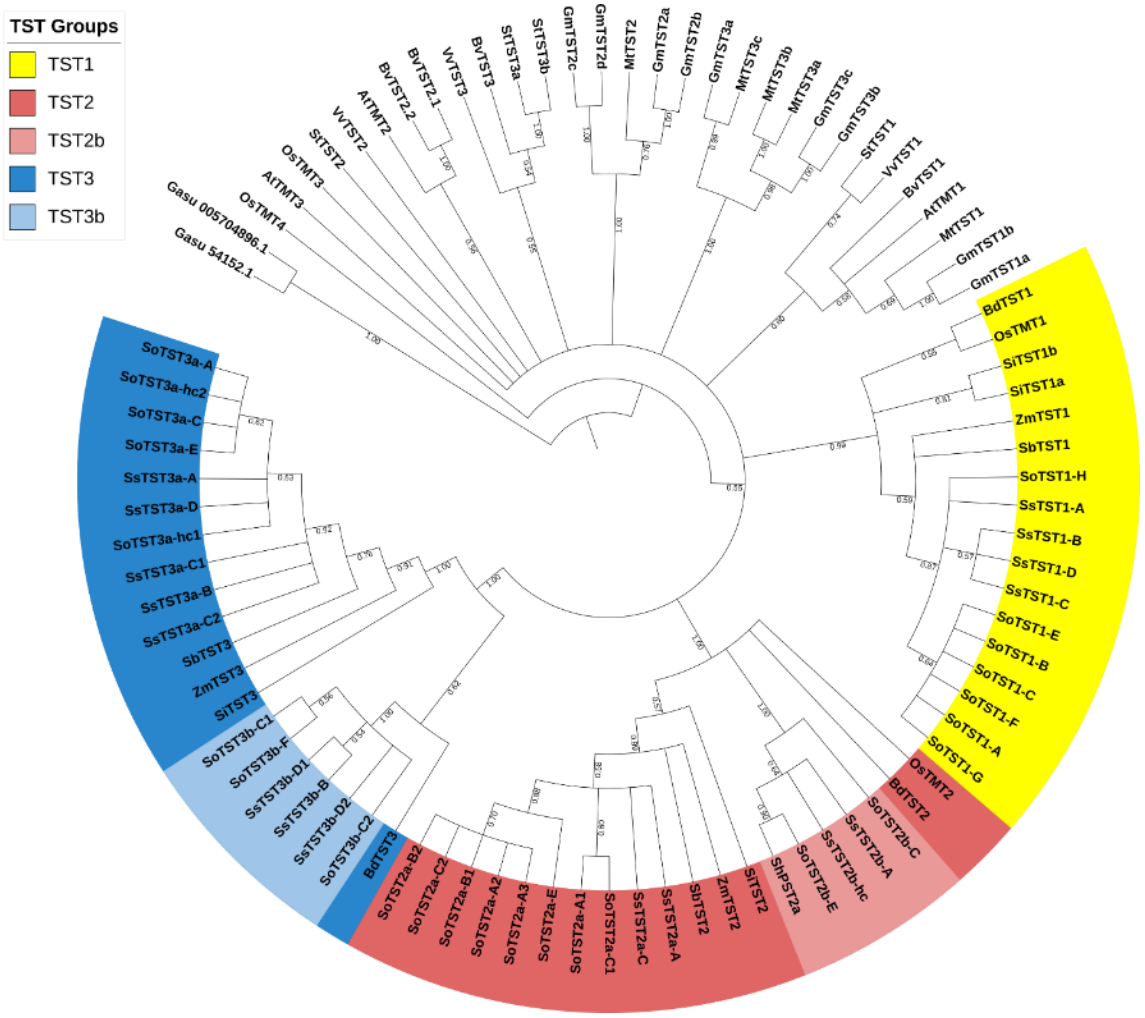


Figure 3.2

Phylogenetic tree of the *S. spontaneum* and *S. officinarum* Tonoplast Sugar Transporter (TST) protein family with other species. The phylogenetic tree was constructed using the maximum likelihood method with TST sequences from various eudicot and monocot species. The branch support values denote the reliability for each internal branch derived from 1000 bootstraps. All branches with bootstrap value below 50 are collapsed. Monocot-specific TST clades containing SsTST and SoTST families are highlighted: TST1 (yellow), TST2 (red), and TST3 (blue). Two distinct subfamily members within the TST2 clade containing TST2a and TST2b (highlighted in light red),

and TST3 clade containing TST3a and TST3b (highlighted in light blue), were identified for *S. spontaneum* and *S. officinarum*. The tree was rooted using orthologous sequences from a red alga *(Galdieria sulphuraria)* as the outgroup. Species abbreviations are defined in the legend to Figure 3.3.1. An additional species *Beta vulgaris (Bv)* is included.

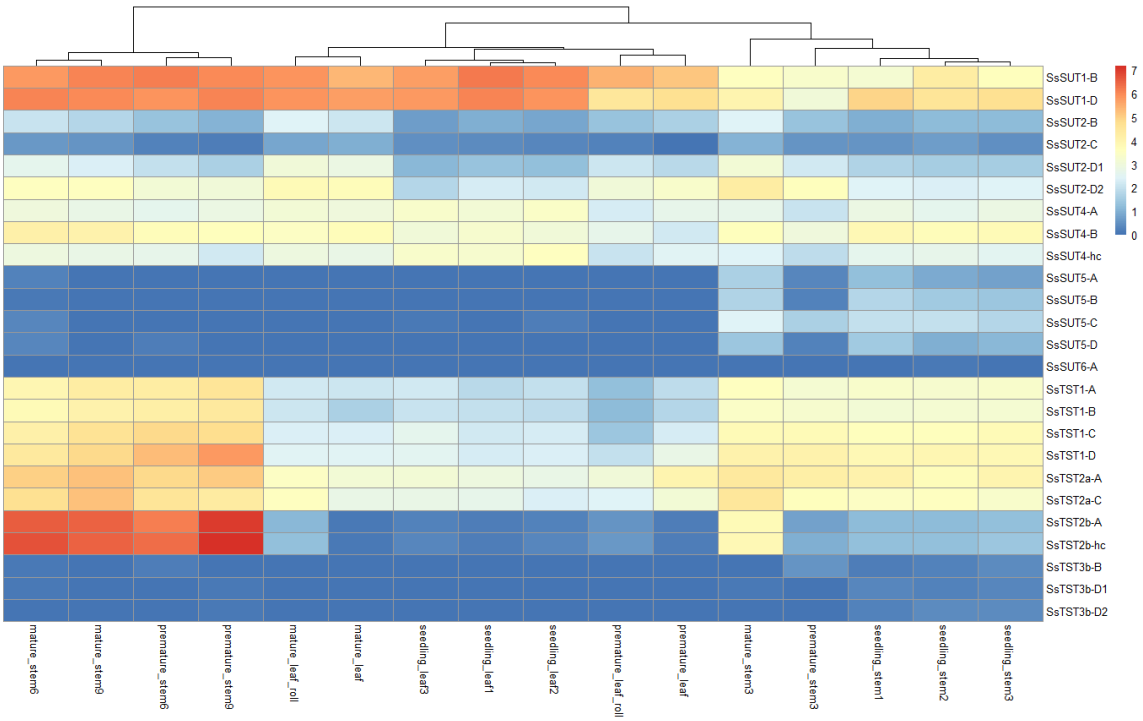


Figure 3.3

Expression levels of *SUTs* and *TSTs* in *S. spontaneum* from various tissues and developmental stages. A heatmap of log2 normalized FPKM values is shown. The expression level of each gene at each developmental stage is represented by the color intensity (blue represents lowest expression whereas red represents highest expression).

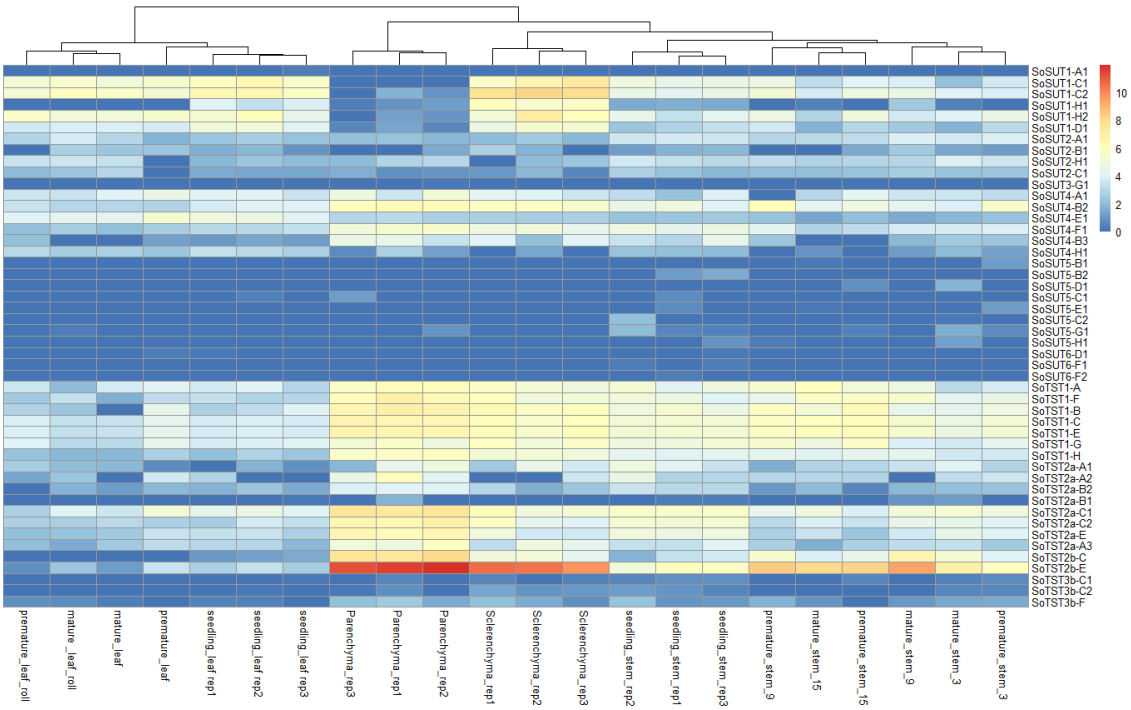


Figure 3.4

Expression levels of *SUTs* and *TSTs* in *S. officinarum* from various tissues and developmental stages. A heatmap of log2 normalized TPM values is shown. The expression level of each gene at each developmental stage is represented by the color intensity (blue represents lowest expression whereas red represents highest expression).

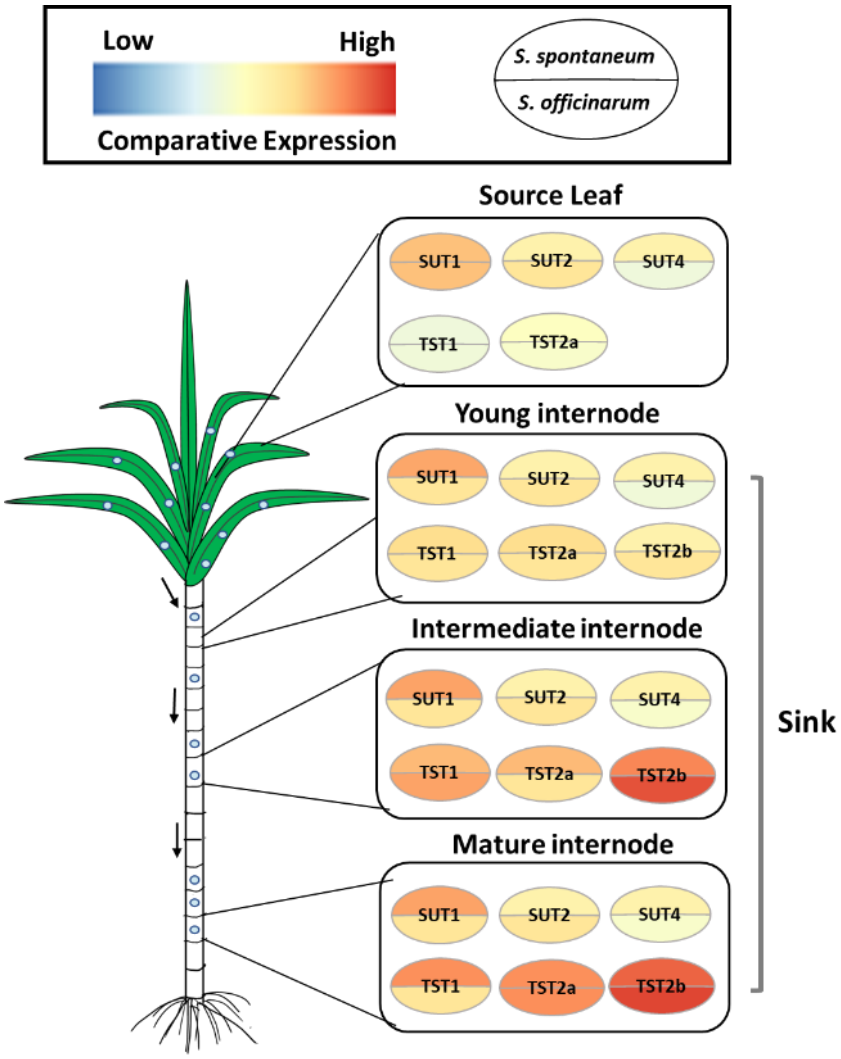


Figure 3.5

Schematic of *SUTs* and *TSTs* expression in sugarcane leaves and stem internodes based on gene expression profiles. The comparative expression patterns of *SUTs* and *TSTs* in source leaves, and in young (internode 3), intermediate (internode 6/9), and mature internodes (internode 9/15) for *S. spontaneum* and *S. officinarum* are shown. Relative expression of each gene is color coded. The path of sucrose transport is shown with sucrose as light blue dots. Sucrose synthesized in the source leaves is mobilized to

the internodes. As the internodes matured, increased gene expression of *TST2a* and *TST2b* were observed in both species whereas *TST1* expression was high in the intermediate internodes for both species.

Chapter 4 : A novel Dna-J-thioredoxin-like domain containing protein is required for carbohydrate partitioning in maize

ABSTRACT

Carbohydrate partitioning is the process by which sugars, primarily sucrose, synthesized in the photosynthetic source tissues (mature leaves) are mobilized to non-photosynthetic (sink) tissues, such as roots, seeds, and developing organs. As all heterotrophic life on earth relies on carbohydrates produced by plants as their primary source of energy, understanding how plants control the allocation of these compounds is crucial. Towards understanding the genetic control of this process, we identified two allelic recessive mutants from EMS (Ethyl methanesulfonate) mutagenized populations*, carbohydrate partitioning defective13 (cpd13)* and *cpd35,* which exhibit reduced plant growth, chlorotic leaves, and hyperaccumulation of soluble sugars and starch in mature leaves. Intriguingly, the mutant leaves exhibit a unique crossbanding pattern of alternating chlorotic and green regions. By using polymorphic DNA markers and whole genome sequencing-based approaches, we mapped the causative mutations to a gene encoding a protein containing DnaJ-like and thioredoxin-like domains. While these domains are not well studied in plants, some DnaJ-like proteins have protein folding activity and are known to be involved in protein quality control in other organisms. Expressing a translational fusion of CPD13 to a red fluorescent protein in tobacco leaves localized the protein to the endoplasmic reticulum, consistent with a potential chaperone function. We hypothesize that the carbohydrate hyperaccumulation in mutant leaves is due to the inability of the defective CPD13 protein to properly interact with or process target proteins.

INTRODUCTION

Carbohydrates are the major source of energy that fuels the growth and development of plants and provides energy for all heterotrophic life forms on earth. In maize and many other crop species, sucrose is the primary form of sugar transported long distance through the phloem (Dhungana and Braun, 2021; Julius et al., 2017; Zimmermann and Ziegler, 1975). The process of moving carbohydrates from the source tissues (mature leaves) to sink tissues (roots, developing leaves, and reproductive structures) is termed whole-plant carbohydrate partitioning (Braun and Slewinski, 2009; Slewinski and Braun, 2010a). Understanding this process is crucial to identify targets for crop improvement; however, only limited studies have explored the genetic architecture controlling carbohydrate partitioning. Towards understanding how it is regulated, several mutants were created via EMS mutagenesis. By studying these mutants, we aim to get a better understanding of the genetic architecture controlling carbohydrate partitioning. Among such mutants, we identified the recessive mutant *carbohydrate partitioning defective13 (cpd13)* along with a second recessive allele, *cpd35*.

To date, a number of mutants defective in carbohydrate partitioning have been reported. These mutants hyperaccumulate carbohydrates and display a set of phenotypes such as chlorotic leaves, and reduced plant growth and fertility. Such mutants can be divided into four different categories: 1) Sugar transporter mutants such as *sucrose transporter1 (sut1)*, *sucrose transporter2 (sut2),* and the triple *SWEET13a*, *b*, and *c* mutant, wild-type proteins of which function as a sucrose transporters (Baker et al., 2016; Bezrutczyk et al., 2018; Carpaneto et al., 2005; Leach et al., 2017; Slewinski et al., 2009); 2) Occlusion mutants such as *sucrose export defective1* (*sxd1*) , and C*arbohydrate*

partitioning defective1 (*Cpd1*), which exhibit ectopic callose deposition at various points along the symplastic sucrose transport pathway, (Julius et al., 2018; Russin et al., 1996); 3) Structural mutants such as *carbohydrate partitioning defective33 (cpd33)* which encodes a protein that localizes to plasmodesmata, or *carbohydrate partitioning defective28/47 (cpd28/47)* which encodes a COBRA protein involved in cell wall development (Julius et al., 2021; Tran et al., 2019); and 4) Mutants in which the mechanisms are not yet fully understood, such as, *psychedelic (psc)*, *tie-dyed1 (tdy1)*, and *tie-dyed2 (tdy2)*, in which the *tdy* loci are hypothesized to function to control transport through plasmodesmata of Companion Cells and Sieve Elements (Baker and Braun, 2008; Braun et al., 2006; Slewinski et al., 2012; Slewinski and Braun, 2010b). Thus, carbohydrate partitioning can be affected by a wide array of genes with diverse functions.

Here, we characterize two novel allelic mutants: *cpd13* and *cpd35*, which hyperaccumulate carbohydrates in their leaves. By utilizing Bulked Segregant Analysis, molecular markers, and whole genome sequencing techniques, we identified the gene responsible for the mutant phenotype and found that *Cpd13* encodes a protein containing DnaJ-like and thioredoxin-like domains. This protein localizes to the Endoplasmic Reticulum (ER), and functions in regulating carbohydrate partitioning. Proteins with Dna-J domains are known to function as molecular chaperons and to be involve in protein folding and maintenance across all species ranging from yeast and mammals to plants (Miernyk, 2001; Ohta and Takaiwa, 2014; Yamamoto et al., 2008). DnaJ proteins belong to the heat shock protein 40 family (hsp40) and are also known as members of the J-protein family. Dna-J proteins engage in various cellular processes, including protein folding of newly synthesized proteins, degradation of misfolded proteins, and

translocation of polypeptides across cellular membranes (Kampinga and Craig, 2010; Ohta and Takaiwa, 2014; Pulido and Leister, 2018). These proteins can localize in the mitochondria, cytosol or ER (Ohta and Takaiwa, 2014; Pulido and Leister, 2018). Dna-J proteins are known to deliver unfolded proteins to the Hsp70 chaperone via the J-domain, which is a conserved domain across many Dna-J proteins. With the exchange of ATP, the Hsp70 can then complete the protein folding process after delivery by Dna-J (Ohta and Takaiwa, 2014; Pulido and Leister, 2018).

Similarly, proteins with thioredoxin-like domains are known to change the activity of target proteins through changes in the redox state of thiol groups (S2 to SH2 or vice versa) (Cunnea et al., 2003). Many proteins from this family have been studied, and among them Protein Disulfide Isomerase (PDI) is a well-characterized member (Wilkinson and Gilbert, 2004). PDI is a key and abundant protein present in the ER that functions as both an enzyme and a molecular chaperone. Mostly studied in yeast and mammalian cells, recent studies have shown that in Arabidopsis and other plants, ER stresses causes upregulation of PDI (Houston et al., 2005; Meyer et al., 2008; Wilkinson and Gilbert, 2004; Zhang et al., 2018). PDI in its reduced or oxidized state can change the protein structure by changing the oxidation state of sulfide residues in target proteins. Furthermore, it can isomerize the sulfide bonds from one folded state to a different folded state (Wang et al., 2015).

RESULTS

cpd13/35 mutants have reduced plant height and early growth defects

The *cpd13* and *cpd35* mutants were identified from EMS mutagenized populations. After outcrossing the mutants to an inbred and self-fertilizing the F1

progeny, segregation analyses of F2 families revealed a Mendelian segregation ratio of 1:3 for mutant: wild-type plants. This demonstrated that the *cpd13* and *cpd35* mutations are recessive. Based on their similar phenotype, we performed complementation crosses to establish *cpd13* and *cpd35* are allelic. Furthermore, we later identified another recessive allele; *mu-il59132.7* from the Mu-illumina database that failed to complement the *cpd13* phenotype (Supplemental Figure S4.1). *cpd13* and its allelic mutants display reduced plant height at maturity (Figure 4.1). Similarly, the plants have growth defects early on, with differences in root growth observed in etiolated seedlings as early as 5 days after germination in both *cpd13* and *cpd35* mutants (Figure 4.1). Furthermore, the root:shoot ratio was also smaller in the *cpd13* mutant compared to the wild type.

cpd13/35 mutants have chlorotic leaves and hyperaccumulate soluble sugars and starch

Many of the mutants with defects of carbon export from leaves exhibit reduced plant height, leaf chlorosis, and anthocyanin accumulation similar to the phenotypes demonstrated by *cpd13* and *cpd35* mutants (Figure 4.2)(Baker and Braun, 2007, 2008; Braun et al., 2006; Julius et al., 2021; Julius et al., 2018; Slewinski and Braun, 2010b; Tran et al., 2019). Intriguingly, in *cpd13/35* the chlorosis and anthocyanin accumulation occur in a crossbanded pattern with alternating regions of green and chlorotic sectors. The chlorosis and anthocyanin accumulation occur in a basipetal manner, i.e., developing from the tip of the leaf towards the base, as the leaves develop (Figure 4.2).

Many of the carbohydrate partitioning defective mutants accumulate large amounts of soluble sugars and starch in their leaves (Baker and Braun, 2007, 2008; Braun et al., 2006; Julius et al., 2021; Julius et al., 2018; Tran et al., 2019). Hence, mature

leaves of *cpd13*, *cpd35* and wild-type siblings were harvested at the end of night, cleared of photosynthetic pigments, and stained with iodine potassium iodide (IKI) for detecting starch. Leaves of all *cpd13* mutant alleles showed dark regions in the chlorotic crossbands compared to the wild type denoting hyperaccumulation of starch (Figure 4.2, Supplemental Figure S4.1).

Furthermore, we quantified soluble sugars (sucrose, glucose, and fructose) as well as starch in *cpd13*, *cpd35*, and wild-type siblings mature leaf tissues collected at the end of night. Leaf regions were sampled from chlorotic tissues for the mutants and a comparable region of leaf was selected in the wild type. We found highly increased levels of starch and all three soluble sugars in the mutants compared with the wild type (Figure 4.2). Thus, *cpd13* and *cpd35* mutants accumulate elevated levels of carbohydrate in their leaves, which we hypothesize is due to a reduced rate of sucrose export from source leaves. Interestingly, *cpd13* and *cpd35* mutants occasionally exude sugary droplets from the leaves and stems, which is similar to the phenotype shown by the *sut1* mutant, which is defective in sugar transport from source leaves (Slewinski et al., 2010) (Supplemental Figure S4.2).

cpd13/35 mutants exhibit reduced photosynthesis

Since hyperaccumulation of carbohydrate in leaves can inhibit photosynthetic processes (Rolland et al., 2006), we measured photosynthetic rate, transpiration rate, and stomatal conductance of the most recently fully matured photosynthetic leaves from mutant and wild-type siblings. The *cpd13* and *cpd35* mutants demonstrated decreased photosynthetic rate, stomatal conductance and transpiration rate compared to the wild-

type plants (Figure 4.3). Hence, we identified reduced photosynthetic processes to be linked with carbohydrate hyperaccumulation in *cpd13* and *cpd35* mutants.

cpd13 and *cpd35* mutant phenotypes are induced by high temperature

Dna-J proteins are heat shock proteins, which are known to be induced by many stress factors, including heat. Hence to understand the effect of high temperature on the mutant phenotype, we grew *cpd13*, *cpd35*, and wild-type sibling plants in growth chambers with varying daytime temperatures; 25 °C for the low daytime temperature versus 35 °C for the high daytime temperature. Both used the same nighttime temperature of 22° C. At two weeks, we observed reduced growth, chlorosis and anthocyanin accumulation in the mutant plants grown at higher temperature compared to the wild-type plants (Supplemental Figure S4.3). However, we did not see any differences among the mutant and the wild-type plants grown at low temperature.

cpd13 and *cpd35* mutants have normal appearing vasculature

In many of the previously characterized mutants defective in carbohydrate partitioning, such as, *Cpd1* and *sxd1*, changes in vasculature morphology, such as lignin deposition or the buildup of callose occluding the flow of sugars were correlated with the carbohydrate hyperaccumulation phenotype (Julius et al., 2018; Russin et al., 1996). Hence, *cpd13* and *cpd35* mutant leaves exhibiting chlorotic phenotypes were harvested along with leaf tissues from wild-type siblings. For the mutant leaves, cross sections from the top, middle, and base regions including the chlorotic regions as well as green regions were observed. We did not observe any obvious morphological changes between the mutant and the wild-type leaf sections when examined under brightfield illumination (Supplemental Figure S4.4). To detect the presence of lignin, we stained the leaf cross

sections with phloroglucinol and observed the tissues under brightfield. We found an increase in lignin associated with the xylem and sclerenchyma cells in the mutant leaves but no ectopic lignin deposition in the phloem of mutant leaf veins (Supplemental Figure S4.4). Similarly, we did not find any ectopic callose deposition (as detected by aniline blue staining) (Supplemental Figure S4.4). Thus, these anatomical and histochemical results suggest that the carbohydrate accumulation phenotype is not associated with changes in the vein anatomy as observable with light microscopy.

CPD13 encodes a protein containing Dna-J and Thioredoxin-like domains

To understand the molecular mechanism that inhibits carbohydrate partitioning in the *cpd13* and *cpd35* mutants, a combination of whole genome sequencing and fine mapping approaches using polymorphic markers were employed (Supplemental Table S4.1). The location of the gene was narrowed down to a 312Kb region on the lower arm of Chromosome 2 containing 18 predicted protein-coding genes. A whole genome sequencing based approach was used to identify the causative mutations for each allele. We isolated DNA from two separate pools consisting of at least 40 mutant plants for each allele. We aligned the *cpd13* and *cpd35* DNA sequences against the Mo17 and B73 reference genomes, respectively, as these are the inbred backgrounds each mutant was generated in. We compared the DNA sequence in the mapped interval to identify any mutations in the genes inside this region. By comparing independent mutations that were unique to *cpd13* and to *cpd35*, we identified a single gene: Zm00001d007009 containing a base pair substitution in each allele. The *cpd13* mutation is a G to A point mutation affecting the splice acceptor site between 7th intron and 8th exon of the gene, whereas *cpd35* is a C to T point mutation causing a gain of stop codon in the 17th exon (Figure 4.

4). Reverse-Transcription-Polymerase Chain Reaction (RT-PCR) of RNA isolated from *cpd13* mutant and wild-type leaf samples revealed a shorter product for *cpd13*. Sequencing the *cpd13* RT-PCR product showed that a different splicing acceptor site was active in *cpd13* transcripts, leading to shift in reading frame and introducing a stop codon 4 amino acids downstream. (Figure 4.4).

The *Cpd13* gene is expressed in almost all tissues at moderate levels, with relatively high expression in the ear primordium, followed by leaves, internodes, embryo, and vegetative meristems (Supplemental Figure S4.5) (Walley et al., 2016). The lowest expression was in mature pollen among the tissues analyzed.

CPD13 localizes to the Endoplasmic Reticulum

Many DnaJ-like domain containing proteins localize to the ER where they function as molecular chaperones (Ohta and Takaiwa, 2014; Pulido and Leister, 2018). Furthermore, searches with the CPD13 amino acid sequence in the Uniprot database and Deeploc prediction software predicted the protein to localize to the ER (Almagro Armenteros et al., 2017). To determine whether CPD13 localizes to the ER, the red fluorescent protein (RFP) mCherry was translationally fused to CPD13 at the C terminus. The construct was expressed under the control of the Cauliflower Mosaic Virus 35S promoter (p35S::CPD13::mCherry) and transiently expressed in stable transgenic tobacco (*Nicotiana benthamiana*) leaves carrying an ER-localized green fluorescent protein (GFP) marker. The RFP signal of the fusion protein was present in the ER network (Figure 4.5). Upon comparison with the ER-GFP signal from the transgenic line, we found that the two fluorescent signals (RFP and GFP) overlap, hence indicating that CPD13 localizes to the ER (Figure 4.5).

CPD13 is hypothesized to be involved in protein folding and chaperone processes

CPD13 has Dna-J-like and thioredoxin-like domains, and the subcellular localization assay indicated that the protein localized to the ER. Hence, we performed analyses on the CPD13 amino acid sequences to predict conserved domains and transmembrane domains using the Uniprot database, CDvist, a sequence-based protein domain search tool, and NCBI conserved domain search tools (Adebali et al., 2014). All programs predicted the presence of DnaJ-like and thioredoxin-like domains with two transmembrane domains, one closer to the N terminus and the other towards the C terminus (Figure 4.4). Based on localization assays, and these predictions, a model for the CPD13 protein is presented (Figure 4.6). We hypothesize that the CPD13 protein interacts with other target proteins to facilitate their folding and processing.

DISCUSSION

The *cpd13* and *cpd35* mutants were identified from EMS mutagenized populations, hyperaccumulate carbohydrates in leaves, and exhibit chlorotic leaves with anthocyanin accumulation. Interestingly, the *cpd13* and *cpd35* mutants exhibit a crossbanding pattern in mature leaves with alternating chlorotic and green sectors. The mutant leaves also exhibit reduced chlorophyll content and low photosynthetic measurements, which are likely due to the accumulation of soluble sugars in the leaves, which are known to downregulate photosynthesis (Rolland et al., 2006). Also, the *cpd13* and *cpd35* mutants have reduced growth and fertility compared to the wild-type siblings. Interestingly, the growth defect is evident as early as 5 days, where we identified *cpd13* and *cpd35* mutants have shorter roots compared to the wild type. Furthermore, we observed sugar exudates on *cpd13* and *cpd35* mutant leaves and stems that is reminiscent

of another carbohydrate partitioning mutant *sut1* (Slewinski et al., 2010)(Supplemental figure 2). This phenotype is expected to occur due to the inhibition of phloem import of sucrose, which causes the buildup of sucrose in the apoplast, and ultimately in the xylem transpiration stream (Slewinski et al., 2010).

Based on the publicly available maize gene expression data, *Cpd13* is broadly expressed across multiple tissue types at moderate levels. This ubiquitous expression pattern of *Cpd13* suggests that it is involved in basic biological and cellular processes. Furthermore, its localization in the ER and the presence of DnaJ-like and thioredoxin-like domains suggest that CPD13 might be involved in protein processing and folding. CPD13 proteins lack the canonical histidine-proline-aspartic acid (HPD) domain that is conserved in DnaJ proteins that directly interact with the hsp70 complex where protein folding occurs (Kampinga and Craig, 2010). However, in human and yeast cells, Dna-J like proteins without the active J domain are involved in de novo protein folding (Kampinga and Craig, 2010). Similarly, many Dna-J-like proteins in plants have been demonstrated to act in a hsp70-indpendent manner (Pulido and Leister, 2018). Hence, it is likely that CPD13 interacts with other protein substrates independently and functions in their quality control. We present a detailed working model for the CPD13 protein and the mutated protein in Figure 4.6. To summarize the model, in the wild type, the DnaJ-thioredoxin-like protein can interact with unfolded target proteins and process it to be folded correctly. However, in the *cpd13* and *cpd35* mutants, the truncated protein, which is lacking the second transmembrane domain, causes the mutated CPD13 protein to not be correctly anchored to the ER membrane and prevents its proper function.

Due to the sugar exudation phenotype in the *cpd13* mutants that is similar to *sut1*, we hypothesize that SUT1 might be a substrate of CPD13 chaperone activity. Studies are underway to address this hypothesis. Similarly, if CPD13 functions as a chaperone, its loss of function will likely cause a buildup of client protein aggregates that have been misfolded or mistargeted. To identify such protein aggregations, the Proteostat Aggresome detection kit (Enzo Life Sciences), which has been widely used in Arabidopsis roots, will be tested (Cho and Kanehara, 2017; Karunadasa et al., 2020; Llamas et al., 2021). I plan to explore if similar assays can be performed with maize roots and leaves.

In conclusion, we have uncovered novel mutations in a DnaJ-thioredoxin-like domain containing protein that causes defects in carbohydrate partitioning. Studies are underway to further dissect the mechanism of how this ER resident protein can lead to backup of sugars and starch in the mature leaves. Once the mechanism is better understood, we can leverage that knowledge to improve plants.

METHODS

Plant materials and growth conditions

Maize (*Zea mays* L.) plants used for the microscopy, and soluble sugar and starch quantifications, and photosynthetic measurements were grown in the field at the South Farm Research Center at the University of Missouri, Columbia, MO. The *cpd13* and *cpd35* mutations were generated through EMS mutagenesis of pollen from the Mo17 and B73 inbreds, respectively, and all the plants analyzed were backcrosses to B73 at least 5 times. A third allele, *Mu-il59132.7,* with a *Mutator (Mu)* transposon insertion was obtained from the Mu-illumina library seed stock at Oregon State University

(http://teosinte.uoregon.edu/mu-illumina/) (Williams-Carrier et al., 2010). Etiolated plants used for early growth measurements were grown for 5 days under 24-hour dark conditions at 28°C in germination paper. Genomic DNA was extracted from plants and sequenced to genotype *cpd13* and *cpd35* homozygous mutants, heterozygotes, and homozygous wild-type individuals. After PCR amplification with allele specific primers (See Supplemental Table S4.1), PCR products were digested with DdeI at 37°C for 1 hour for *cpd13* mutants, and with TaqI at 65°C for *cpd35* mutants. The enzyme digested PCR products were loaded into a 4% super fine resolution (SFR) agarose gel and imaged. *Nicotiana benthamiana* plants used for transient protein expression via agroinfiltration were grown in a growth chamber under 14 h light and 10 h dark conditions at 18-25°C and 55-65% relative humidity for 4-5 weeks (Tran and Braun, 2017).

For the temperature response studies, *cpd13* and *cpd35* mutants and their wild-type siblings were grown in growth chambers at the East Campus Plant Growth Facility at the University of Missouri with 50% humidity, 22°C nighttime temperature for 10 hours and either 25°C or 35°C daytime temperature for 14 hours with the lights gradually increasing from 500 µmol m^{-2} s^{-1} to 1200 µmol m^{-2} s^{-1} in the first hour and decreasing from 1200 µmol m^{-2} s^{-1} to 500 µmol m^{-2} s^{-1} in the 13^{th} hour.

Mapping and whole genome DNA sequencing

Mapping population generation, bulked segregant analysis (BSA), and fine mapping was performed as described previously (Settles et al., 2014; Tran et al., 2019). Sequences of primers used for creating mapping markers can be found in Supp. Table 1. Whole genome sequencing was performed on DNA isolated from pools of *cpd13* and *cpd35* mutant plants. Leaf samples for DNA extraction were collected into two separate

pools consisting of at least 40 mutant plants for each allele. DNA was extracted and purified using the DNeasy Plant Mini Kit (Qiagen, USA), and sequenced at Psomagen, Inc. 150 bp Paired-end sequencing was performed with a Novaseq6000 (Illumina, USA). The raw reads were filtered and trimmed using Trimmomatic (Bolger et al., 2014). The reads were aligned against the B73 or Mo17 reference genome using Bowtie2 (Langmead and Salzberg, 2012), and the resulting SAM file was converted to BAM files with Samtools (Li et al., 2009). Then single nucleotide polymorphisms (SNPs) and Indels compared to the chromosomal region of interest were identified using Mpileup, and the predicted effects of these SNPs on gene function were annotated by SNPeff (Cingolani et al., 2012). From these analyses, we identified a single candidate gene containing a unique SNP in both mutant alleles.

To confirm the identified candidate gene from whole genome sequencing, we amplified the candidate sequence from DNA independently isolated again from both mutant alleles by PCR (Supplemental table S1). Diffinity RapidTips (Diffinity Genomic, USA) were used to purify the PCR products, which were sent for Sanger sequencing to GENEWIZ Inc., USA. The sequences of the PCR product were assembled and aligned against the *Cpd13* gene sequence using Lasergene sequence analysis software (DNASTAR Inc., USA).

Starch staining

The *cpd13*, *cpd35*, and *Mu-il59132.7* mutant and wild-type leaves were collected before sunrise. The leaves were cleared in ethanol, rinsed in water, and stained with 1% IKI as previously described (Baker and Braun, 2007).

Measurements of Sugar and Starch levels

Samples from mature source leaves of mutant and wild-type plants were harvested at 5:30 AM from field grown plants and immediately frozen in liquid nitrogen then kept at -80 C until ready to be processed. Soluble sugars and starch were extracted and quantified as previously described using high-performance anion-exchange (HPAE) chromatography (ICS-5000, Thermo-Fisher Scientific, USA) (Leach and Braun, 2016).

Photosynthesis, and gas exchange measurements

Photosynthesis rate, stomatal conductance, and transpiration rate were measured as previously described on 8 week old plants (Tran et al., 2019).

Histochemical analyses

Free-hand sections from fresh leaves were used to examine vein structure under bright-field and UV-light illumination as previously described (Baker and Braun, 2008). For callose staining, free hand sections from fresh leaves were incubated in Aniline blue solution (0.05% in 0.1 M K_2HPO_4 buffer, pH 9) for 5 min, then observed under bright-field and UV illumination. For lignin staining, free-hand sections from fresh leaves were incubated with phloroglucinol (0.0068 M in 13.7% HCl) for 20 min before observing under bright-light illumination. The samples were analyzed with a Nikon Eclipse 80i epifluorescence microscope (Nikon Instruments Inc., USA). All images from an experiment were captured using identical microscope and camera settings.

Subcellular localization analyses

The coding sequence for the CPD13 protein (Zm00001d007009_T001) was codon optimized for expression in *Nicotiana bethamiana* and synthesized in the pMK-T vector backbone (LifeTechnologies, USA). The sequence was recombined into pDONR-zeo vector and then ultimately into a modified pEarleyGate101 vector containing mCherry to generate the p35S::CPD13::mCherry translational fusion. The constructs were transformed into Agrobacterium tumefaciens GV3101. Agroinfiltration of tobacco leaves was performed as described previously (Tran et al., 2019).

References

Adebali, O., Ortega, D.R., Zhulin, I.B., 2014. CDvist: a webserver for identification and visualization of conserved domains in protein sequences. Bioinformatics 31(9), 1475-1477.

Almagro Armenteros, J.J., Sønderby, C.K., Sønderby, S.K., Nielsen, H., Winther, O., 2017. DeepLoc: prediction of protein subcellular localization using deep learning. Bioinformatics 33(21), 3387-3395.

Baker, R.F., Braun, D.M., 2007. *tie-dyed1* functions non-cell autonomously to control carbohydrate accumulation in maize leaves. Plant Physiol. 144(2), 867-878.

Baker, R.F., Braun, D.M., 2008. *Tie-dyed2* functions with *Tie-dyed1* to promote carbohydrate export from maize leaves. Plant Physiol. 146(3), 1085-1097.

Baker, R.F., Leach, K.A., Boyer, N.R., Swyers, M.J., Benitez-Alfonso, Y., Skopelitis, T., Luo, A., Sylvester, A., Jackson, D., Braun, D.M., 2016. Sucrose transporter *ZmSut1* expression and localization uncover new insights into sucrose phloem loading. Plant Physiol. 172(3), 1876-1898.

Bezrutczyk, M., Hartwig, T., Horschman, M., Char, S.N., Yang, J., Yang, B., Frommer, W.B., Sosso, D., 2018. Impaired phloem loading in *zmsweet13a,b,c* sucrose transporter triple knock-out mutants in *Zea mays*. New Phytol. 218(2), 594-603.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15), 2114-2120.

Braun, D.M., Ma, Y., Inada, N., Muszynski, M.G., Baker, R.F., 2006. *tie-dyed1* regulates carbohydrate accumulation in maize leaves. Plant Physiol. 142(4), 1511-1522.

Braun, D.M., Slewinski, T.L., 2009. Genetic control of carbon partitioning in grasses: roles of *Sucrose Transporters* and *Tie-dyed* loci in phloem loading. Plant Physiol. 149(1), 71-81.

Carpaneto, A., Geiger, D., Bamberg, E., Sauer, N., Fromm, J., Hedrich, R., 2005. Phloem-localized, proton-coupled sucrose carrier ZmSUT1 mediates sucrose efflux under the control of the sucrose gradient and the proton motive force*. J. Biol. Chem. 280(22), 21437-21443.

Cho, Y., Kanehara, K., 2017. Endoplasmic Reticulum stress response in Arabidopsis Roots. Front. Plant Sci. 8, 144-144.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin) 6(2), 80-92.

Cunnea, P.M., Miranda-Vizuete, A., Bertoli, G., Simmen, T., Damdimopoulos, A.E., Hermann, S., Leinonen, S., Huikko, M.P., Gustafsson, J.A., Sitia, R., Spyrou, G., 2003. ERdj5, an endoplasmic reticulum (ER)-resident protein containing DnaJ and thioredoxin domains, is expressed in secretory cells or following ER stress. J. Biol. Chem. 278(2), 1059-1066.

Dhungana, S.R., Braun, D.M., 2021. Sugar transporters in grasses: Function and modulation in source and storage tissues. J. Plant Physiol. 266, 153541.

Houston, N.L., Fan, C., Xiang, J.Q.-Y., Schulze, J.-M., Jung, R., Boston, R.S., 2005. Phylogenetic analyses identify 10 classes of the protein disulfide isomerase family in plants, including single-domain protein disulfide isomerase-related proteins. Plant Physiol. 137(2), 762-778.

Julius, B.T., Leach, K.A., Tran, T.M., Mertz, R.A., Braun, D.M., 2017. Sugar transporters in plants: New insights and discoveries. Plant Cell Physiol. 58(9), 1442-1460.

Julius, B.T., McCubbin, T.J., Mertz, R.A., Baert, N., Knoblauch, J., Grant, D.G., Conner, K., Bihmidine, S., Chomet, P., Wagner, R., Woessner, J., Grote, K., Peevers, J., Slewinski, T.L., McCann, M.C., Carpita, N.C., Knoblauch, M., Braun, D.M., 2021. Maize *Brittle Stalk2-Like3*, encoding a COBRA protein, functions in cell wall formation and carbohydrate partitioning. Plant Cell 33(10), 3348-3366.

Julius, B.T., Slewinski, T.L., Baker, R.F., Tzin, V., Zhou, S., Bihmidine, S., Jander, G., Braun, D.M., 2018. Maize *Carbohydrate partitioning defective1* impacts carbohydrate distribution, callose accumulation, and phloem function. J. Exp. Bot. 69(16), 3917-3931.

Kampinga, H.H., Craig, E.A., 2010. The HSP70 chaperone machinery: J proteins as drivers of functional specificity. Nat. Rev. Mol. Cell Biol. 11(8), 579-592.

Karunadasa, S.S., Kurepa, J., Shull, T.E., Smalle, J.A., 2020. Cytokinin-induced protein synthesis suppresses growth and osmotic stress tolerance. New Phytol. 227(1), 50-64.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9(4), 357-359.

Leach, K.A., Braun, D.M., 2016. Soluble sugar and starch extraction and quantification from maize (*Zea mays*) leaves. Current Protocols in Plant Biology 1(1), 139-161.

Leach, K.A., Tran, T.M., Slewinski, T.L., Meeley, R.B., Braun, D.M., 2017. *Sucrose transporter2* contributes to maize growth, development, and crop yield. J. Integr. Plant Biol. 59(6), 390-408.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Subgroup, G.P.D.P., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16), 2078-2079.

Llamas, E., Torres-Montilla, S., Lee, H.J., Barja, M.V., Schlimgen, E., Dunken, N., Wagle, P., Werr, W., Zuccaro, A., Rodríguez-Concepción, M., Vilchez, D., 2021. The intrinsic chaperone network of Arabidopsis stem cells confers protection against proteotoxic stress. Aging Cell 20(8), e13446.

Meyer, Y., Siala, W., Bashandy, T., Riondet, C., Vignols, F., Reichheld, J.P., 2008. Glutaredoxins and thioredoxins in plants. Biochim. Biophys. Acta 1783(4), 589-600.

Miernyk, J.A., 2001. The J-domain proteins of *Arabidopsis thaliana*: an unexpectedly large and diverse family of chaperones. Cell Stress Chaperones 6(3), 209-218.

Ohta, M., Takaiwa, F., 2014. Emerging features of ER resident J-proteins in plants. Plant Signal. Behav. 9(3), e28194-e28194.

Pulido, P., Leister, D., 2018. Novel DNAJ-related proteins in *Arabidopsis thaliana*. New Phytol. 217(2), 480-490.

Rolland, F., Baena-Gonzalez, E., Sheen, J., 2006. Sugar Sensing and Signaling in plants: Conserved and novel mechanisms. Annu. Rev. Plant Biol. 57(1), 675-709.

Russin, W.A., Evert, R.F., Vanderveer, P.J., Sharkey, T.D., Briggs, S.P., 1996. Modification of a specific class of plasmodesmata and loss of sucrose export ability in the *sucrose export defective1* maize mutant. Plant Cell 8(4), 645-658.

Settles, A.M., Bagadion, A.M., Bai, F., Zhang, J., Barron, B., Leach, K., Mudunkothge, J.S., Hoffner, C., Bihmidine, S., Finefield, E., Hibbard, J., Dieter, E., Malidelis, I.A., Gustin, J.L., Karoblyte, V., Tseung, C.-W., Braun, D.M., 2014. Efficient molecular marker design using the MaizeGDB Mo17 SNPs and indels track. G3 Genes|Genomes|Genetics 4(6), 1143-1145.

Slewinski, T.L., Baker, R.F., Stubert, A., Braun, D.M., 2012. *Tie-dyed2* encodes a Callose Synthase that functions in vein development and affects symplastic trafficking within the phloem of maize leaves Plant Physiol. 160(3), 1540-1550.

Slewinski, T.L., Braun, D.M., 2010a. Current perspectives on the regulation of whole-plant carbohydrate partitioning. Plant Sci. 178(4), 341-349.

Slewinski, T.L., Braun, D.M., 2010b. The *Psychedelic* genes of maize redundantly promote carbohydrate export from leaves. Genetics 185(1), 221-232.

Slewinski, T.L., Garg, A., Johal, G.S., Braun, D.M., 2010. Maize *SUT1* functions in phloem loading. Plant Signal. Behav. 5(6), 687-690.

Slewinski, T.L., Meeley, R., Braun, D.M., 2009. *Sucrose transporter1* functions in phloem loading in maize leaves. J. Exp. Bot. 60(3), 881-892.

Tran, T.M., Braun, D.M., 2017. An inexpensive, easy-to-use, and highly customizable growth chamber optimized for growing large plants. Current Protocols in Plant Biology 2(4), 299-317.

Tran, T.M., McCubbin, T.J., Bihmidine, S., Julius, B.T., Baker, R.F., Schauflinger, M., Weil, C., Springer, N., Chomet, P., Wagner, R., Woessner, J., Grote, K., Peevers, J., Slewinski, T.L., Braun, D.M., 2019. Maize *carbohydrate partitioning defective33* encodes an MCTP protein and functions in sucrose export from leaves. Mol. Plant 12(9), 1278-1293.

Walley, J.W., Sartor, R.C., Shen, Z., Schmitz, R.J., Wu, K.J., Urich, M.A., Nery, J.R., Smith, L.G., Schnable, J.C., Ecker, J.R., Briggs, S.P., 2016. Integration of omic networks in a developmental atlas of maize. Science 353(6301), 814-818.

Wang, L., Wang, X., Wang, C.-c., 2015. Protein disulfide–isomerase, a folding catalyst and a redox-regulated chaperone. Free Radical Biol. Med. 83, 305-313.

Wilkinson, B., Gilbert, H.F., 2004. Protein disulfide isomerase. Biochim. Biophys. Acta 1699(1), 35-44.

Williams-Carrier, R., Stiffler, N., Belcher, S., Kroeger, T., Stern, D.B., Monde, R.A., Coalter, R., Barkan, A., 2010. Use of Illumina sequencing to identify transposon insertions underlying mutant phenotypes in high-copy Mutator lines of maize. Plant J. 63(1), 167-177.

Yamamoto, M., Maruyama, D., Endo, T., Nishikawa, S., 2008. *Arabidopsis thaliana* has a set of J proteins in the endoplasmic reticulum that are conserved from yeast to animals and plants. Plant Cell Physiol. 49(10), 1547-1562.

Zhang, Z., Liu, X., Li, R., Yuan, L., Dai, Y., Wang, X., 2018. Identification and functional analysis of a Protein Disulfide Isomerase (*At*PDI1) in *Arabidopsis thaliana*. Front. Plant Sci. 9.

Zimmermann, M., Ziegler, H., 1975. List of sugars and sugar alcohols in sieve-tube exudates, in: Zimmermann, M., Milburn, H. (Eds.), Encyclopedia of Plant Physiology. Springer, New York, pp. 480–503.

TABLES

Table S4. 1

List of Primers used for mapping, genotyping, and sequencing

Primer Name	Primer Sequence 5'-3'	Purpose
cpd13-F1	GCAATCCACTTCCCTTACTG	Candidate gene sequencing and genotyping PCR with enzyme digest
cpd13-R1	GGATCATTTGGCCGTTGA	
cpd35-F2	GCACTAGAAGGTAGGAAACC	
cpd35-R2	AGTTTGCAGGTGTAGATTGA	
cpd35-E15-R	TGCAACTGCCTTCAAAGAT	RT-PCR
cpd13-E7-R	CCTCGGTGTGCTCAATTT	RT-PCR
cpd13-F2	GCGAAATAACCAGCCACT	RT-PCR
cpd35-F1	CCTTGTACCTTGCCACTAAC	RT-PCR
IDP8680 F	GGTCCCTTGATGTCCATGC	Mapping
IDP8680 R	GAAGCAGGTCGTGATGTTCC	Mapping
INS.41853 F	TGTTGGGTTCGGTTGAAG	Mapping
INS..41853 R	GTGTCCGGATCTATTTCTGAC	Mapping
umc1633 F	GTCCTTCCTCTCCTTCGTGCATA	Mapping
umc1633 R	CAGAGGCTGTTGTTCCCCAC	Mapping
DEL.51560 set3 F	GGCACCGTATCCGCTCTTTA	Mapping
DEL.51560 set3 R	TCCAGGTGGGATGGGAGAT	Mapping
INS.41789 set2 F	TCGCAGCCCTGCAAACAA	Mapping
INS.41789 set2 R	GCCGAGGACATCATCAAGCAA	Mapping
INS.41801 set1 F	CGCGCAATTAATAGTGCAAA	Mapping
INS.41801 set1 R	CTTTCGGTCGAAACCAGTAT	Mapping

FIGURES

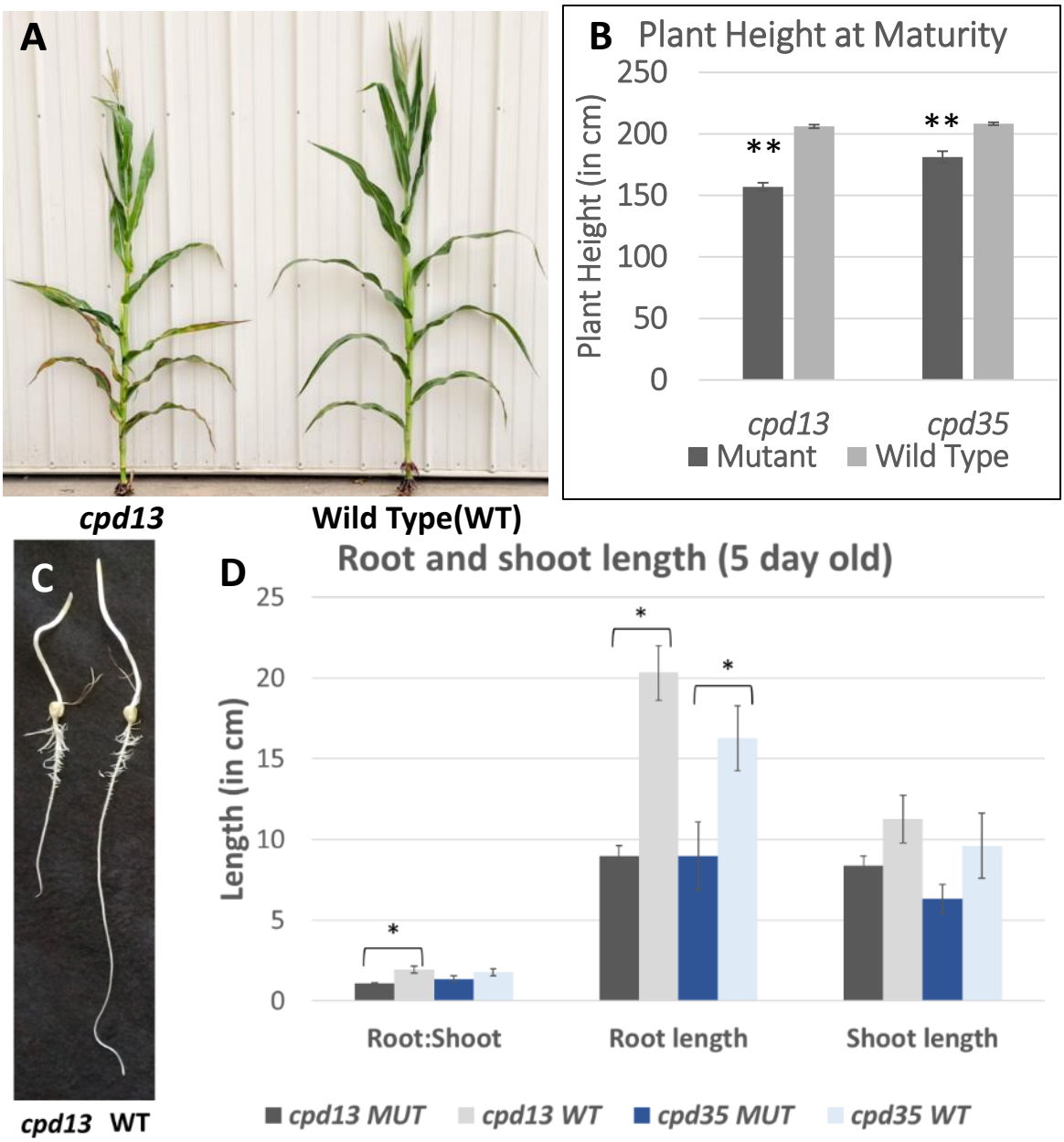
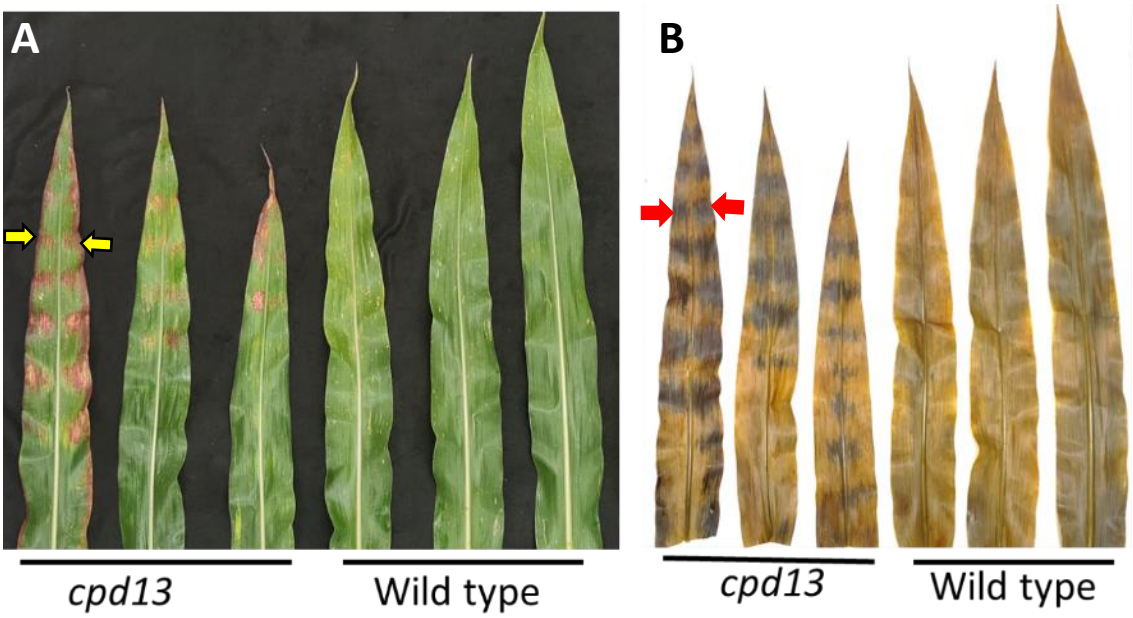


Figure 4.1

cpd13 and *cpd35* mutants exhibit reduced growth at maturity and early stages. (A) Image of *cpd13* and wild-type (WT) plants. (B) Height measurements of *cpd13* and *cpd35* mutant plants and wild-type siblings at maturity. (C) 5-day old seedlings of *cpd13* and

wild-type plants grown in the dark (D) Measurements of root and shoot length of 5 day old seedlings grown in the dark **denotes significant difference based on t-test at $\alpha = 0.01$, $n \geq 28$ whereas * denotes significant difference based on t-test at $\alpha = 0.05$, $n \geq 3$



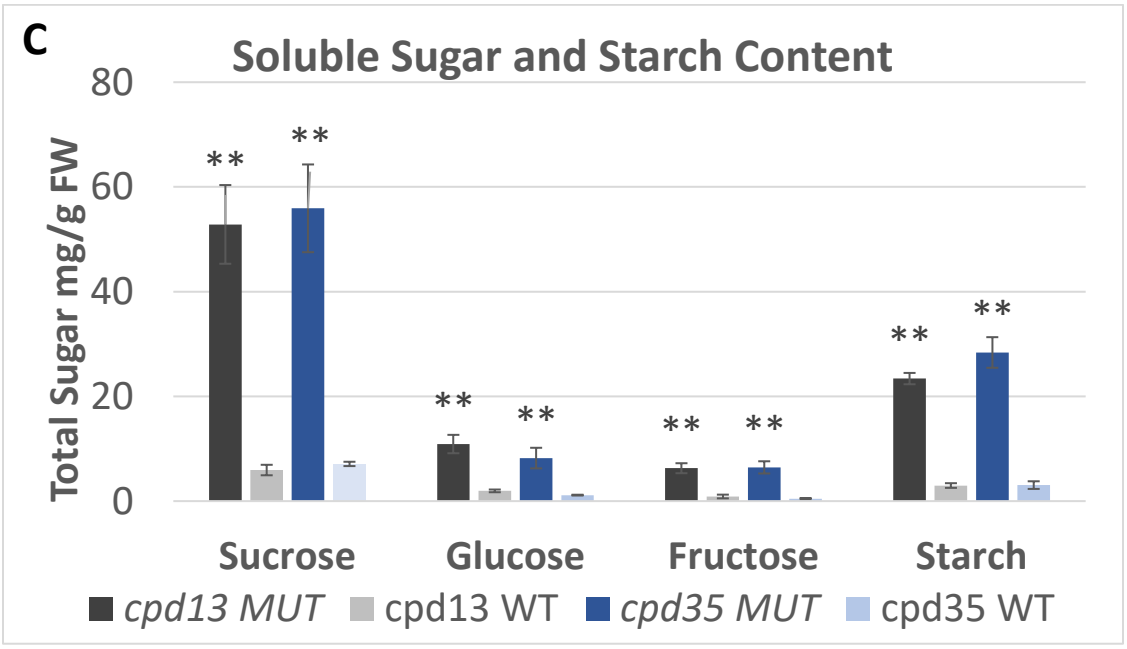


Figure 4.2

cpd13 mutant leaves exhibit chlorosis and anthocyanin accumulation and hyperaccumulate soluble sugars and starch in the crossbanded regions. (A) *cpd13* mutant and wild-type (WT) leaves ranging from mature to younger (left to right); yellow arrows

indicate chlorotic regions with anthocyanin accumulation. The younger leaves start with anthocyanin accumulation and chlorosis at the tip and the phenotype progresses towards the base as they mature. (B) Leaves from panel A after starch staining. Red arrows point to the regions with starch accumulation. (C) Soluble sugars and starch (glucose equivalents) levels in *cpd13* and *cpd35* mutant and wild-type leaves. ** denotes significant difference at $\alpha = 0.01$, $n = 3$ based on a t-test between mutant and wild type of each allele. Error bars denote standard error.

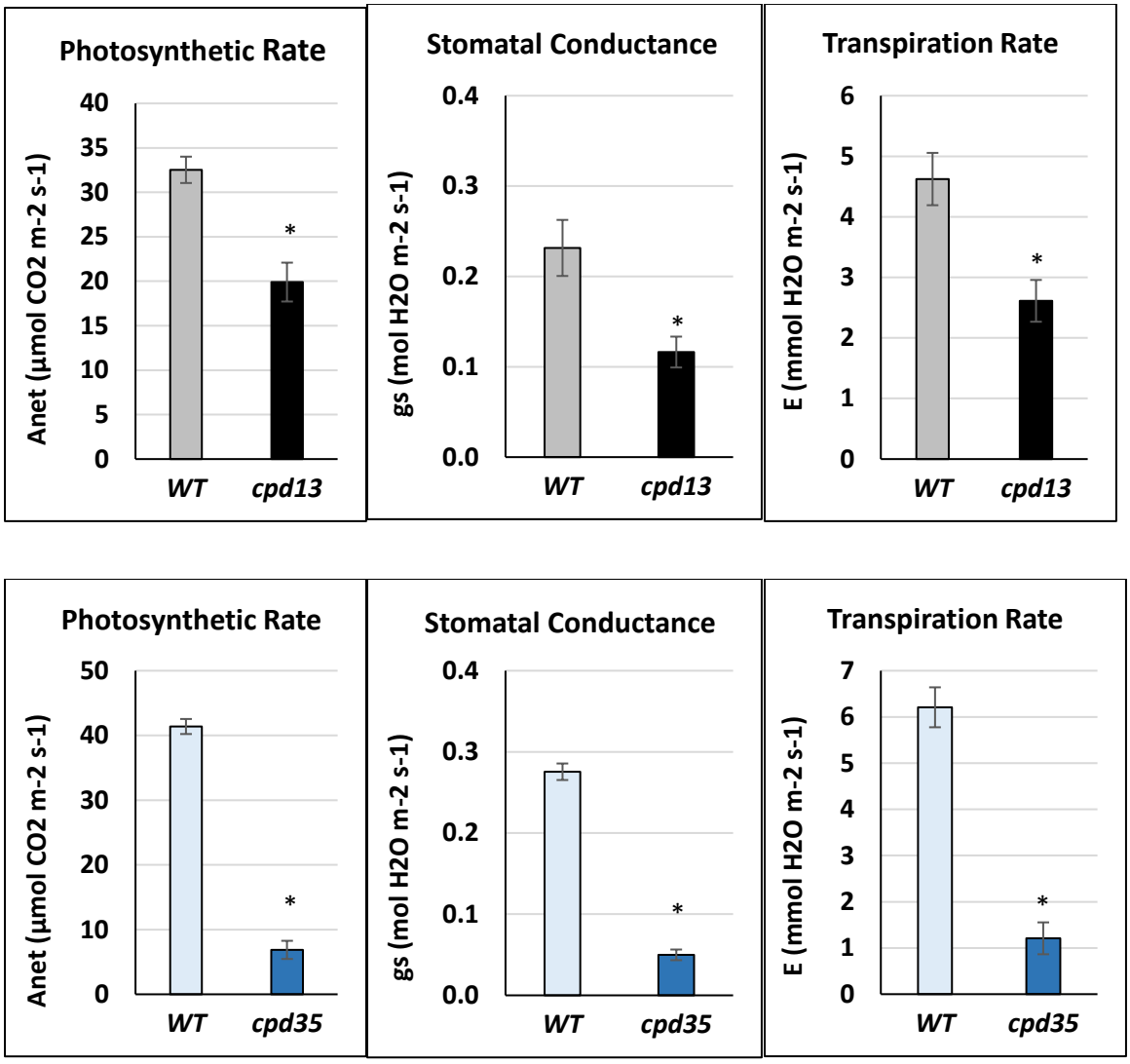


Figure 4.3

cpd13 and *cpd35* mutants have reduced photosynthetic rate, stomatal conductance, and transpiration rates. Average measurements for each parameter are shown for each category. Error bars denote standard error. * denotes significant difference based on a t-test at $\alpha = 0.05$, $n = 8$

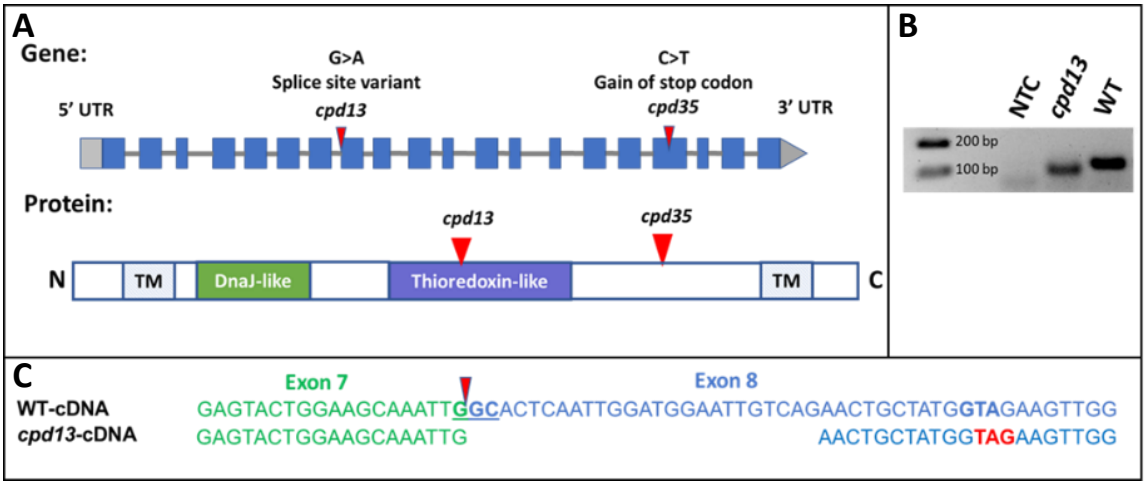


Figure 4.4

Representation of *cpd13* and *cpd35* mutations in Dna-J-thioredoxin-like gene. (A) Gene model (top) and protein model (bottom) of the DnaJ-thioredoxin-like gene which is mutated in the *cpd13* and *cpd35* mutants. The sites of mutations in *cpd13* and *cpd35* mutants are denoted by red arrows. (B) RT-PCR gel showing differences in band size of *cpd13* cDNA and the WT cDNA. NTC = no added template (water) control. (C) Sequences of *cpd13* and WT cDNA samples. The splice site in shown by red arrow, which is where *cpd13* mutation lies. The stop codon introduced due to the frame shift is indicated in red letters. UTR = Untranslated region, TM = Transmembrane domain, N = NH_2 terminus, and C = COOH terminus of the protein.

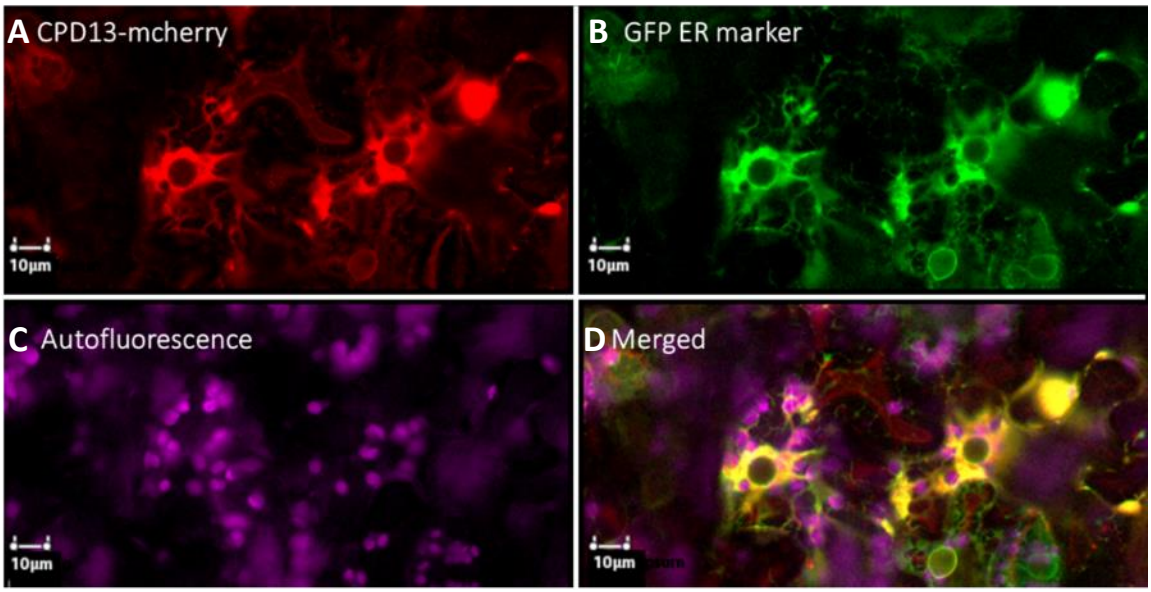


Figure 4.5

CPD13 localizes to the Endoplasmic Reticulum. (A) Transient expression of p35S::CPD13::mCherry in *Nicotiana benthamiana* leaves stably expressing a (B) ER-GFP marker. (C) Chlorophyll autofluorescence. (D) Merged signals from all channels. Yellow regions indicate the overlap of signals from both CPD13-mCherry and the ER-GFP marker.

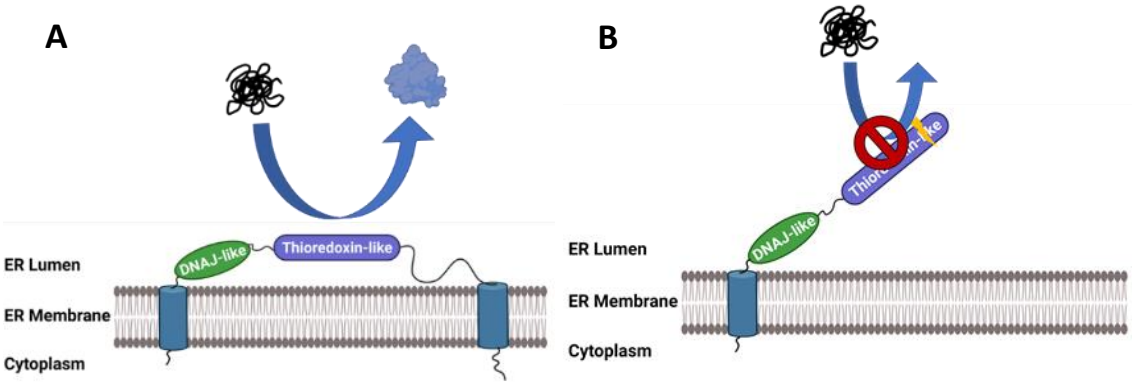


Figure 4.6

Working model of (A) the Dna-J-thioredoxin-like protein in the wild type, and of the (B) defective protein in *cpd13* and *cpd35* mutants. In the wild-type plants, the Dna-J-thioredoxin-like protein can interact with unfolded target proteins and process them to be folded correctly. In the *cpd13* and *cpd35* mutants, the DNA-J-thioredoxin-like protein is truncated, causing the loss of the second transmembrane domain as well as truncation of the thioredoxin-like domain (in the case of *cpd13*), which may cause the protein to not be anchored properly and unable to interact with target proteins. This loss of interaction prevents the target proteins from being folded and processed properly, which is hypothesized to cause an aggregation of unfolded proteins in the ER. This protein accumulation in that ER may interfere with the normal trafficking of proteins involved in sugar transport, causing the buildup of carbohydrates in the leaves and inducing the carbohydrate partitioning defective phenotype that we observe in the mutants.

Supplemental Figures

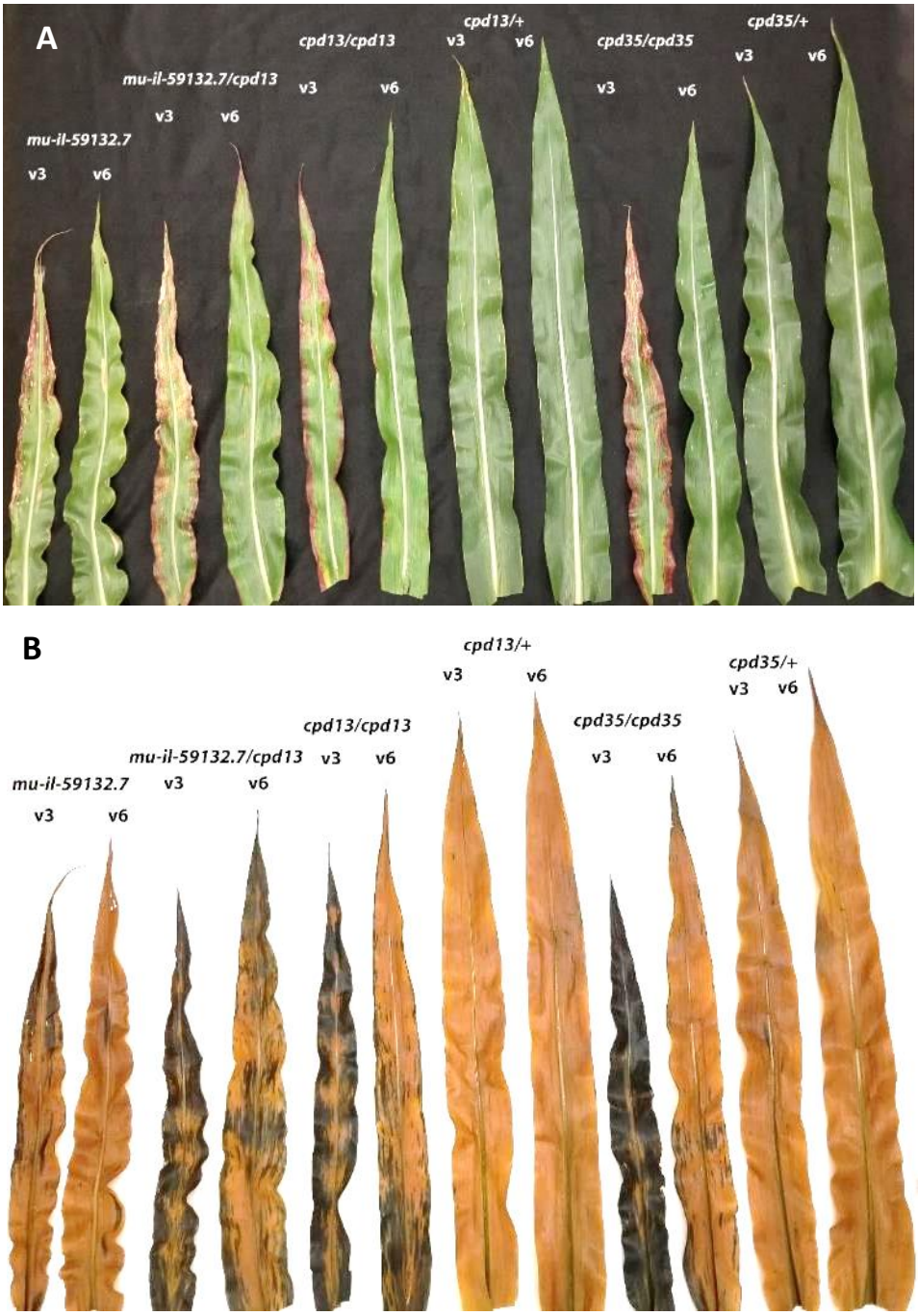
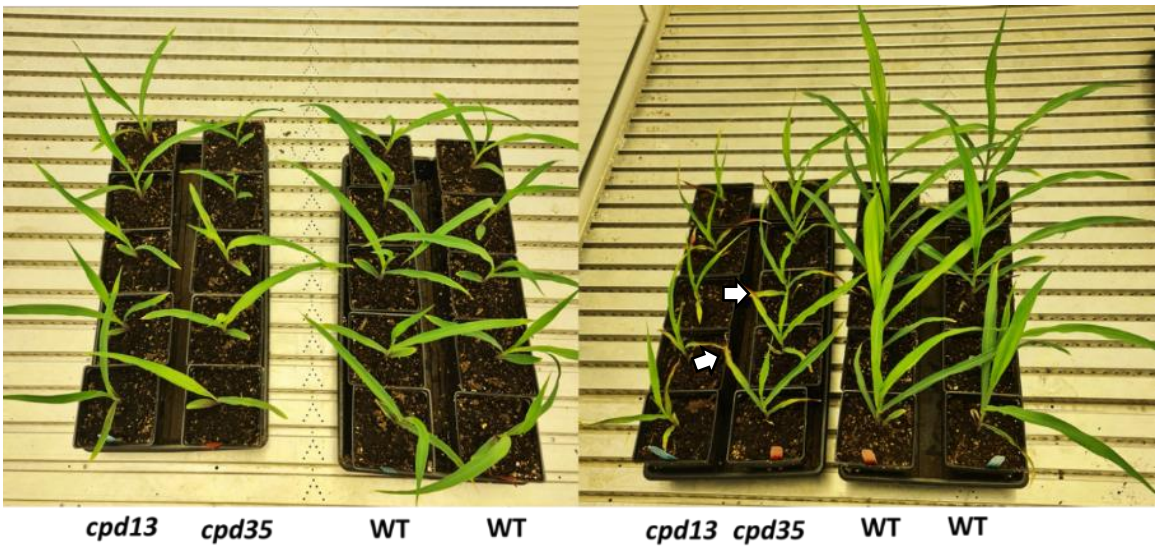


Figure S4. 1

cpd13 and its allelic mutants (A) before and (B) after starch staining. *cpd13* and its allelic mutants exhibit leaf chlorosis and starch accumulation phenotypes. The leaves were collected at various stages of phenotype progression (v3 and v6) and stained for starch as described in the methods section. The *Mu.il-59132.47* allele failed to complement the *cpd13* mutation and showed starch accumulation and chlorosis in the F1 progeny from crossing with *cpd13* mutants.

Figure S4. 2

cpd13 mutant leaves and stems occasionally exude sugary droplets (denoted by red arrows).



Low Temperature (25°C) **High Temperature (35°C)**

Figure S4. 3

cpd13 and *cpd35* mutant plants with their wild-type siblings grown in growth chambers at low temperature (25°C daytime, left) and high temperature (35°C daytime, right). Regions with chlorosis and anthocyanin accumulation are denoted by white arrows.

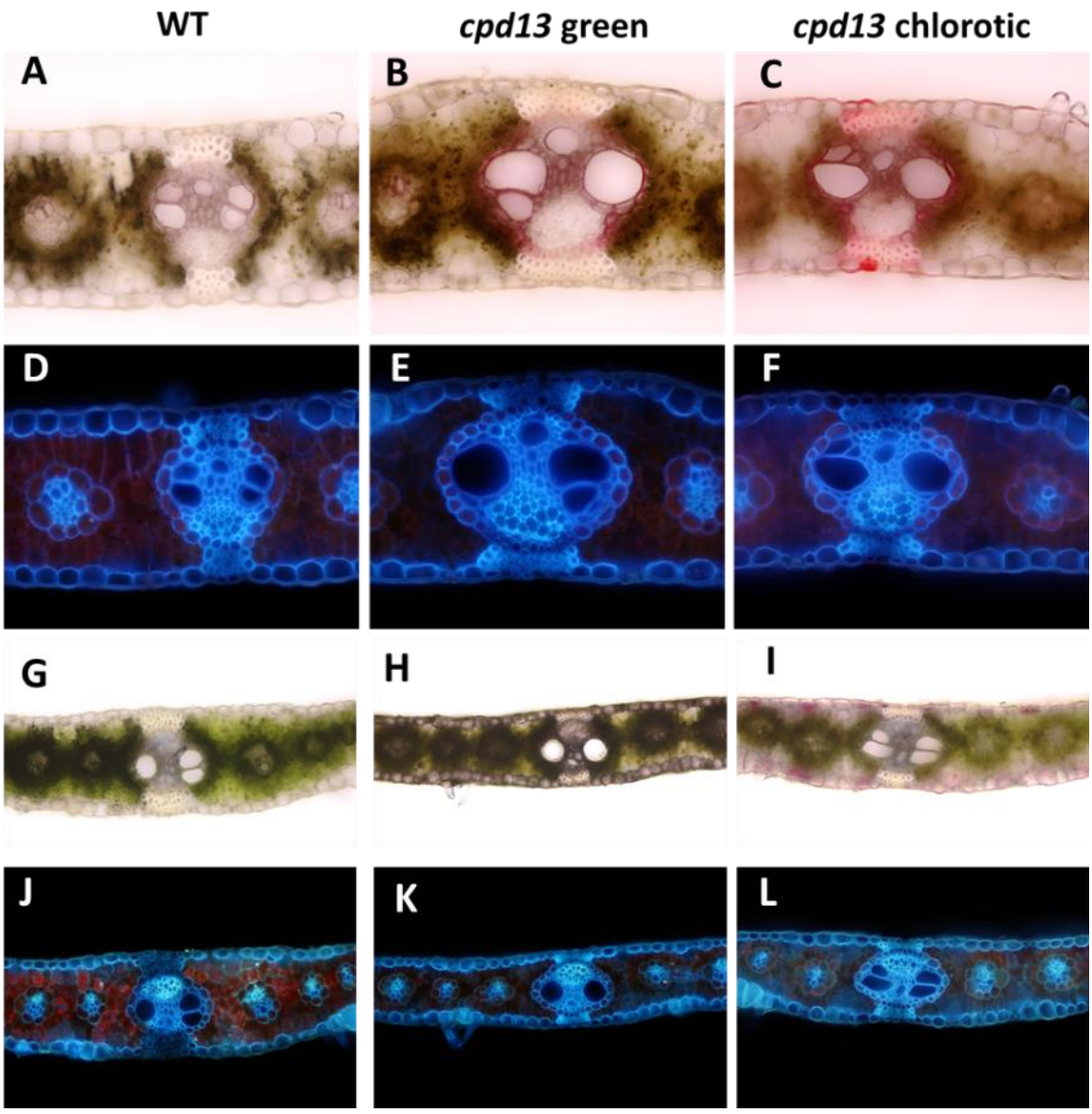


Figure S4. 4

cpd13 mutant leaves have normal appearing vasculature. Cross sections from mature leaves of wild-type plants (A, D, G, J), *cpd13* green sectors (B,E,H,K), and *cpd13* chlorotic sectors (C, F, I, L) are shown. (A, B, C) Phloroglucinol-HCl stained leaf cross sections under brightfield. Red stain denotes presence of lignin. (D, E, F) Same leaf sections viewed under UV illumination. (G, H, I) Aniline blue staining of leaf cross

sections under bright-field. (J, K, L) Same leaf sections imaged under UV illumination. No callose deposits were detected, which would have appeared as bright blue-white dots.

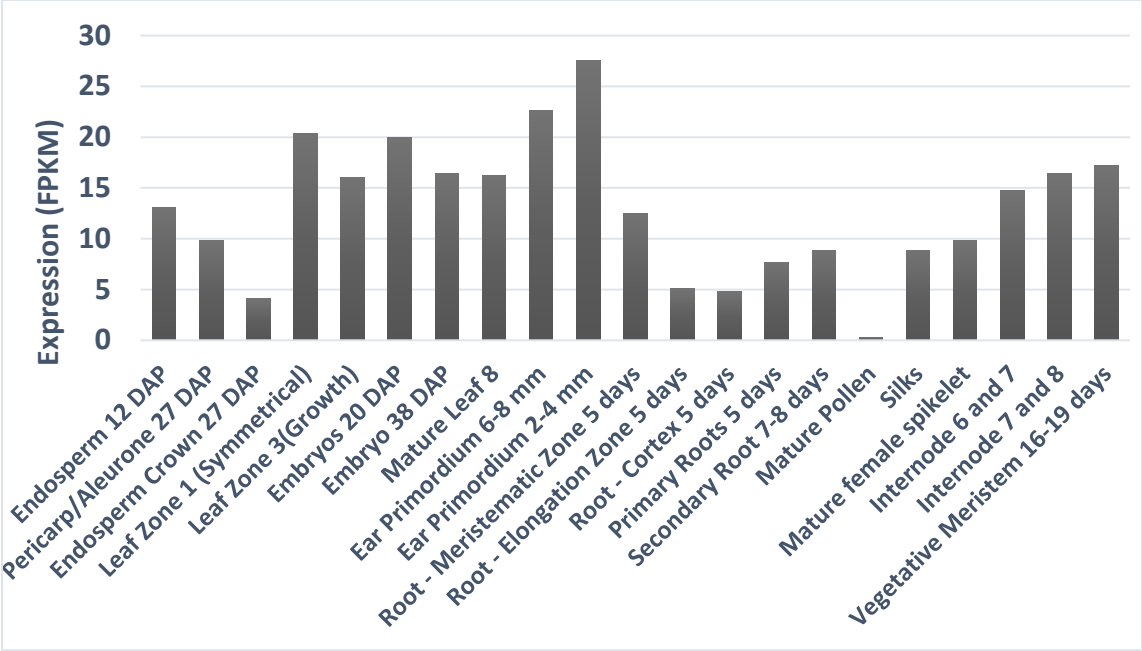


Figure S4. 5

Expression Levels of *Zm00001d007009* transcripts in various tissues determined by RNA-seq (Adapted from Walley et. al, 2016).

Chapter 5 : Maize *carbohydrate partitioning defective60* mutant exhibits lignification of phloem and defects in carbohydrate transport

INTRODUCTION

In the previous chapter, I presented the *cpd13* and *cpd35* mutants, which were two mutants out of the large catalog of carbohydrate partitioning defective mutants generated by EMS mutagenesis, and I elaborated the importance of carbohydrate partitioning (Chapters 1 and 4). Towards gaining a better understanding of the genetic control of carbohydrate partitioning, I also undertook the characterization of *carbohydrate partitioning defective60 (cpd60)* and *cpd87* mutants.

cpd60 and *cpd87* are recessive mutations and are allelic, as determined by complementation tests. We further identified two additional alleles *cpd87-like* and *pr14-cw-334,* which failed to complement the *cpd60* mutant. As with other mutants defective in carbohydrate partitioning, we observed a suite of phenotypes such as dwarfed plants, reduced fertility, chlorotic leaves, anthocyanin accumulation, and hyperaccumulation of starch in *cpd60* and its allelic mutants (Baker and Braun, 2008; Braun et al., 2006; Julius et al., 2021; Julius et al., 2018; Slewinski and Braun, 2010; Tran et al., 2019). Interestingly, we also observed ectopic lignin in various phloem cell types, including sieve elements, which are the conduits through which sucrose is transported long distance. Here, I discuss the characterization of *cpd87* and *cpd60* mutants in relation to carbohydrate partitioning.

RESULTS

Plants with *cpd60* and its allelic mutations are dwarfed and exhibit reduced fertility and leaf chlorosis

The plants with *cpd60* and its allelic mutations display reduced height and a greatly reduced tassel (Figure 5.1, Supplemental Figure S5.1). The homozygous recessive mutants rarely make ears, and even when present are severely reduced and not fertile. Furthermore, the leaves of the mutants are chlorotic and accumulate anthocyanins (Figure 5.1, Supplemental figure S5.1). The chlorosis and anthocyanin accumulation occurs in a basipetal manner, i.e., developing from the tip of the leaf towards the base, as the leaves develop (Figure 5.1). Hence, we observe the development of chlorosis and anthocyanin in the leaves as they mature.

cpd60/87 mutants hyperaccumulate soluble sugars and starch

Since high levels of starch accumulate in mature leaves of many other carbohydrates partitioning defective mutants, *cpd60* and *cpd87* leaves were also examined for starch accumulation by staining with IKI (Braun et al., 2006; Julius et al., 2021; Julius et al., 2018; Tran et al., 2019). Leaves of both *cpd60* and *cpd87* mutants showed dark staining in the chlorotic regions compared to the wild type denoting hyperaccumulation of starch (Figure 5.2, Supplemental figure S5.2). Furthermore, we quantified the soluble sugars and starch levels in the mature leaves of *cpd60* and *cpd87* mutants and their wild-type siblings collected at the End of Night (EON). We observed highly elevated levels of sucrose, glucose, fructose, and starch in the mutant leaves compared to the wild type (Figure 5.2, Supplemental Figure S5.2). Hence, *cpd60* and *cpd87* mutants hyperaccumulate soluble sugars and starch in their leaves.

Sucrose export is reduced in *cpd60* mutant leaves

We observed high levels of sucrose in *cpd60* mutant leaves, which we hypothesize is due to the decreased export of sucrose from the mature leaves as observed in other mutants defective in carbohydrate partitioning (Julius et al., 2021; Julius et al., 2018; Slewinski et al., 2009; Tran et al., 2019). To test whether *cpd60* leaves have reduced export of sucrose, we used radioactively labeled ^{14}C-sucrose for transport assays. The labeled sucrose was applied to the tip of a mature leaf on an intact plant and allowed to transport for 1 h. The leaf was then excised and autoradiographed using phosphor plates. In the wild-type leaves, considerable amount of ^{14}C-sucrose was transported down the length of the blade, but in the *cpd60* mutant leaves, the transport of the labeled sucrose was reduced (Figure 5.3). This indicates that sucrose export is inhibited in the *cpd60* mutant leaves.

***cpd60/87* mutants exhibit reduced photosynthesis**

As it is a well-known phenomenon that hyperaccumulation of sugars in the leaves inhibits photosynthesis (Rolland et al., 2006), we measured the photosynthetic rate, transpiration rate, and stomatal conductance of the most recently fully matured photosynthetic leaves from mutant and wild-type siblings. The *cpd60* and *cpd87* mutants demonstrated decreased photosynthetic rate, stomatal conductance, and transpiration rate compared to the wild-type plants (Figure 5.4). Hence, we identified reduced photosynthetic processes to be linked with carbohydrate hyperaccumulation in *cpd60* and *cpd87* mutants.

cpd60/87 mutants exhibit lignified phloem cells

As described in Chapter 4, in many of the previously characterized mutants defective in carbohydrate partitioning, such as *Cpd1* and *sxd1*, changes in the anatomy of the vasculature, such as lignin deposition or buildup of callose occluding the flow of sugars in the phloem were correlated with the carbohydrate hyperaccumulation phenotype (Julius et al., 2018; Russin et al., 1996). Hence, we examined *cpd60* (data not shown) and *cpd87* mutant leaf tissue sections along with leaf tissues harvested from wild-type siblings. Immature leaves (not expressing the phenotype) and mature leaves (expressing the phenotype) of mutants and similar aged leaves of wild-type plants were harvested, sectioned, and stained using Phloroglucinol for visualizing lignin. Ectopic lignin was found in phloem tissues of major veins and minor veins of mature mutant leaves but absent in immature leaves (Figure 5.5). Staining for callose using aniline blue was also performed on similar leaf cross sections, but no differences were observed between the mutant and wild type plants (data not shown). Such ectopic lignin deposition in the phloem of mature leaves has previously been observed in other *cpd* mutants as well (Braun lab, unpublished).

Mapping the *cpd60* mutant and its allelic mutations

The *cpd60, cpd87, cpd87-like* and *pr14-cw334* mutants originated in a B73 background, either spontaneously or from EMS mutagenesis. A mapping population was generated by backcrossing these mutants to the Mo17 background in order to identify the causative mutation for the mutant phenotype. The mutation responsible for the *cpd60* phenotype has been mapped to the long arm of Chromosome 1 by Bulked Segregant Analysis (BSA). By using polymorphic markers, we fine mapped the causative mutation

to a 600 kb region, which is predicted to contain eight protein coding sequences (Table 5.1). To further narrow down the region, we performed whole genome sequencing on all four mutants; *cpd60*, *cpd87*, *cpd87-like*, and *pr14-cw-334*, but we could not identify any unique Single Nucleotide Polymorphisms (SNPs) in the coding region of these eight genes in any of the mutants within the mapped region. However, for all of these mutants, there were SNPs in the intergenic regions. While the mutations in these intergenic regions could be affecting many genes, we narrowed down our candidate gene list by looking for DNA regions containing independent new SNPs nearby a gene in the majority of the mutants. For *cpd87*, *cpd87-like* and *pr14-cw334* mutants, most of the high scoring SNPs were present in the intergenic region between *Zm00001d032298* and *Zm00001d032300*. These SNPs were predicted to affect the expression of *Zm00001d032298*, *Zm00001d032299*, and *Zm00001d032300*, however, no SNPs in *cpd60* were present in the same region. For *cpd60*, we only found a SNP nearby the *Zm00001d032293* gene, but no SNPs nearby that gene in the other alleles, hence we did not consider that gene further. For other genes in the mapping interval, we did not find a region nearby each gene that had unique SNPs in all or most of the mutants, lowering our ranking of these being the gene responsible for the phenotype.

Among the genes that had SNPs nearby in multiple mutants, Zm*00001d032298* encodes *trehalose6-phosphate phosphatase1 (tpp1)*, which belongs to the *trehalose6-phosphate phosphatase* gene family known to be involved in sugar signaling and controlling carbohydrate metabolism (Nuccio et al., 2015; Smeekens, 2015). Additionally, Zm*00001d032300* encodes an uncharacterized member of the Major Facilitator Superfamily (MFS) gene family, which is the same large superfamily that

Sucrose Transporters and many other sugar transporters are part of, and many of these genes are involved in transporting carbohydrates (Niño-González et al., 2019). Zm*00001d032298* is expressed in mature leaves and embryo whereas Zm*00001d032300* is expressed in almost all tissues including leaves, vegetative meristem, and roots at some levels (Table 5.1). *Zm00001d032299* is not expressed in any tissues in multiple transcriptome datasets, hence we did not analyze it further (Table 5.1) (Johnston et al., 2014; Stelpflug et al., 2016; Walley et al., 2016). To identify the gene mutated in *cpd60* and its allelic mutants, we generated CRISPR-Cas9 knockouts of Zm*00001d032298* (*tpp1)* and Zm*00001d032300* (*MFS*). T0 plants that had a base pair deletion in an exon of Zm*00001d032298 (tpp1)* causing a frameshift did not display any chlorosis or anthocyanin accumulation in their leaves when grown to maturity in both greenhouse or field settings. Currently, T0 plants that have large deletions (80 bp) or single base deletions in the second exon of Zm*00001d032300* (*MFS*) are growing in the greenhouse; however, they are at early growth stages, ranging from 3 leaf stage to v4, and a clear *cpd* mutant phenotype is not evident yet.

DISCUSSION

Here, we describe the recessive mutants *cpd60*, *cpd87*, and their allelic mutations, which display reduced sucrose export and carbohydrate accumulation in their mature leaves. Furthermore, we observed reduced plant height, decreased fertility, chlorosis and anthocyanin accumulation in the leaves, which are all phenotypes associated with defects in carbohydrate partitioning as observed in other mutants (Julius et al., 2021; Tran et al., 2019). Interestingly, we observed ectopic lignification of phloem cell walls in the mature leaves of the mutant plants displaying the chlorotic phenotype. Such lignification of

phloem was observed in *Cpd1* mutants (Julius et al., 2018) as well as in other *cpd* mutants in the Braun lab that are defective in carbohydrate partitioning (unpublished). We observed the ectopic lignin deposition only in mature regions of the leaves in both major and minor veins but not in immature leaves. This suggests that the lignin accumulation could be a secondary effect of hyperaccumulated carbohydrates instead of being the cause for carbohydrate accumulation phenotype. Further studies are needed to understand this response as a result of the defects in carbohydrate partitioning.

Based on the whole genome sequencing, we have identified intergenic regions between a few genes, which contains unique SNPs for each allele. Particularly, some candidates such as *Tpp1*, which is involved in sugar signaling (Nuccio et al., 2015; Smeekens, 2015), and *MFS*, which is predicted to transport carbohydrates (Niño-González et al., 2019), are being pursued based on the presence of SNPs near these genes. The current hypothesis is that these SNPs in the intergenic region may cause changes in the expression levels of the genes in the vicinity. Quantitative assays measuring the transcript levels, such as qPCR or RNA-Seq between the mutant and the wild-type leaf tissues might help answer this question. Similarly, the generation of CRISPR/Cas9 knockouts of the candidate genes will be instrumental in confirming these candidate genes. For the *tpp1* CRISPR/Cas9 edited plants, we were only able to recover a single mutation and did not observe any phenotype in T0 plants carrying a base pair deletion in the first exon. For the *MFS* gene, regenerated plants following transformation, are currently being screened for *cpd* phenotypes.

The identification of the gene responsible for the *cpd60* phenotype will provide new insights into the genetic regulation of sugar metabolism and allocation in maize.

With this knowledge, we can potentially improve maize productivity and translate our understanding of carbohydrate partitioning to other crop species, such as, sorghum and sugarcane, for genetic improvements to increase food yield and biofuel production.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

All the plants used for the studies were grown at the University of Missouri South Farm Agricultural Experiment Station in Columbia, MO during summer 2018, 2019, or 2020.

Histochemical analyses

Sample preparation and Phloroglucinol and aniline blue staining were performed as described in previous chapter (Chapter 4). The samples were analyzed with a Nikon Eclipse 80i epifluorescence microscope (Nikon Instruments Inc., USA). All images from an experiment were captured using identical microscope and camera settings.

Mapping and whole genome DNA sequencing

Mapping population generation, bulked segregant analysis (BSA), and fine mapping was performed as described previously (Settles et al., 2014; Tran et al., 2019). Sequences of primers used for creating mapping markers can be found in Table 5.2. Whole genome sequencing was performed on DNA isolated from pools of *cpd60, cpd87, cpd87-like, and pr14-cw-334* mutant plants. Leaf samples for DNA extraction were collected into four separate pools consisting of at least 40 mutant plants for each allele. DNA was extracted and purified using the DNeasy Plant Mini Kit (Qiagen, USA), and sequenced at Psomagen, Inc. 150 bp Paired-end sequencing was performed with a

Novaseq6000 (Illumina, USA). All other follow up analyses were performed as described in previous chapter (Chapter 4).

Starch staining

The *cpd60, cpd87, cpd87-like, and pr14-cw332* mutant and wild-type leaves were collected before sunrise. The leaves were cleared in ethanol, rinsed in water, and stained with 1% IKI as previously described (Baker and Braun, 2007).

Measurements of Sugar and Starch levels

Samples from mature source leaves of mutant and wild-type plants were harvested at 5:30 AM from field grown plants and immediately frozen in liquid nitrogen then kept at -80° C until ready to be processed. Soluble sugars and starch were extracted and quantified as previously described using high-performance anion-exchange (HPAE) chromatography (ICS-5000, Thermo-Fisher Scientific, USA) (Leach and Braun, 2016).

Photosynthesis and gas exchange measurements

Photosynthesis rate, stomatal conductance, and transpiration rate were measured as previously described on 8 week old plants (Tran et al., 2019).

Radioactively labeled sucrose transport assays

Radioactively labeled sucrose transport assays were performed as previously described on 4-5-week-old plants (Slewinski et al., 2009; Tran et al., 2017; Tran et al., 2019). ^{14}C-sucrose was purchased from PerkinElmer (USA). 2.54 cm of the tip of mature leaf six was cut quickly and dipped into 2 ml of 1 mM unlabeled sucrose spiked with 200 μCi of ^{14}C-sucrose. The leaf was removed from the sucrose solution and cleaned twice with wet Kimwipes followed by dry Kimwipes after 3 minutes. The leaf was then allowed to translocate the labeled sucrose for one hour. Then, a 25.4 cm segment

(measured from the cut site) was excised from the leaf, taped on filter paper, and exposed. The leaves were first dried at 80°C on a gel drier and then exposed to a phosphor plate for 10 days. The phosphor plates were scanned with a GE Typhoon FLA 9000 scanner (GE Healthcare, USA).

CRISPR-Cas9 gRNA design and cloning

The guide RNA spacer sequences were designed and selected as described (Char et al., 2019). For each gene to be edited, we identified two sites located in exons that could be targeted by two different pairs of gRNA sequences, which we called gRNAa and gRNAb, respectively. For each gRNA target site, two complementary oligonucleotides were annealed with overhangs. The first set of annealed oligos (gRNAa) were cloned into the BtgZI digested pgRNA1 vector as described (Char et al., 2019). After the first set of oligos (gRNAa) were inserted into the plasmid, the second set of annealed oligos (gRNAb) was ligated at the BsaI restriction site. The positive clones carrying both set of annealed oligos (both gRNAa and gRNAb) were identified by digesting with BamHI and BsaI and checking for a specific digestion pattern. Positive cones were sequenced with the forward primer (U6P-F1b) or reverse primer (pENTR4-R) to confirm the accuracy of the first and second spacer sequences, i.e. insertion of gRNAa and gRNAb, respectively (Char et al., 2019). The positive clones were submitted for transformation of B73 embryos and were transformed using biolistic gene gun delivery. The target gene was PCR amplified from T0 plants and Sanger sequencing was performed to identify the specific mutations generated. T0 plants with the edits or with the Cas9 construct were backcrossed to the B73 inbred line. The sequences for gRNA and any primers used are provided in Table S5.1.

References

Baker, R.F., Braun, D.M., 2007. *tie-dyed1* functions non-cell autonomously to control carbohydrate accumulation in maize leaves. Plant Physiol. 144(2), 867-878.

Baker, R.F., Braun, D.M., 2008. *Tie-dyed2* functions with *Tie-dyed1* to promote carbohydrate export from maize leaves. Plant Physiol. 146(3), 1085-1097.

Braun, D.M., Ma, Y., Inada, N., Muszynski, M.G., Baker, R.F., 2006. *tie-dyed1* regulates carbohydrate accumulation in maize leaves. Plant Physiol. 142(4), 1511-1522.

Char, S.N., Li, R., Yang, B., 2019. CRISPR/Cas9 for mutagenesis in rice, in: Kumar, S., Barone, P., Smith, M. (Eds.), Transgenic Plants: Methods and Protocols. Springer New York, New York, NY, pp. 279-293.

Johnston, R., Wang, M., Sun, Q., Sylvester, A.W., Hake, S., Scanlon, M.J., 2014. Transcriptomic analyses indicate that maize ligule development recapitulates gene expression patterns that occur during lateral organ initiation. Plant Cell 26(12), 4718-4732.

Julius, B.T., McCubbin, T.J., Mertz, R.A., Baert, N., Knoblauch, J., Grant, D.G., Conner, K., Bihmidine, S., Chomet, P., Wagner, R., Woessner, J., Grote, K., Peevers, J., Slewinski, T.L., McCann, M.C., Carpita, N.C., Knoblauch, M., Braun, D.M., 2021. Maize *Brittle Stalk2-Like3*, encoding a COBRA protein, functions in cell wall formation and carbohydrate partitioning. Plant Cell 33(10), 3348-3366.

Julius, B.T., Slewinski, T.L., Baker, R.F., Tzin, V., Zhou, S., Bihmidine, S., Jander, G., Braun, D.M., 2018. Maize *Carbohydrate partitioning defective1* impacts carbohydrate distribution, callose accumulation, and phloem function. J. Exp. Bot. 69(16), 3917-3931.

Leach, K.A., Braun, D.M., 2016. Soluble sugar and starch extraction and quantification from maize (*Zea mays*) leaves. Current Protocols in Plant Biology 1(1), 139-161.

Niño-González, M., Novo-Uzal, E., Richardson, D.N., Barros, P.M., Duque, P., 2019. More transporters, more substrates: The Arabidopsis Major Facilitator Superfamily revisited. Mol. Plant 12(9), 1182-1202.

Nuccio, M.L., Wu, J., Mowers, R., Zhou, H.-P., Meghji, M., Primavesi, L.F., Paul, M.J., Chen, X., Gao, Y., Haque, E., Basu, S.S., Lagrimini, L.M., 2015. Expression of trehalose-6-phosphate phosphatase in maize ears improves yield in well-watered and drought conditions. Nat. Biotechnol. 33(8), 862-869.

Rolland, F., Baena-Gonzalez, E., Sheen, J., 2006. Sugar Sensing and Signaling in plants: Conserved and novel mechanisms. Annu. Rev. Plant Biol. 57(1), 675-709.

Russin, W.A., Evert, R.F., Vanderveer, P.J., Sharkey, T.D., Briggs, S.P., 1996. Modification of a specific class of plasmodesmata and loss of sucrose export ability in the *sucrose export defective1* maize mutant. Plant Cell 8(4), 645-658.

Settles, A.M., Bagadion, A.M., Bai, F., Zhang, J., Barron, B., Leach, K., Mudunkothge, J.S., Hoffner, C., Bihmidine, S., Finefield, E., Hibbard, J., Dieter, E., Malidelis, I.A., Gustin, J.L., Karoblyte, V., Tseung, C.-W., Braun, D.M., 2014. Efficient molecular marker design using the MaizeGDB Mo17 SNPs and indels track. G3 Genes|Genomes|Genetics 4(6), 1143-1145.

Slewinski, T.L., Braun, D.M., 2010. The *Psychedelic* genes of maize redundantly promote carbohydrate export from leaves. Genetics 185(1), 221-232.

Slewinski, T.L., Meeley, R., Braun, D.M., 2009. *Sucrose transporter1* functions in phloem loading in maize leaves. J. Exp. Bot. 60(3), 881-892.

Smeekens, S., 2015. From leaf to kernel: Trehalose-6-phosphate signaling moves carbon in the field. Plant Physiol. 169(2), 912-913.

Stelpflug, S.C., Sekhon, R.S., Vaillancourt, B., Hirsch, C.N., Buell, C.R., de Leon, N., Kaeppler, S.M., 2016. An expanded maize gene expression atlas based on RNA Sequencing and its use to explore root development. The Plant Genome 9(1), plantgenome2015.2004.0025.

Tran, T.M., Hampton, C.S., Brossard, T.W., Harmata, M., Robertson, J.D., Jurisson, S.S., Braun, D.M., 2017. In vivo transport of three radioactive [18F]-fluorinated deoxysucrose analogs by the maize sucrose transporter ZmSUT1. Plant Physiol. Biochem. 115, 1-11.

Tran, T.M., McCubbin, T.J., Bihmidine, S., Julius, B.T., Baker, R.F., Schauflinger, M., Weil, C., Springer, N., Chomet, P., Wagner, R., Woessner, J., Grote, K., Peevers, J., Slewinski, T.L., Braun, D.M., 2019. Maize *carbohydrate partitioning defective33* encodes an MCTP protein and functions in sucrose export from leaves. Mol. Plant 12(9), 1278-1293.

Walley, J.W., Sartor, R.C., Shen, Z., Schmitz, R.J., Wu, K.J., Urich, M.A., Nery, J.R., Smith, L.G., Schnable, J.C., Ecker, J.R., Briggs, S.P., 2016. Integration of omic networks in a developmental atlas of maize. Science 353(6301), 814-818.

TABLES

Table 5.1

Candidate genes for *cpd60* and its alleles in the mapped region along with the transcript expression levels based on RNA-Seq (Adapted from Walley et. al, 2016). Blue denotes low expression whereas red denotes high expression levels.

gene_name	Annotation	chromosome	start	Endosperm 12 DAP	Pericarp/Aleurone 27 DAP	Endosperm Crown 27 DAP	Leaf Zone 1 (Symmetrical)	Leaf Zone 3(Growth)	Embryos 20 DAP	Embryo 38 DAP	Mature Leaf 8	Ear Primordium 6-8 mm	Ear Primordium 2-4 mm	Root - Meristematic Zone 5 days	Root - Elongation Zone 5 days	Root - Cortex 5 days	Primary Roots 5 days	Secondary Root 7-8 days	Mature Pollen	Silks	Mature female spikelet	Internode 6 and 7	Internode 7 and 8	Vegetative Meristem 16-19 days
Zm00001d032292	Zf-FLZ_dom	1	220640969	2.7	12.4	0.7	4.3	6	32	46.1	53.5	2.9	4.1	10.2	27.8	126.1	38.4	35.1	0.2	36.3	18.9	33.3	23.3	7.8
Zm00001d032293	WD40_repeat	1	220711403	19.9	19.2	18	30.1	24	31.8	20.1	22.1	52.1	44.3	6.9	14	26.3	18.5	17.5	81.5	23.3	31.1	25.2	24	27.4
Zm00001d032295	AP2/ERF_dom	1	220776886	18.5	27.4	14.4	6.8	9.6	80.7	70.3	29.5	13.4	13.8	19.5	8.4	54.6	20.1	19.8	0.1	63.8	28.5	7	10	17.7
Zm00001d032297	UDP_glucos_trans	1	220858595	0.3	2.8	0.1	5	10.2	4.7	3.4	4.2	3.3	3	0.6	2.8	1.9	1.5	1.2	0.1	3.1	1.6	6.6	3.3	1.9
Zm00001d032298	Trehalose_PPase	1	220971335	2	0.7	6.4	0.7	1.8	10.6	10.2	14.6	6.3	5.2	3.8	4.4	0.4	1.7	4.7	0	2.4	7.1	8.3	3.3	3.4
Zm00001d032299	programmed cell death	1	221079602	0	0.1	0	0	0.1	0.1	0.1	0.1	0.1	0	0.1	0	0	0	0.1	0	0	0	0	0	0
Zm00001d032300	MFS	1	221083224	11.5	19.4	7	5.8	5.5	5.2	7.2	12	4.5	4.6	9.8	13	8.6	12.7	12.6	0.1	8.6	15.3	12	11.8	19.7
Zm00001d032301	SAM-dependent MTase	1	221117789	6.5	2.7	2.5	8.5	4	11.2	10.9	29.6	7.7	6.7	1.4	0.4	0.6	0.5	0.4	0	1	0.8	1.4	1.1	4.5

Table 5.2

List of selected primers used for mapping and genotyping *cpd60* and *cpd87* mutants.

Primer Name	Sequence (5'-3')	Purpose
INS.14047-F1	CATTAGTCGGTTCACCCATACC	Mapping and Genotyping
INS.14047-R1	ACCATTGAAATCGCTGTGAGTA	Mapping and Genotyping
DEL.17222-F1	TCTCAATTGGTGCCCGTTC	Mapping
DEL.17222-R1	TCTACACTGTACAGCCCTAACA	Mapping
DEL.17223-F1	AGCCAGATGCCATAGTTGATAG	Mapping
DEL.17223-R1	TCAGTCTTAACTCACCAAGGTTC	Mapping
DEL.17230-F1	GGGTCGAGTTATATGGCTAGTTC	Mapping and Genotyping
DEL.17230-R1	GGGTTTGCGTTCAAAGTGATAA	Mapping and Genotyping

Table S5. 1

List of spacer sequences oligonucleotides and sequencing and PCR primers to confirm Cas9 edits.

gRNA spacer or Primer	Sequence (5'-3')
MFS-gRNAaF	TGTTGCGATCTCGCGGGGCCTCAA
MFS-gRNAaR	AAACTTGAGGCCCCGCGAGATCGC
MFS-gRNAbF	GTGTGATGACCACACCCGTGGCG
MFS-gRNAbR	AAACCGCCACGGGTGTGGTCATC
tpp1-gRNAaF	TGTTGAGTTGGGCGGTACCGCCA
tpp1-gRNAaR	AAACTGGCGGTACCGCCCAACTC
tpp1-gRNAbF	GTGTGCTGGCCATGTTCGACCAGC
tpp1-gRNAbR	AAACGCTGGTCGAACATGGCCAGC
U6P-F1b	CGTTGAGGGGAGACAGGTTTAG
pENTR4-R	TGGGTCTAGATATCTCGAGTG
U6T-R	CTGCAGAATTGCCCTTCGAAG
cas9-pBUE-F	GAGGTACACAAGCACTAAGG
M13 Reverse	CAGGAAACAGCTATGAC

FIGURES

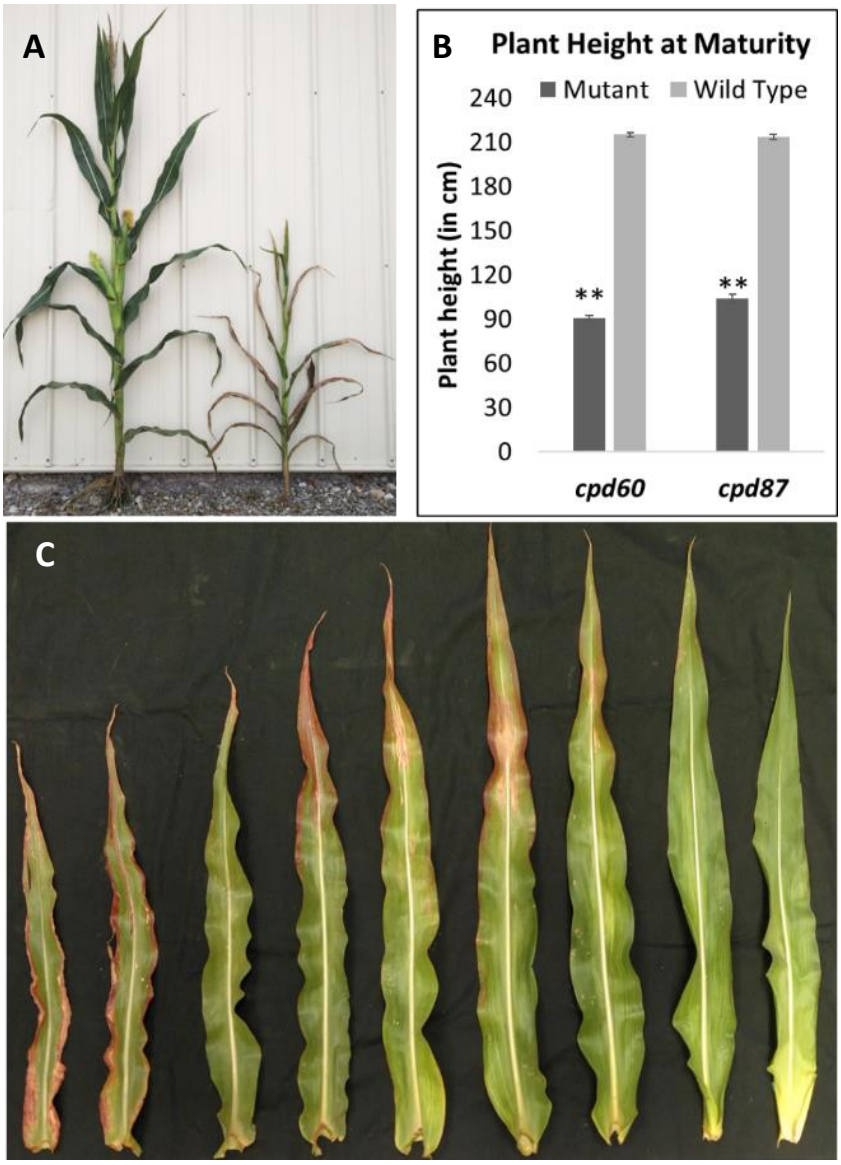


Figure 5.1

cpd60 and *cpd87* mutants are dwarfed and display chlorosis, reduced fertility, and anthocyanin accumulation in their leaves. (A) Field grown wild-type and *cpd60* mutant plants at maturity. (B) Height measurements of *cpd60* and *cpd87* mutants and their wild-type siblings at maturity. (C) Leaves collected from *cpd60* mutant ranging from mature

(left) to youngest (right). **denotes significant difference based on a t-test at $\alpha = 0.01$, $n \geq 28$

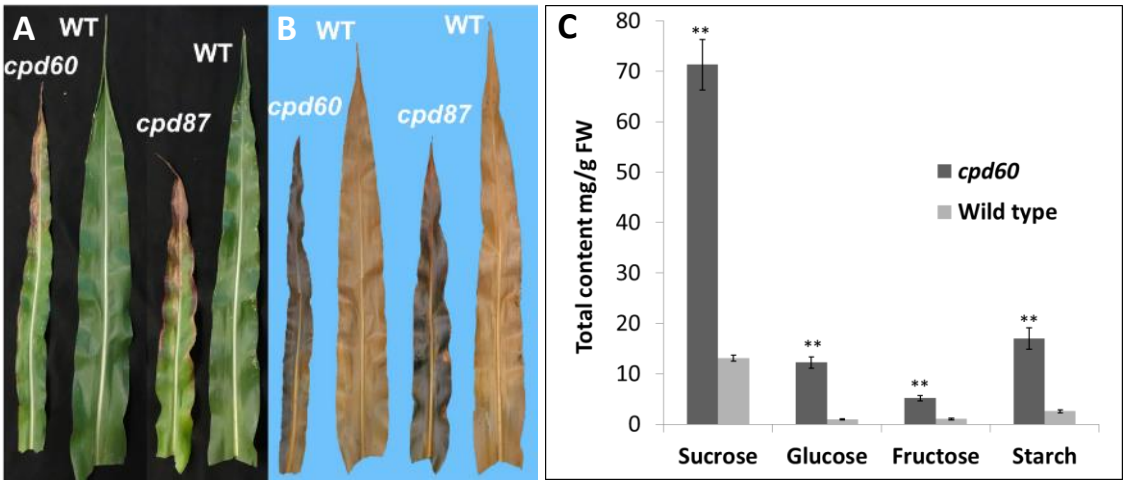
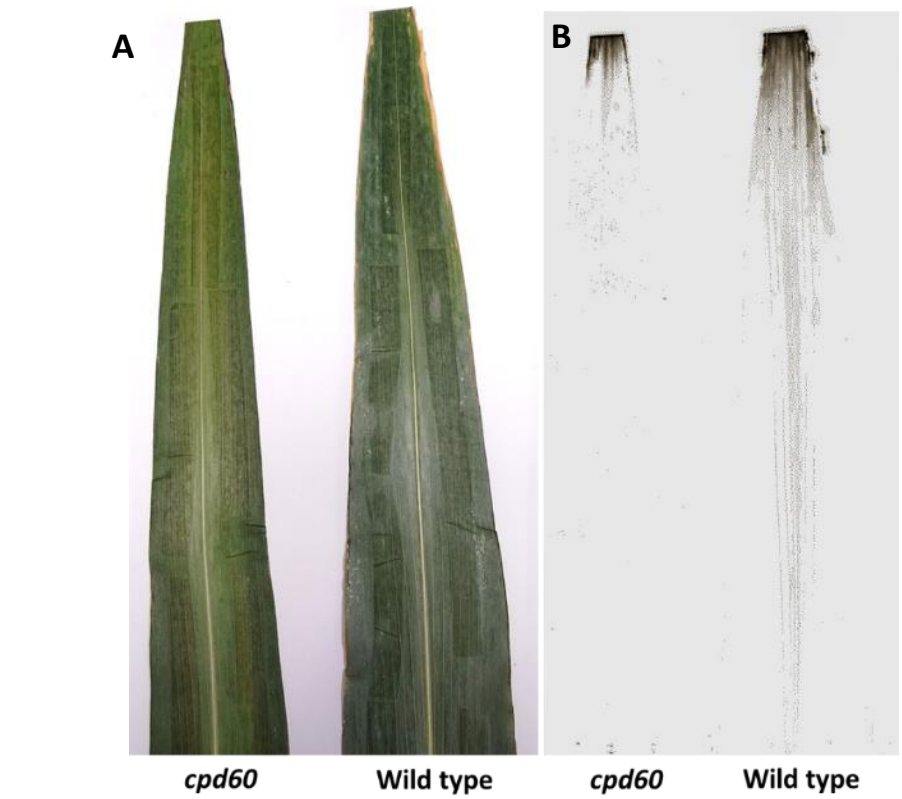


Figure 5.2

cpd60 and *cpd87* mutant leaves hyperaccumulate soluble sugars and starch. *cpd60* and *cpd87* leaves and their wild-type siblings collected at the EON. (A) Before starch staining. (B) After starch staining. (C) Quantification of soluble sugars and starch in *cpd60* mutant and wild-type leaf sections collected at EON is presented along with error bars representing standard error. ** denotes significant difference based on a t-test at $\alpha = 0.01$, $n = 6$


A
B
cpd60
Wild type
cpd60
Wild type

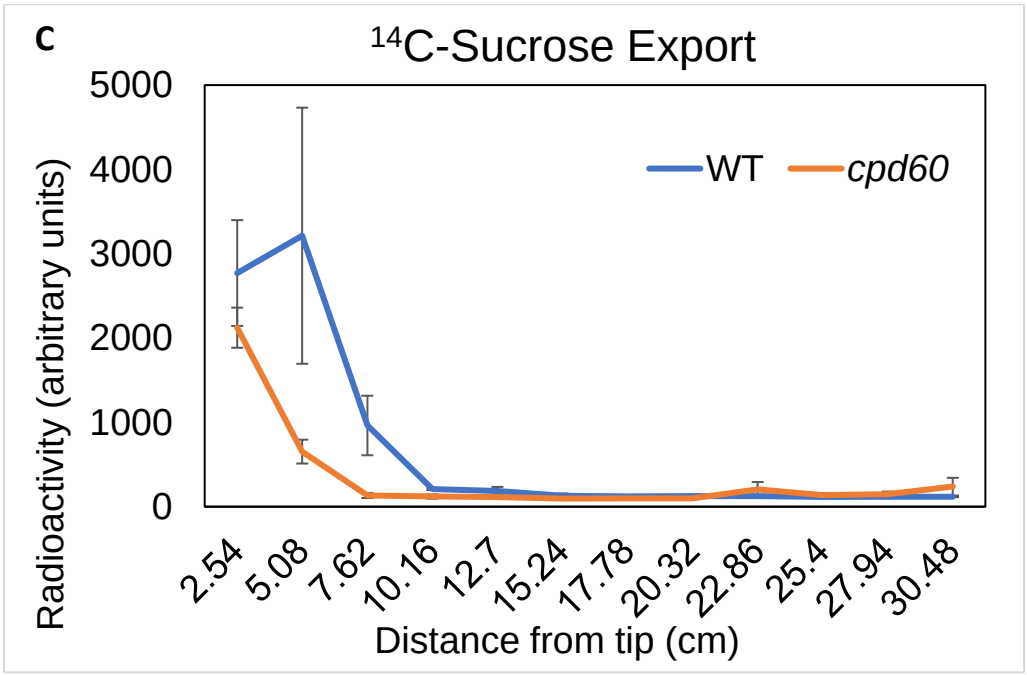

C
14C-Sucrose Export
5000
4000
3000
2000
1000
0
Radioactivity (arbitrary units)
WT
cpd60
2.54
5.08
7.62
10.16
12.7
15.24
17.78
20.32
22.86
25.4
27.94
30.48
Distance from tip (cm)

Figure 5.3

The transport of ^{14}C-sucrose is reduced in mature leaves of *cpd60* mutants. (A) Dried *cpd60* and wild-type (WT) leaves. (B) Autoradiography showing ^{14}C-sucrose distribution in *cpd60* and WT leaves. (C) Quantification of ^{14}C-sucrose intensity along the leaf blade measured at distance (in cm away from the application site (n=3)

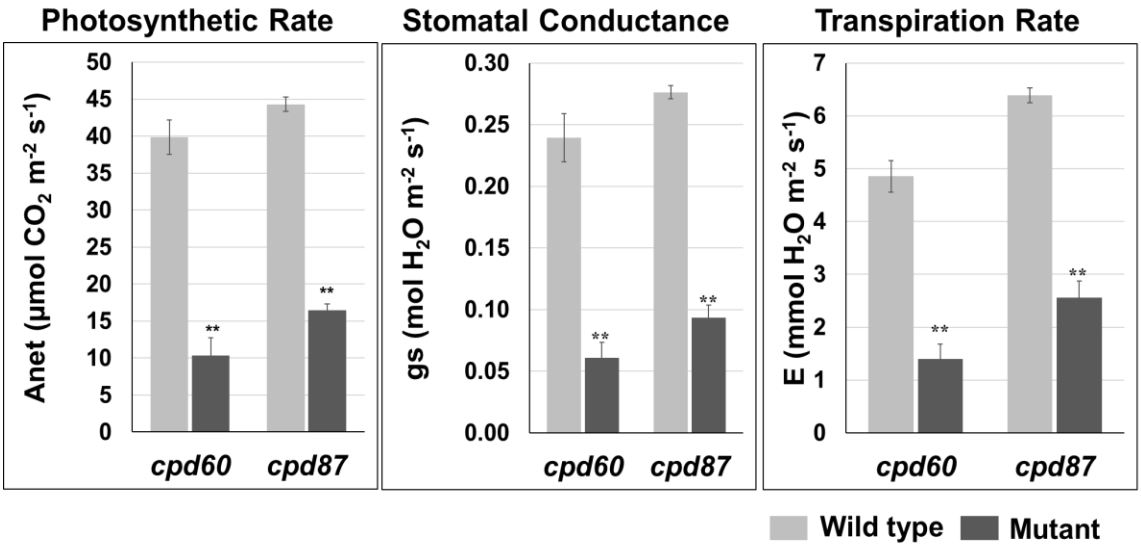


Figure 5.4

cpd60 and *cpd87* mutants have reduced photosynthetic rate, stomatal conductance, and transpiration rates. Average measurements for each parameter are shown for each category. Error bars denote standard error. ** denotes significant difference at $\alpha = 0.01$, $n \geq 4$ based on a t-test between mutant and wild type of each allele.

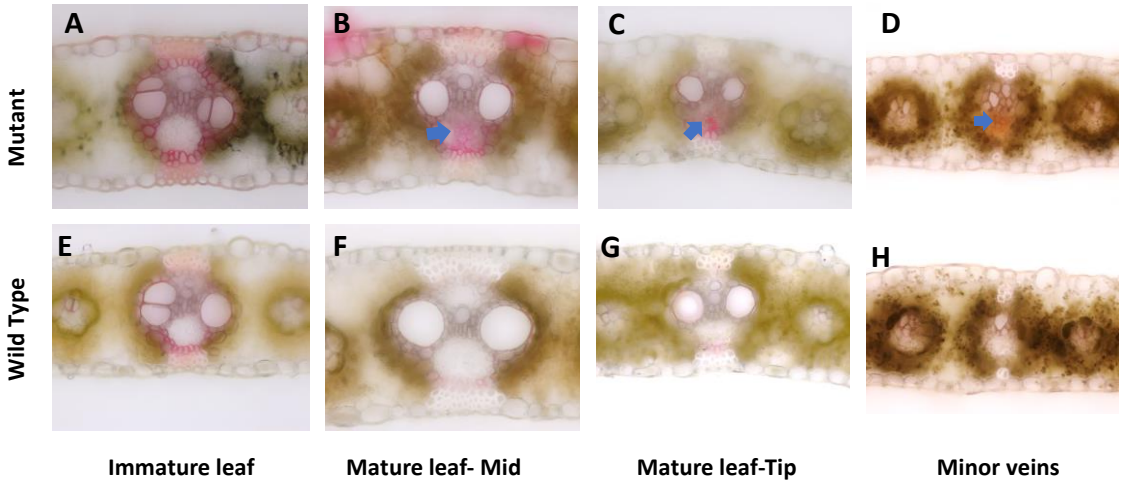


Figure 5.5

cpd87 mutants exhibit ectopic lignin in the phloem cell walls in major and minor veins of mature leaves. Phloroglucinol stained leaf cross sections from immature (A) *cpd87* mutants and (E) wild type, mid regions of a mature leaf from (B) *cpd87* mutants and (F) wild type, tip region of a mature leaf from (C) *cpd87* mutants and (G) wild type, and minor veins of mature leaves of (D) *cpd87* mutants (H) wild type. The blue arrows denote ectopic lignin in the phloem cell walls of the mutants. Lignin depositions were only present in xylem vessels and hypodermal sclerenchyma cells of the wild type.

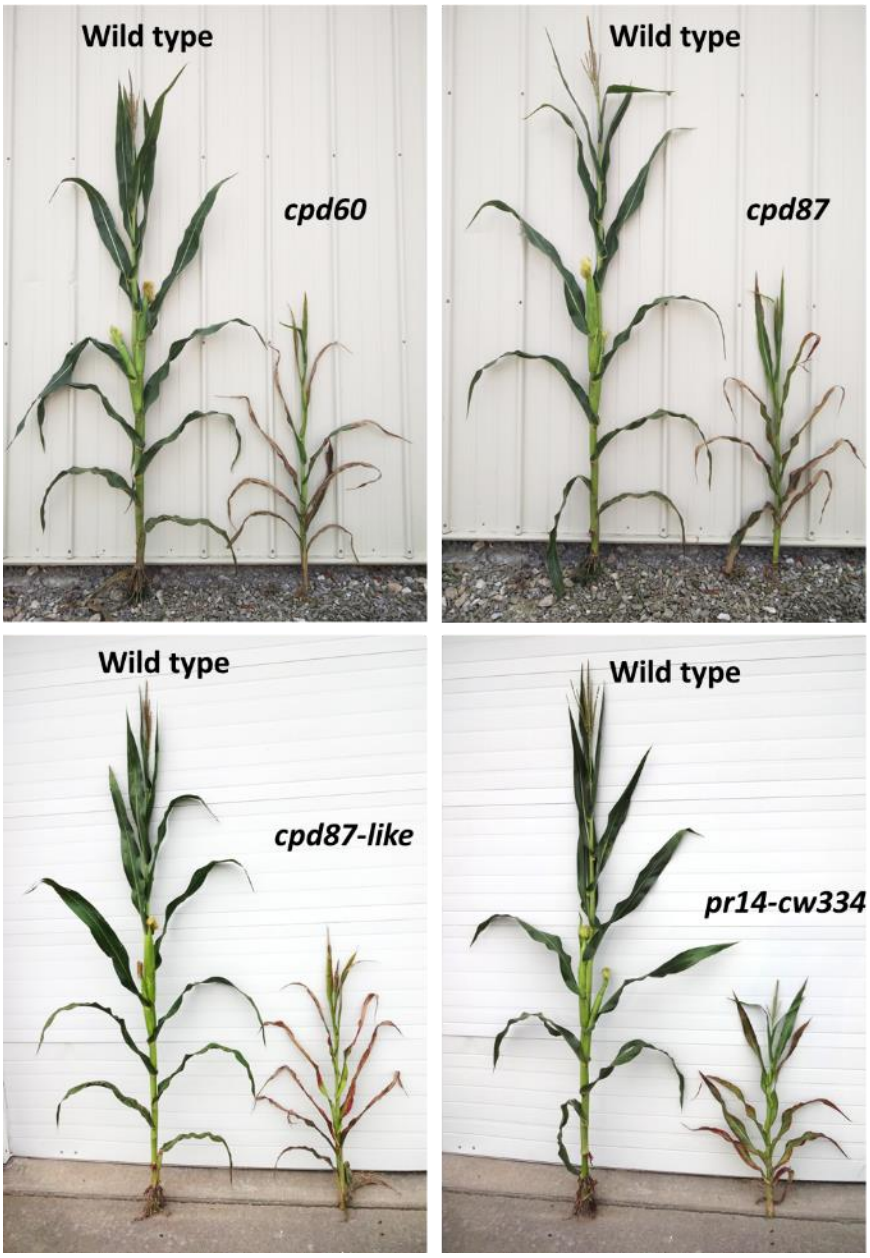


Figure S5. 1

cpd60 and its allelic mutants exhibit decreased plant height and reduced fertility. All mutants and their wild-type siblings are labeled. The wild-type plants have well developed tassels and ears, but the mutants have severely reduced tassels and ears. The mutants display chlorotic leaves and anthocyanin accumulation.

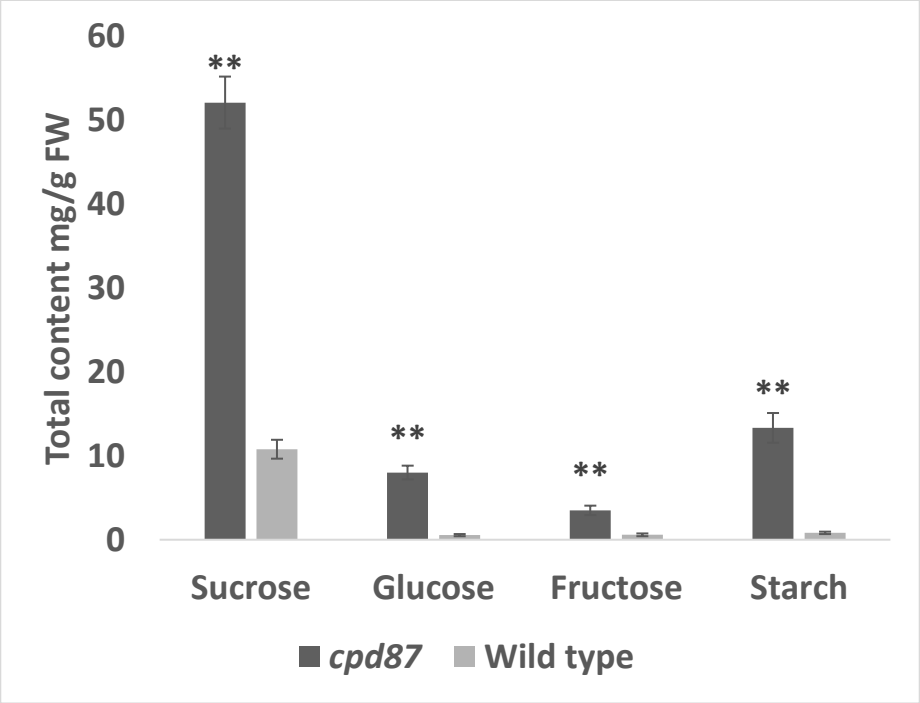


Figure S5. 2

Soluble sugar and starch levels in *cpd87* mutant and wild-type leaves. Quantification of soluble sugars and starch in *cpd87* mutant and wild-type mature leaf tissues collected at EON is presented along with error bars representing standard error. ** denotes significant difference based on t-test at $\alpha = 0.01$, $n = 6$

Chapter 6 : Conclusion and Future Directions

The research presented in the previous chapters have contributed to our understanding of the control of carbohydrate partitioning in maize and sugarcane. With the characterization and cloning of *cpd13/35* mutants, I have uncovered the function of a novel gene that has previously not been studied in relation to carbohydrate partitioning. Similarly, with the characterization of *cpd60/87* mutants, I have uncovered eight new candidate genes that have not yet been studied in maize, at least one of which is involved in carbohydrate partitioning. Furthermore, I annotated Sucrose Transporters (SUTs) and Tonoplast Sugar Transporters (TSTs) in the recently sequenced *S. spontaneum* and *S. officinarum* genomes and compared their transcript levels expression to elucidate which specific *SUTs* and *TSTs* might contribute to the ability of modern sugarcane cultivars to store high amounts of sugars in their stems. This work has established a firm foundation for understanding the genes important for sucrose transport and accumulation in modern sugarcane varieties.

To date there has been no report of a protein with DnaJ-like and thioredoxin-like domains influencing carbohydrate partitioning in any plant species, and the *cpd13* and *cpd35* mutations provide tools to study the mechanism in which a mutated DnaJ-like and thioredoxin-like domain containing protein results in decreased carbohydrate transport. I also uncovered the effect of high temperature on the mutant phenotype, which presents an interesting avenue of exploring abiotic stress factors and their effect on carbohydrate partitioning. In the following section, I outline the directions upon which my research can be built to advance our understanding of the genetic control of carbohydrate partitioning in maize, sugarcane, and other crop species.

Future studies on *carbohydrate partitioning defective13/35* mutants

The identification of the *Carbohydrate partitioning defective13* gene has opened up a new avenue for exploring the function of DnaJ-like and thioredoxin-like proteins in carbohydrate partitioning. DnaJ-like proteins are known to be involved in de novo protein folding (Kampinga and Craig, 2010) and have been demonstrated to act in a hsp70-indpendent manner in plants (Pulido and Leister, 2018), while the thioredoxin-like proteins are known to change the redox state of thiol groups and change the structure of proteins. We hypothesize that in the *cpd13* mutant, the truncated protein encoded cannot perform proper protein folding of its client proteins, which could lead to mistargeting of the protein or aggregation in the ER. These client proteins could be proteins that are directly involved in transporting sugars or proteins involved in plasmodesmata formation or permeability, which can affect the symplasmic transport of sugars (Baker et al., 2016; Tran et al., 2019; Wu et al., 2018; Zhang et al., 2017).

To test my protein aggregation hypothesis, I propose to use the Proteostat Aggresome Detection (Enzo Life Sciences) kit on *cpd13* and wild-type samples to visualize and quantify the aggregated proteins as I outlined in Chapter 4. Longer-term, future protein-protein interaction studies will aid in the identification of CPD13 interacting proteins and will be crucial in further advancing our knowledge of the regulation of carbohydrate partitioning. For that purpose, CPD13 specific antibodies could be developed and used for pulldown assays. As an alternate approach, yeast two-

hybrid screening could be used to identify possible protein-protein interactions. As a targeted approach, crosses of *cpd13* mutants with the major sugar transporter SUT1-YFP (Baker et al., 2016) and an ER resident protein PDI-YFP (Krishnakumar et al., 2014) have been generated based on the fact that *cpd13* leaves exhibit sugar exudates in the leaves similar to *sut1* mutant (Slewinski et al., 2010) and the fact that CPD13 localizes to the ER. These lines will be analyzed this summer.

To test whether symplasmic movement of sugars is perturbed in *cpd13* mutants, a phloem-mobile tracer, such as carboxyfluorescein diacetate (CFDA), could be used as previously described (Baker et al., 2016; Grignon et al., 1989; Julius et al., 2018; Ma et al., 2009). CFDA is cell permeable and is cleaved to form carboxyfluorescein (CF) once it enters the cell. CF is a fluorescent molecule that is transported through sieve tubes in a similar manner to sucrose and can be used to trace the path of symplasmic transport. The results would elucidate the involvement of CPD13 in symplasmic transport of sugars.

I found that the *cpd13* mutant phenotype is induced by high temperature. It would be valuable to explore the effect of other abiotic stresses and chemicals that induce ER stresses to further understand the gene's function. Based on our knowledge that CPD13 localizes to the ER, various other stress factors such as Tunicamycin, an ER stress inducer, or other stresses such as high light and salinity could be used to explore the response in the mutant and wild-type plants.

Future studies on *carbohydrate partitioning defective60/87* mutants

We have identified several candidate genes with SNPs in the intergenic regions in *cpd60*/*87* mutants and their allelic mutants. CRISPR/Cas9 mutants have been generated

to study two of the candidate genes identified from whole genome sequencing and fine mapping experiments. It is exciting to speculate that mutations in the intergenic regions alter the expression of a MFS gene that would induce the carbohydrate defective phenotype (and potentially identify a new type of sugar transporter); however, the evidence of other genes being the candidate cannot be excluded. To follow up on the candidate genes, qPCR assays could be performed on all the genes in the region to identify if any particular gene transcript is reduced or missing in the mutant. If different genes other than the ones being currently CRISPR edited are identified from the qPCR analysis, they could be characterized by creating additional CRISPR mutants or studying known mutants in the gene. Seed stock resources such as the Uniform Mu. Mu-illumina, BonnMu or Ac/Ds lines available in the MaizeGDB database could also be explored (Marcon et al., 2020; Portwood et al., 2018; Williams-Carrier et al., 2010).

Analysis of sugar transporters in *Saccharum* species

Most modern sugarcane varieties are hybrids of *S. spontaneum*, the wild relative of modern sugarcane, which is low in sugar accumulation but highly disease resistant, and *S. officinarum*, the high sugar accumulating sugarcane (Roach, 1972; Zhang et al., 2018). I analyzed the genomes of these varieties for *SUTs* and *TSTs* and hypothesized their function based on transcriptomic data on Chapter 3. However, to further elucidate the functions of these genes and confirm the in-silico analyses, additional molecular characterizations are necessary. Furthermore, the study of the expansion of gene families in these polyploid species could provide insights into gene evolution in polyploids, which could provide novel angles for future researchers.

References

Baker, R.F., Leach, K.A., Boyer, N.R., Swyers, M.J., Benitez-Alfonso, Y., Skopelitis, T., Luo, A., Sylvester, A., Jackson, D., Braun, D.M., 2016. Sucrose transporter *ZmSut1* expression and localization uncover new insights into sucrose phloem loading. Plant Physiol. 172(3), 1876-1898.

Grignon, N., Touraine, B., Durand, M., 1989. 6(5)Carboxyfluorescein as a tracer of phloem sap translocation. Am. J. Bot. 76(6), 871-877.

Julius, B.T., Slewinski, T.L., Baker, R.F., Tzin, V., Zhou, S., Bihmidine, S., Jander, G., Braun, D.M., 2018. Maize *Carbohydrate partitioning defective1* impacts carbohydrate distribution, callose accumulation, and phloem function. J. Exp. Bot. 69(16), 3917-3931.

Kampinga, H.H., Craig, E.A., 2010. The HSP70 chaperone machinery: J proteins as drivers of functional specificity. Nat. Rev. Mol. Cell Biol. 11(8), 579-592.

Krishnakumar, V., Choi, Y., Beck, E., Wu, Q., Luo, A., Sylvester, A., Jackson, D., Chan, A.P., 2014. A maize database resource that captures tissue-specific and subcellular-localized gene expression, via fluorescent tags and confocal imaging (Maize Cell Genomics Database). Plant and Cell Physiology 56(1), e12-e12.

Ma, Y., Slewinski, T.L., Baker, R.F., Braun, D.M., 2009. *Tie-dyed1* encodes a novel, phloem-expressed transmembrane protein that functions in carbohydrate partitioning. Plant Physiol. 149(1), 181-194.

Marcon, C., Altrogge, L., Win, Y.N., Stöcker, T., Gardiner, J.M., Portwood, J.L., II, Opitz, N., Kortz, A., Baldauf, J.A., Hunter, C.T., McCarty, D.R., Koch, K.E., Schoof, H., Hochholdinger, F., 2020. BonnMu: A sequence-indexed resource of

transposon-induced maize mutations for functional genomics studies Plant Physiol. 184(2), 620-631.

Portwood, J.L., II, Woodhouse, M.R., Cannon, E.K., Gardiner, J.M., Harper, L.C., Schaeffer, M.L., Walsh, J.R., Sen, T.Z., Cho, K.T., Schott, D.A., Braun, B.L., Dietze, M., Dunfee, B., Elsik, C.G., Manchanda, N., Coe, E., Sachs, M., Stinard, P., Tolbert, J., Zimmerman, S., Andorf, C.M., 2018. MaizeGDB 2018: the maize multi-genome genetics and genomics database. Nucleic Acids Res. 47(D1), D1146-D1154.

Pulido, P., Leister, D., 2018. Novel DNAJ-related proteins in *Arabidopsis thaliana*. New Phytol. 217(2), 480-490.

Roach, B., 1972. Nobilisation of sugarcane, Proc Int Soc Sugar Cane Technol. pp. 206-216.

Slewinski, T.L., Garg, A., Johal, G.S., Braun, D.M., 2010. Maize *SUT1* functions in phloem loading. Plant Signal. Behav. 5(6), 687-690.

Tran, T.M., McCubbin, T.J., Bihmidine, S., Julius, B.T., Baker, R.F., Schauflinger, M., Weil, C., Springer, N., Chomet, P., Wagner, R., Woessner, J., Grote, K., Peevers, J., Slewinski, T.L., Braun, D.M., 2019. Maize *carbohydrate partitioning defective33* encodes an MCTP protein and functions in sucrose export from leaves. Mol. Plant 12(9), 1278-1293.

Williams-Carrier, R., Stiffler, N., Belcher, S., Kroeger, T., Stern, D.B., Monde, R.A., Coalter, R., Barkan, A., 2010. Use of Illumina sequencing to identify transposon insertions underlying mutant phenotypes in high-copy Mutator lines of maize. Plant J. 63(1), 167-177.

Wu, S.-W., Kumar, R., Iswanto, A.B.B., Kim, J.-Y., 2018. Callose balancing at plasmodesmata. J. Exp. Bot. 69(22), 5325-5339.

Zhang, J., Zhang, X., Tang, H., Zhang, Q., Hua, X., Ma, X., Zhu, F., Jones, T., Zhu, X., Bowers, J., Wai, C.M., Zheng, C., Shi, Y., Chen, S., Xu, X., Yue, J., Nelson, D.R., Huang, L., Li, Z., Xu, H., Zhou, D., Wang, Y., Hu, W., Lin, J., Deng, Y., Pandey, N., Mancini, M., Zerpa, D., Nguyen, J.K., Wang, L., Yu, L., Xin, Y., Ge, L., Arro, J., Han, J.O., Chakrabarty, S., Pushko, M., Zhang, W., Ma, Y., Ma, P., Lv, M., Chen, F., Zheng, G., Xu, J., Yang, Z., Deng, F., Chen, X., Liao, Z., Zhang, X., Lin, Z., Lin, H., Yan, H., Kuang, Z., Zhong, W., Liang, P., Wang, G., Yuan, Y., Shi, J., Hou, J., Lin, J., Jin, J., Cao, P., Shen, Q., Jiang, Q., Zhou, P., Ma, Y., Zhang, X., Xu, R., Liu, J., Zhou, Y., Jia, H., Ma, Q., Qi, R., Zhang, Z., Fang, J., Fang, H., Song, J., Wang, M., Dong, G., Wang, G., Chen, Z., Ma, T., Liu, H., Dhungana, S.R., Huss, S.E., Yang, X., Sharma, A., Trujillo, J.H., Martinez, M.C., Hudson, M., Riascos, J.J., Schuler, M., Chen, L.-Q., Braun, D.M., Li, L., Yu, Q., Wang, J., Wang, K., Schatz, M.C., Heckerman, D., Van Sluys, M.-A., Souza, G.M., Moore, P.H., Sankoff, D., VanBuren, R., Paterson, A.H., Nagai, C., Ming, R., 2018. Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. Nat. Genet. 50(11), 1565-1573.

Zhang, Z., Ruan, Y.-L., Zhou, N., Wang, F., Guan, X., Fang, L., Shang, X., Guo, W., Zhu, S., Zhang, T., 2017. Suppressing a putative sterol carrier gene reduces plasmodesmal permeability and activates Sucrose Transporter genes during cotton fiber elongation. Plant Cell 29(8), 2027-2046.

Appendix A: Interaction between induced and natural variation at *oil yellow1* delays reproductive maturity in maize

Note: The information in this chapter was published under the title:

Khangura, R.S., Venkata, B.P., Marla, S.R., Mickelbart, M.V., **Dhungana, S.R.**, Braun, D.M., Dilkes, B.P., Johal, G.S., 2020. Interaction Between Induced and Natural Variation at *oil yellow1* Delays Reproductive Maturity in Maize. G3 Genes|Genomes|Genetics 10(2), 797-810.

https://doi.org/10.1534/g3.119.400838

Contributions: S.R.D. quantified and analyzed the soluble sugar and starch levels in leaves presented in Figure A.7.

ABSTRACT

We previously demonstrated that maize (*Zea mays*) locus *very oil yellow1 (vey1)* encodes a putative cis-regulatory expression polymorphism at the magnesium chelatase subunit I gene (aka *oil yellow1*) that strongly modifies the chlorophyll content of the semi-dominant *Oy1-N1989* mutants. The *vey1* allele of Mo17 inbred line reduces chlorophyll content in the mutants leading to reduced photosynthetic output. *Oy1-N1989* mutants in B73 reached reproductive maturity four days later than wild-type siblings. Enhancement of *Oy1-N1989* by the Mo17 allele at the *vey1* QTL delayed maturity further, resulting in detection of a flowering time QTL in two bi-parental mapping populations crossed to *Oy1-N1989*. The near isogenic lines of B73 harboring the *vey1* allele from Mo17 delayed flowering of *Oy1-N1989* mutants by twelve days. Just as previously observed for chlorophyll content, *vey1* had no effect on reproductive maturity in the absence of the *Oy1-N1989* allele. Loss of chlorophyll biosynthesis in *Oy1-N1989* mutants and enhancement by *vey1* reduced CO_2 assimilation. We attempted to separate the effects of photosynthesis on the induction of flowering from a possible impact of chlorophyll metabolites and retrograde signaling by manually reducing leaf area. Removal of leaves, independent of the *Oy1-N1989* mutant, delayed flowering but surprisingly reduced chlorophyll contents of emerging leaves. Thus, defoliation did not completely separate the identity of the signal(s) that regulates flowering time from changes in chlorophyll content in the foliage. These findings illustrate the necessity to explore the linkage between metabolism and the mechanisms that connect it to flowering time regulation.

KEYWORDS flowering time; photosynthesis; epistasis, setaria; sorghum

INTRODUCTION

The onset of flowering in angiosperms has been a key focus for plant biologists working on ornamental, horticultural, and other crop species (Lang 1952; Zeevaart 1962; Searle 1965). The onset of reproductive development in angiosperms is affected by a change in meristem identity. The vegetative to floral transition of meristems commits plant development to production of floral organs and sexual reproduction. The integration of signals to correctly time this transition is key to plant fitness. Unsurprisingly, endogenous and environmental cues regulate flowering time (Amasino and Michaels 2010; Cho *et al.* 2017; Minow *et al.* 2018). A critical environmental cue is the duration of the light period, or photoperiod. The photoperiodic responses of plants influence the vegetative to floral transition and the mechanisms of this response have been a focus of intensive research for over a century (Klebs, 1918). Multiple non-photoperiodic cues as well as endogenous signals, sometimes called the autonomous pathway, are also critical to floral transition. Endogenous signals, including hormones and the carbohydrate status of the plant, can also play a critical role in the regulation of flowering time (Corbesier *et al.* 1998; Moghaddam and Ende 2013). But it can be difficult to fully separate endogenous and environmental influences as some environmental factors, such as light quality, alter hormone biosynthesis (Lang 1957; Evans and Poethig 1995; Mutasa-Göttgens and Hedden 2009), and light powers photosynthesis and thereby carbohydrate status (Chen *et al.* 2004). These stimuli converge through the same floral integrators (named FLOWERING LOCUS T (FT) and FLOWERING LOCUS D (FD) in *Arabidopsis thaliana)* for which orthologs

have been identified in many flowering plants (Abe *et al.* 2005; Wigge *et al.* 2005; Corbesier *et al.* 2007; Meng *et al.* 2011; Zhang *et al.* 2016). Accumulation of FT and FD gene products trigger the vegetative shoot apical meristems to acquire the competency to become inflorescence meristems and produce flowers in part via the activation of MADS-box transcription factors that control meristem identity through APETALA1 (Abe *et al.* 2005; Wigge *et al.* 2005). In Arabidopsis, FT is regulated by CONSTANS (CO) in response to both circadian regulation and photoperiodic responses, and CO regulates the MADS-box transcription factor SUPPRESSOR OF CONSTANS1 (SOC1) through FT (Samach *et al.* 2000; Yoo *et al.* 2005).

Maize was domesticated from teosinte (*Zea mays* ssp. *parviglumis* or spp. *mexicana*) in Central America (Doebley *et al.* 1997; Wang *et al.* 1999). Strong selection on time to reproductive maturity contributed to the adaptation of maize to different latitudes (Salvi *et al.* 2007; Huang *et al.* 2017; Swarts *et al.* 2017). Flowering of teosinte is promoted by short-day conditions. In contrast, temperate maize germplasm is relatively day-neutral and flowering is primarily under the control of the autonomous pathway (Coles *et al.* 2010). Mutant studies have identified loci critical to flowering in maize including: *indeterminate1* (*id1*; Colasanti et al. 1998); *early phase change* (*epc*; Vega et al. 2002); *delayed flowering1* (*dlf1*; Muszynski et al. 2006); the cis-element polymorphism *vegetative transitioning1* (*vgt1*; Salvi et al. 2007) that regulates a downstream APETALA2-like transcription factor *zmrap2.7*; *zea mays mads4* (*zmm4*; Danilevskaya et al. 2008); *zmcct10* (Hung *et al.* 2012); *zea mays centroradiales8* (*zcn8*; Meng et al. 2011); *zea mays mads1* (*zmmads1*; Alter et al. 2016); and *zea mays mads69* (Liang *et al.* 2019). Many of these loci encode the maize orthologs of genes identified as

regulators of flowering in Arabidopsis. For example, *dlf1* and *zcn8* encode homologs of the Arabidopsis flowering time determinants FD and FT, respectively. *id1* encodes a zinc-finger transcription factor acting upstream of both DLF1 and ZCN8 (Kozaki *et al.* 2004; Muszynski *et al.* 2006; Meng *et al.* 2011). *zmm4* is an activator of flowering that is part of a conserved syntenic pair of MADS box genes in the grasses, with *zmm24* as the neighboring gene, and encodes one of two maize paralogs of the wheat flowering time and vernalization response locus VRN1 (Danilevskaya *et al.* 2008). *zmm4* acts downstream of *dlf1* and *id1* in the control of flowering time in maize. *zmmads1* is a functional homolog of the Arabidopsis flowering time and circadian rhythm regulator *soc1* (Alter *et al.* 2016). Several QTL studies have used the convenient phenotype of days to reproductive maturity as a proxy for flowering time and identified alleles controlling this trait in maize (Buckler *et al.* 2009; Coles *et al.* 2010; Steinhoff *et al.* 2012; Bouchet *et al.* 2017). While this trait is convenient it is determined by both the days to floral transition of the meristem and the growth rate of the stem and emergence and maturation of floral structures. Nevertheless, many natural variants controlling days to reproductive maturity in maize map to *bona fide* flowering time loci identified by mutant studies including alleles of *zmmads69*, *zmcct10*, *zcn8*, *dlf1*, and *vgt1* (Muszynski *et al.* 2006; Salvi *et al.* 2007; Meng *et al.* 2011; Hung *et al.* 2012; Guo *et al.* 2018; Liang *et al.* 2019).

One important endogenous signal that contributes to flowering time is the carbohydrate allocation status (Ohto *et al.* 2001; Seo *et al.* 2011; Eveland and Jackson 2012; Wahl *et al.* 2013; Cho *et al.* 2018). In maize, mutants that are compromised in either sugar export from source tissues or loading sucrose into the phloem flower later

than their congenic wild-type siblings (Braun *et al.* 2006; Baker and Braun 2008; Ma *et al.* 2008; Slewinski *et al.* 2009; Slewinski and Braun 2010). This is not limited to maize, as starch-deficient Arabidopsis mutants exhibit delayed flowering (Corbesier *et al.* 1998). Trehalose-6-phosphate (T6P) has been implicated as a reporter of the energy status and Arabidopsis mutants effected in this metabolite also exhibit altered flowering time (Paul 2008; Wahl et al. 2013; Seo et al. 2011). T6P and sucrose are positively correlated, and low levels of T6P results in delayed flowering in Arabidopsis (Wahl *et al.* 2013). Remarkably, the carbohydrate status and transcriptional regulatory genes controlling flowering time may be directly linked in maize. The *id1* flowering time mutants alter carbohydrate partitioning in leaves and accumulate more sucrose and starch (Coneva et al. 2012). As a result, ID1 has been proposed to act as a carbohydrate status sensor that influences flowering time in maize (Coneva, Zhu, and Colasanti 2007; Minow et al. 2018). Remarkably, the promoter of the T6P biosynthetic gene *trehalose 6-phosphate synthase1* (*tps1*) is a predicted target of the ID1 DNA-binding protein and low levels of T6P were observed in *id1* mutants (Minow *et al.* 2018).

If sugars are critical for floral transitioning in plants, then manipulation of photosynthesis should alter flowering. Magnesium chelatase (MgChl) is a hetero-oligomeric enzyme complex comprised of subunits I, D, and H. This enzyme catalyzes the first committed step of chlorophyll biosynthesis by conversion of protoporphyrin IX (PPIX) into magnesium-PPIX (Walker and Weinstein 1991; Gibson et al. 1995). The I subunit of MgChl is encoded by *oil yellow1* (*oy1*) in maize (Sawers *et al.* 2006). The OY1-N1989 mutant protein carries a L176F amino acid substitution that results in the protein acting as a competitive inhibitor of MgChl complex function, and decouples

ATPase and Mg^{2+} chelatase activity (Hansson *et al.* 1999, 2002; Sawers *et al.* 2006). As a result, homozygous *Oy1-N1989* mutants are seedling lethal with no chlorophyll accumulation but are viable to reproductive maturity in heterozygous condition (Sawers *et al.* 2006, Khangura *et al.* 2019).

We previously identified a cis-acting expression polymorphism at the *oy1* locus associated with a QTL called *very oil yellow1* (*vey1*) that modifies the chlorophyll content of semi-dominant *Oy1-N1989* mutants (Khangura *et al.* 2019). The *vey1* QTL was proposed to modulate the chlorophyll content of heterozygous *Oy1-N1989/+* mutants via cis-regulatory differences resulting in differential accumulation of the product encoded by the wild-type allele at *oy1*. The Mo17 allele at *vey1* (*vey1*^{Mo17}) was associated with lower abundance of OY1 transcripts, whereas the B73 allele at *vey1* (*vey1*^{B73}) is associated with higher accumulation of OY1. The effect of *vey1* on chlorophyll content is only visible in the presence of *Oy1-N1989*, indicating that this natural variant has a cryptic effect on the function of the MgChl complex.

In this study, we used controlled crosses to segregate *Oy1-N1989* and the modifier alleles at *vey1* (*vey1*^{B73} and *vey1*^{Mo17}) to generate populations of maize with a range of chlorophyll contents. We used this variation in chlorophyll to explore the effects of chlorophyll content on net CO_2 assimilation. These changes in chlorophyll accumulation resulted in changes in net CO_2 assimilation and photosynthetically-fixed carbon accumulation. Remarkably, we noticed that flowering time across material with differing photosynthetic rates and chlorophyll contents was dramatically different. We observed that reduced chlorophyll accumulation was associated with a delay in flowering time. Similar to the cryptic effects of *vey1* on chlorophyll content, the

modifier allele had no effect on flowering time in the absence of the *Oy1-N1989* mutant allele. Chlorophyll content was consistently associated with earlier flowering and partial rescue of chlorophyll accumulation in the *Oy1-N1989* mutant by the *vey1* QTL accelerated flowering in the mutants but had no effect on wild-type siblings. In addition to measurements of net CO_2 assimilation, the premature senescence of maize leaves, induced by sugar accumulation following sink removal, was also reduced by *Oy1-N1989* and further reduced by *vey1* alleles that decrease chlorophyll content and net CO_2 assimilation. The effect of reduced photosynthate accumulation on flowering time was not specific to *Oy1-N1989* as mechanical removal of leaves, to reduce plant leaf area, also delayed flowering time. Thus, all of our results are consistent with an integrative measure of carbon assimilation linking energy status and flowering time in maize.

Materials and Methods

Plant materials

Our previously described stock of the *Oy1-N1989* mutant allele is maintained in the B73 background and is propagated by crossing heterozygous mutants (*Oy1-N1989/+*) to wild-type siblings (Khangura *et al.* 2019). The B73 introgressed stock of *Oy1-N1989* was used for crossing to various mapping populations. A total of 216 intermated B73 × Mo17 population (IBM-RILs), and 251 synthetic 10 doubled haploid lines (Syn10-DH) were crossed as ear-parents with *Oy1- N1989/+*:B73 pollen. The pollen of *Oy1-N1989/+*:B73 plants were also crossed on to the ears of 35 B73-Mo17 near-isogenic lines (BM-NILs) for QTL validation. Tables S1-S3 contain the full list of IBM-RILs, Syn10-DH, and BM-NILs used to develop F_1 hybrid populations.

Field trials

All of the field experiments described in this study were conducted at the Purdue Agronomy Center for Research and Education (ACRE) in West Lafayette, Indiana. Each cross was evaluated as a single plot of 12-16 plants. Each plot derived from crosses with *Oy1-N1989/+*:B73 tester segregated for both mutant and wild-type siblings in approximately 1:1 ratio. Seeds were sown with a tractor-driven seed planter with plot length of 3.84 meters (m), alley length of 0.79 m, and inter-row spacing of 0.79 m. Standard crop management practices at Purdue in terms of fertilizer, pest, and weed control for growing field maize were adopted.

Each experiment was divided into blocks. Progenies from crosses of B73 and Mo17 ears to *Oy1-N1989/+*:B73 pollen were used as parental checks in each block. Parental checks were randomized within each block. The IBM-RILs F_1 populations were evaluated as a single replication in 2013. Syn10-DH F_1 populations were evaluated in 2016 with two full replications planted in randomized complete block design (RCBD). Hybrid progenies of *Oy1-N1989/+*:B73 with BM-NILs were screened in 2016 with five replications planted in RCBD.

Field phenotyping and data collection

The mutant plants in each plot were identified visually as pale plants. The chlorophyll content in mutant and wild-type plants was approximated using a CCM-200 plus (Opti-Sciences Inc., Hudson, NH) as described in Khangura *et al.* 2019. We previously demonstrated correlation of 0.94 between CCM-200 plus values (CCM) and chlorophyll *a*, chlorophyll *b*, and total chlorophyll contents (Khangura *et al.* 2019), and

therefore used CCM as a proxy for chlorophyll content in the materials described here. CCMI refers to CCM measurements at 25-30 days after sowing and CCMII refers to measurements at 45-50 days after sowing. Mutants were tagged between the V5- V7 stages of development. Tagging is necessary as suppression of the *Oy1-N1989* mutant phenotype by *vey1* interferes with visual classification of mutant and wild-type siblings at maturity. Reproductive maturity in each F1 population was recorded separately on the wild-type and mutant plants. The date at which roughly half of the wild-type or mutant plants in a given plot were shedding pollen and had emerging silks from the primary ear was recorded as the date of anthesis and date of silking, respectively, for a given genotype. The dates of anthesis and silking for both wild-type and mutant genotypes were then subtracted from the date of planting to obtain respective wild-type or mutant days to anthesis (DTA), and days to silking (DTS). Further, the difference between DTA and DTS was used to derive the anthesis-silking interval (ASI); ASI = DTA-DTS. Wild-type and mutant trait values are denoted with a prefix WT and MT, respectively. Ratio and differences of these flowering time traits were also calculated as MT/WT and WT-MT, respectively.

A total of 15 F1 populations derived from B73-like NILs × *Oy1-N1989*/+:B73 cross were used to study induced leaf senescence. Seven of these B73-like NILs carried an introgression of *vey1* from Mo17 ($vey1^{Mo17}$), whereas the other eight NILs had the B73 genotype at *vey1* ($vey1^{B73}$).

These NIL populations were planted in the field with at least two replications of each genotype planted in RCBD and two times separated by two weeks. The procedure for this experiment was adapted from Sekhon *et al.* 2012. Briefly, primary and secondary

ears of both wild-type and mutant plants were covered with shoot bags before silk emergence. After 3-4 days of tassel shedding, roughly half of the shoot bags were removed, and these ears were allowed to open pollinate. Staggered rows of B73, in addition to the pollen shed within the row fully pollinated exposed ears. The day of shoot exposure was marked as 0 days after anthesis (DAA). Photographs were taken on the same date using plants from both planting dates to permit display of differences in the effect of DAA on phenotype severity.

Genotypic and gene expression data

The genotypic data and other public datasets on various mapping populations used in this study have been described previously (Khangura *et al.* 2019). Briefly, the public marker dataset for IBM- RILs was obtained from MaizeGDB with 2,178 markers (Sen *et al.* 2010). The markers were reduced to 2,156 after removing duplicate variants, with ~13.3 percent of missing data in the final dataset. Genotypic data consisting of 6611 SNPs for Syn10-DH lines was obtained from Liu et al. 2015. This dataset had no missing genotypes and was used as such for QTL analyses. Genotypes of the B73-Mo17 Near Isogenic Lines (BM-NILs) used for QTL validation were obtained from Eichten *et al.* 2011. Expression data of *oy1* locus in IBM-RILs were obtained from a public repository of the National Science Foundation grant (GEPR: Genomic Analyses of shoot meristem function in maize; NSF DBI-0820610; https://download.maizegdb.org/GeneFunction_and_Expression/ShootApicalMeristem/). This data consists of the expression of maize genes in the tissue derived from the shoot apex of 14 days old IBM-RILs seedlings. The expression data from each gene is normalized to reads per kilobase of transcript per million mapped reads (RPKM).

Allele-specific expression (ASE) assay

Three replications of B73-Mo17 near-isogenic lines (BM-NILs) × *Oy1-N1989/+*:B73 F1 progenies were grown in the field. Mutant siblings derived from four B73-like NILs crossed with *Oy1- N1989/+*:B73 were selected for the ASE experiment. These NILs consisted of two B73-like NILs (b094 and b189) with $vey1^{Mo17}$, and other two B73-like NILs (b135 and b185) with $vey1^{B73}$ genotype. Leaf tissue was harvested from the top fully-expanded leaf at the V3 developmental stage from the mutant siblings of the four B73-like NIL F1 plots. For each biological replicate, tissue was pooled from 4-5 seedlings to make one sample. The samples were stored at -80 ˚C until needed. The procedure of total RNA isolation, cDNA synthesis, and the ASE assay has been described previously in detail (Khangura *et al.* 2019). Briefly, one µg of DNase treated total RNA was used to synthesize cDNA. PCR was conducted using the forward oligonucleotide 5'- TCACCGTCTGCAATGTCGCCGCTC -3' and reverse oligonucleotide 5'- AGTATGCCCCTGTTGGCCTTGGCG -3' under following reaction conditions with 30 cycles of polymerization (94˚C for 30s, 56˚C for 30s, 72 ˚C for 30s and final extension for 2 minutes) to amplify the targeted region of OY1. The primer pair used in this assay flanked the SNP that causes the L176F amino acid substitution in the *Oy1-N1989* mutant allele. PCR products were sequenced on a MiSeq instrument (Illumina, San Diego, CA) at the Purdue Genomics Core Facility. Reads were aligned to a small reference sequence of B73 derived from targeted PCR region using the GATK packages (DePristo *et al.* 2011). Read counts derived from GATK was used to calculate allele-specific expression. Genomic DNA derived from B73 × Mo17 F1 hybrids resulted in roughly 1:1 read counts demonstrating no bias in the assay.

Statistical analyses

Exploratory data analysis was done using JMP 13.0 (SAS Institute Inc. 2016). The pairwise correlations were calculated using the Pearson correlation coefficient. The average values of various traits from line-cross populations, IBM-RILs and Syn10-DH, were subjected to QTL analyses. QTL detection was done using a single interval mapping via the EM algorithm using the function "scanone" in R/qtl, a software package implemented in R (Broman *et al.* 2003; R core Team 2013). Similar results were obtained with composite interval mapping function "cim" in R/qtl (data not shown).

Defoliation assay

The defoliation experiments were performed using maize inbred B73, sorghum (*Sorghum bicolor*) inbred BTx623, and green foxtail (*Setaria viridis*) inbred A10.1. These experiments were conducted in a greenhouse using mogul base high-pressure sodium lamps (1000 Watts) as the supplemental light source for L:D cycle of 16:8 hours, with the temperature set at 28ºC (day-time) and 20ºC (night-time). The maize inbred line B73 was defoliated at V3 leaf stage. All the leaves with a fully visible leaf collar were cut slightly above the ligule. Sorghum plants with three to four fully opened leaves (~20 days after sowing in the greenhouse) were defoliated in a similar way. All fully expanded leaves at ~15 days after planting, including those on tillers, were removed in *Setaria* plants. The time to reproductive maturity of both defoliated and undisturbed controls was recorded on maize as described above. For sorghum and *Setaria*, the date of head emergence, rather than anthesis, on every plant was recorded and deducted from the date of planting to obtain days to heading.

Non-structural carbohydrate (NSC) quantification

The soluble sugars and starch were quantified from the mutant siblings of four B73-NILs × *Oy1- N1989/+*:B73 F_1 population. Two of these NILs carried $vey1^{Mo17}$, while the other two had $vey1^{B73}$ genotype. Plants were grown in the field with three replications using a RCBD. The top fully- expanded leaf at the V3 stage was harvested at 1:00 PM and transferred to liquid nitrogen. Leaf tissues were stored at -80°C until needed. Leaf samples were ground into a fine powder and ~100 mg of the powder was used to extract sucrose, glucose, fructose, and starch using a previously described method (Leach and Braun 2016). The quantification of these NSC was done using a previously described method (Leach *et al.* 2017). Briefly, high-performance anion exchange (HPAE) chromatography (ICS-5000, Thermo-Fisher Scientific) was used to analyze the neutral fraction of the extract. Sugar standards were used to construct a standard curve, and samples were diluted to ensure that the detected values fell within the scope of this curve.

Gas-exchange measurements

Gas-exchange measurements were taken on field-grown plants during the summer of 2017 at the Purdue ACRE farm. The gas-exchange measurements were taken on the third leaf on plants at the V3 stage between 11 AM and 1 PM using a LICOR LI-6400XT open photosynthesis system (LI- COR Inc., Lincoln, NE, USA). The F_1 progenies used for these measurements consisted of four independent B73-NILs × *Oy1-N1989/+*:B73 cross, in which two NILs carried $vey1^{Mo17}$ introgression and the other two carried $vey1^{B73}$ allele. Plants were grown in a RCBD of three replicated blocks. For each genotype, nine plants were measured as three replicates in each of the three blocks.

The following instrument conditions were maintained throughout the measurement period: an artificial light source with an intensity of 1700 µmol photosynthetically active radiation (PAR) m^{-2} s^{-1}, air temperature of ~31 ºC, CO_2 concentration of 400 mL L^{-1}, air flow of 400 µmol s^{-1}, and relative humidity of 50-60%. Leaf temperatures varied from 32-34ºC during the measurements.

Chlorophyll fluorescence measurements were taken on the same leaves used for gas- exchange using a Handy PEA (Hansatech Instruments Ltd., Norfolk, UK). Leaves were dark- adapted for 20-30 min using leaf clips before taking measurements. The saturation pulse rate of 3000 µmol m^{-2} s^{-1} was used to measure the emission of chlorophyll fluorescence. The initial chlorophyll fluorescence yield (F_0), the variable chlorophyll fluorescence yield (F_V), and the maximum chlorophyll fluorescence yield (F_m) were recorded. The maximum photochemical efficiency of PSII in dark-adapted leaves was obtained by calculating the ratio of F_V/F_m.

Data availability

All phenotypic data from the QTL and NIL populations are attached to this manuscript as supplemental tables S1-S12 and available via figshare. All marker data was previously used in Khangura *et al.* 2019 and made available to the public via figshare (https://doi.org/10.25387/G3.7370948). All the seed stocks described in this study are available upon request.

RESULTS

Negative effect of Oy1-N1989 on time to reproductive maturity is exacerbated by Mo17

While preparing the material for our previous study (Khangura *et al.* 2019), we noticed that *Oy1- N1989* exhibited a consistent delay in flowering, as measured by the days to silking and days to pollen anthesis. Heterozygous *Oy1-N1989* mutant plants in the Mo17 × B73 hybrid genetic background flower up to two-weeks later than wild-type siblings (Figure A.1; Table S4). The *Oy1- N1989* mutants also flower later in an isogenic inbred B73 background; however, the delay is only four days. By contrast, wild-type B73 × Mo17 F_1 hybrid plants flower earlier than the wild-type B73 inbred plants. Maize is protandrous and tassels mature earlier than the ear-inflorescence. The effect of *Oy1-N1989* and flowering time was similar for both anthesis and silk emergence. The window of difference in maturity of the tassel and ear inflorescence, measured as anthesis-silking interval (ASI), is used as an indicator of plant stress in maize (Bolanos and Edmeades 1996). The ASI was wider in the wild-type siblings compared to the mutants in Mo17 × B73 hybrid background (Figure A.1G) and not discernably different in the B73 inbred background. Thus, the delay in flowering does not seem to be due to a generic stress effect due to lower chlorophyll contents.

Delayed reproductive maturity of Oy1-N1989 mutants in B73 × Mo17 mapping populations maps to *vey1*

If the effect of genetic background on flowering time in *Oy1-N1989* mutants is due to variation in the accumulation of chlorophyll, we expect that the previously described *vey1* QTL from Mo17 should make this more severe (Khangura *et al.* 2019). To identify the genetic basis of the flowering time variation in *Oy1-N1989* mutants and test the effect of *vey1*, we recorded flowering time in wild-type and mutant F_1 progenies from the crosses between *Oy1-N1989/+*:B73 pollen-parent with IBM-RILs and Syn10-

DH lines. Hereafter, the hybrid populations developed from these crosses will be referred to as the IBM-RILs F_1 and Syn10-DH F_1 populations (Figure A.2).

Pairwise correlations were calculated between previously reported chlorophyll index measures (Khangura *et al.* 2019) and flowering time traits collected from the same plots (Tables S5 and S6). The chlorophyll index was measured at two time points CCMI (25-30 days after sowing) and CCMII (45-50 days after sowing). In the IBM-RILs crosses, wild-type CCMII displayed a weak but significant negative correlation with wild-type DTA and DTS. Similarly, in the Syn10-DH crosses wild-type CCMI displayed a significant weak negative correlation with wild-type DTA and DTS. This indicates that the phenomena observed in our mutants, reduced chlorophyll content associated with a delay in flowering time was true in the wild-type populations as well, but much less obvious. The variation in chlorophyll content in wild-type plants was not predictive of either mutant CCM or flowering time in the mutants (Figure A.3, Table S5, and S6) indicating that the variation in CCM was not under the same control in the mutant and wild-type subpopulations. A dramatic and obvious negative correlation was observed between CCM trait values (CCMI and CCMII) and flowering time in the mutant siblings in both IBM-RILs and Syn10-DH F_1 populations. As was observed in the *Oy1-N1989/+* B73 inbred stock and B73 × Mo17 hybrids *Oy1-N1989/+* mutants, mutants in these test-cross populations also showed a clear increase in mean values for days to anthesis and silking compared to wild-type siblings (Figure A.4). The frequency distribution plot of days to anthesis in mutant siblings of IBM-RILs and Syn10-DH F_1 populations displayed a bimodal distribution, suggesting a single polymorphic locus segregating between B73 and

Mo17 was the basis of flowering time variation in *Oy1-N1989* mutants (Figure A.5). QTL mapping detected a single QTL on chromosome 10 at similar linkage positions for all mutant derived flowering traits (Figure A.5; Tables S7 and S8). This corresponds to the *vey1* locus that we previously described as a major-effect QTL that controls chlorophyll biosynthesis only in the presence of the *Oy1-N1989* allele (Khangura *et al.* 2019). QTL mapping for various direct and derived mutant flowering time traits such as days to flower in the mutant heterozygotes (MT_DTA and MT_DTS), ratio of days to flower derived from mutant and wild- type siblings (Ratio_DTA and Ratio_DTS), and the difference in days to flower between the wild-type and mutant siblings (Diff_DTA and Diff_DTS) all detected *vey1* in both mapping populations and exceeding permutation-estimated significance thresholds (alpha<0.05). The *vey1* QTL explained ~40-48% of the phenotypic variation for these flowering traits in the IBM-RIL crosses, and >65% variation in Syn10-DH F_1 crosses. As expected, the *Oy1-N1989* enhancing $vey1^{Mo17}$ allele was associated with a delay in flowering time in the mutant hybrid siblings (Tables S7 and S8). In addition to *vey1*, an additional QTL controlling both mutant anthesis and silking was detected in the Syn10-DH F_1 population on chromosome 2 which we call *other oil yellow1 flowering time locus1* (*oof1*). This QTL explained ~7-8% variation in flowering time with the Mo17 allele at this locus resulting in a delay of 2-3 days in reproductive maturity of mutant siblings.

Because of the very large effect of *vey1*, and the mild segregation distortion at this locus in both bi-parental populations (Khangura *et al.* 2019), weak QTL might be detected by spurious linkage between chr2 and chr10 markers. To test this, we carried out multiple regressions using top markers at *oof1* and *vey1* as independent variables

and mutant DTA in the Syn10-DH as a dependent variable (data not shown). In this analysis, both *vey1* and *oof1* remained significant factors in the multiple regression and explained 62% and 2% of the variance in flowering time, respectively. Inclusion of a *vey1* × *oof1* interaction term did not improve the fit of the model, did not eliminate the significance of the *oof1* term, and the interaction term was not a significant variable. Moreover, we did not detect additional QTL by including these as covariates in an additional genome-wide scan. Therefore, we propose that *oof1* is a novel QTL contingent upon the *Oy1-N1989* mutation and genetically independent of *vey1*.

The ASI of mutant and wild-type siblings in the test cross populations were not discernably different, just as we observed in the mutant parents. QTL mapping for this trait did not detect any loci controlling ASI in either the mutants or the wild type siblings. In addition, we did not detect any QTL for flowering time in wild-type siblings (Tables S7 and S8). Thus, *Oy1-N1989* was epistatic to both *vey1* and *oof1* QTLs suggesting a role for each locus in controlling either photosynthesis or chlorophyll metabolites in the regulation of flowering time.

We further validated the effect of the *vey1* critical region using a set of NIL that vary at the *vey1* QTL from the recurrent background. F1 progeny of these NIL and *Oy1-N1989*/+ produced matched wild-type and *Oy1-N1989*/+ heterozygous mutant NIL F1 hybrids. The $vey1^{Mo17}$ allele delayed flowering time in *Oy1-N1989/+* mutants crossed to both B73 and Mo17 recurrent backgrounds when compared to NILs carrying $vey1^{B73}$ (Figure A.6, Figure S3, and Table S3). No effect of *vey1* introgression from either parent was visible on the flowering traits of the wild-type siblings. The reproductive maturity of mutant B73-like NILs carrying $vey1^{Mo17}$, and reciprocal

introgression of *vey1*^{B73} into Mo17 background displayed the opposite effect on flowering time in the F1 mutants (Figure A.6). Thus, these results clearly show the single locus effect of *vey1* on flowering time in maize in *Oy1-N1989*-contingent manner in both isogenic inbred (B73) and hybrid (Mo17 × B73) background.

Expression polymorphism in B73-like NILs is consistent with cis-acting regulatory polymorphism at *vey1*

Our previous study looking at the suppression of *Oy1-N1989* mutant phenotype using chlorophyll accumulation identified a cis-eQTL at *oy1* in the IBM-RILs population (Khangura *et al.* 2019). Normalized expression (expressed as RPKM) of OY1 derived from 14 days old shoot apices of IBM-RILs (Li et al., 2013; 2018) were used for this analysis. The top marker, isu085b, used in the detection of this cis-eQTL was also one of the top significant markers for mutant flowering time traits (Figure S1 and Table S9). Regression of OY1 expression and flowering traits collected in IBM-RILs F1 population identified a significant linear relationship between gene expression in wild-type inbred lines and flowering time measurements from mutant F1 siblings (Figure S2 and Table S9). Roughly 21% of the variation in mutant DTA could be explained by OY1 expression in the IBM-RILs shoot apices. OY1 expression did not predict any variation in wild-type DTA.

The allele specific-expression (ASE) assay in our previous study identified bias in expression with the wild-type *oy1* allele from Mo17 displaying lower expression than the wild-type *oy1* allele from B73 (Khangura *et al.* 2019). Our previous ASE work compared mutant plants in two different genetic backgrounds (inbred vs hybrid) which can

complicate the interpretations. To overcome this limitation, an ASE assay was designed to test bias at *oy1* using near-isogenic lines in the B73 background. These B73-like NILs consisted of two independent NILs with *vey1*^{Mo17} introgression, and two independent NILs with B73 genotype at *vey1*. Consistent with the previous ASE results, a significantly greater proportion of expression was derived from the *Oy1-N1989* mutant allele when the wild-type *oy1* allele was contributed by *vey1*^{Mo17} introgression as compared to the isogenic mutant siblings carrying the wild-type *oy1* allele from *vey1*^{B73} introgression (Table S10).

Net CO2 assimilation and sugar metabolism is reduced in *Oy1-N1989* mutants in *vey1*- dependent manner

We measured net CO_2 assimilation, sub-stomatal CO_2, photosystem II fluorescence, and photosynthate accumulation in enhanced and suppressed *Oy1-N1989/+* mutants. A previous study in maize has shown that reduction in chlorophyll levels in the leaves leads to a reduction in photosynthetic rate (Huang *et al.* 2009). A similar reduction in photosynthetic rate should be displayed by *Oy1-N1989* mutants and the Mo17 allele should show a greater reduction in photosynthesis compared to the B73 allele. We tested this using F1 progenies derived from the same four B73-like NILs used for the ASE experiment. The negative effect of *vey1*^{Mo17} introgression on chlorophyll accumulation was visible in the *Oy1-N1989* mutant allele background (Figure A.6). As expected, photosynthetic rate (A) was reduced in mutants as compared to wild-type siblings, and mutants were modified further by the *vey1* genotype (Table A.1). Photosystem II efficiency (Fv/Fm) measurements indicated higher photo-oxidative damage to the photosystem in enhanced mutant plants compared to the

suppressed mutants. Wild-type siblings of all four B73- like NIL F_1 progenies showed no statistically significant difference for chlorophyll and gas- exchange measurements. This indicates that, just as for chlorophyll content, the *vey1* QTL affects net CO_2 assimilation in the presence of the *Oy1-N1989* mutant allele. No differences in stomatal conductance (gs) or transpiration (E) were observed in the mutant plants, however, severe mutant plants showed significantly higher accumulation of intracellular CO_2 (Ci) compared to the suppressed mutants suggesting the failure of the enhanced mutant plants to uptake CO_2 (Table A.1). These differences in photosynthetic rate should result in a decrease of non-structural carbohydrate accumulation in the photosynthetic leaf tissue. The levels of non-structural carbohydrates were determined from leaves of the same mutant plants that were used for gas- exchange measurements. Levels of sucrose, glucose, fructose, and starch were measured in these samples. All of these showed a significant reduction in the enhanced mutants compared to the suppressed mutants in the B73 isogenic background (Figure A.7 and Table S11). Lower levels of sugars and starch in the *Oy1-N1989/+* mutant heterozygotes enhanced by a $vey1^{Mo17}$ allele is consistent with the observation of lower chlorophyll levels and photosynthetic rates in these genotypes compared to the suppressed mutant NILs.

Defoliation of *Zea mays, Sorghum bicolor,* and *Setaria viridis* delays reproductive maturity

The existing literature suggests sugars and carbohydrate metabolism play an important role in regulating flowering time in plants (Ohto *et al.* 2001; Seo *et al.* 2011; Wahl *et al.* 2013). We hypothesized that removal of source tissue should mimic the

sugar starvation observed in *Oy1-N1989* mutants. We used mechanical defoliation to reduce photosynthetic surplus of the plant. The choice of this treatment was intended to deprive plants of leaf area and photosynthate early in development to separate the block in chlorophyll biosynthesis and the loss of photosynthate, which are coupled in the *Oy1-N1989* mutant study. We conducted this experiment using wild-type inbred strains of maize and two other monocot species: *Sorghum bicolor* (sorghum) and *Setaria viridis* (green foxtail). Each species was defoliated at an early vegetative stage when only few leaves had fully expanded. Defoliation delayed flowering in all three species (Figure A.8). Maize, sorghum, and green foxtail displayed a delay in flowering by about 13, 7, and 4 days, respectively, compared to the control plants. Remarkably, one-week post-defoliation, newly-emerged leaves displayed a pale leaf color. Chlorophyll estimation using CCM found a reduction in the leaf greenness in the defoliated treatments compared to the control samples (Figure A.8). The newly-emerged leaves of defoliated maize, sorghum, and green foxtail showed ~35%, ~48%, and ~58% reduction in CCM, respectively, compared to control plants. Subsequent leaves emerging from the defoliated plants displayed normal leaf color suggesting recovery of the plants from defoliation. We propose that early season defoliation results in the removal of source tissue that might be critical for vegetative to floral transition in grasses but a direct effect of chlorophyll accumulation cannot be ruled out.

Leaf senescence is suppressed by *Oy1-N1989* mutants in a *vey1*-dependent manner

Leaf senescence can be induced in maize by sucrose accumulation in the leaves. This can be accomplished genetically by disrupting sucrose transport (Braun *et al.* 2006; Baker and Braun 2008; Slewinski *et al.* 2009) or by preventing the maize ears

from acting as a sink (Allison and Weinmann 1970; Sekhon *et al.* 2012, 2019). Previous studies have shown that maize leaf senescence caused by pollination prevention or ear removal before pollination is genotype- dependent (Ceppi et al. 1987). We tested the effect of sugar accumulation in mutant B73-like NILs on induced leaf senescence by pollination prevention. Given the variation in leaf sugar in *Oy1- N1989/+* (Figure A.7) we expect the mutants exhibit less or later senescence following pollination prevention and modulation of this effect by *vey1*. We observed that 30 days-after-anthesis (DAA), the top leaves of unpollinated wild-type B73 plants showed complete senescence with only a few green patches. Unpollinated *Oy1-N1989/+* F_1 mutant plants crossed to the *vey1*^{B73} NIL background were green and showed only a few patches of anthocyanin accumulation and cell death on the top leaves at 30 DAA (Figure A.9 and Table S12). Consistent with the lower chlorophyll and NSC accumulation, unpollinated *Oy1-N1989/+* mutant F_1 plants crossed to the B73 NIL background containing the *vey1*^{Mo17} introgression did not show any sign of leaf senescence at 30 DAA. By 42 DAA, unpollinated *Oy1-N1989/+* mutants with the *vey1*^{B73} allele and all unpollinated wild-type plants showed leaf senescence. Even at 42 DAA, the enhancement of *Oy1-N1989/+* by the *vey1*^{Mo17} allele resulted in substantially less cell death and anthocyanin accumulation.

DISCUSSION

Our previous work identified a modifier, *vey1*, that affects the chlorophyll accumulation in *Oy1- N1989/+* heterozygotes (Khangura *et al.* 2019). The *vey1* polymorphism(s) are common natural variant(s) linked to the *oy1* locus of maize. We proposed that *vey1* results from cis-acting regulatory polymorphisms that affect the

expression of OY1. The suppression of the mutant, and accumulation of chlorophyll, follows the proportion of wild-type and mutant transcript levels (Table S10 and Khangura *et al.* 2019). In this work, we describe a delay in reproductive maturity in the *Oy1-N1989* mutants and demonstrate that *vey1* encodes a strong modifying QTL altering the flowering time of mutant siblings in all mapping populations (Figures A.5 and A.6; Tables S7 and S8). Just as the detection of *vey1* for CCM was contingent on the *Oy1-N1989* mutant allele in the background, there was no effect of the *vey1* genotype on flowering time in wild-type siblings. We observed the same marker, isu085b, had the strongest effect on chlorophyll content, OY1 transcript abundance in shoot apices (Khangura *et al.* 2019), and variation in flowering time of IBM-RILs × *Oy1-N1989/+* mutant F_1 mutant siblings (Table S9 and Figure S2). Taking these observations together, we propose that the cis-acting eQTL at *oy1* is affecting chlorophyll level in *Oy1-N1989* mutants, and that the alteration in photosynthesis through perturbed chlorophyll metabolism affects flowering time variation in mutant siblings in these populations.

Our previous work on *vey1* has focused on the cryptic nature of the variation, and the interaction between the *vey1* QTL and the *Oy1-N1989* mutant allele. The experiments presented here also suggest chlorophyll content as a heretofore unappreciated correlate of flowering time in wild-type maize plants (Figure A.3; Tables S5 and S6). We observed that CCM values in both test cross populations, were negatively correlated with wild-type days to reproductive maturity. This suggests a role for the determinants of variation in chlorophyll contents in regulation of flowering time in maize perhaps via changes to photosynthetic output or signaling. As no QTL were

detected for CCM in the wild-type siblings in our previous work (Khangura *et al.* 2019), it also suggests that the mechanism responsible for the covariation between CCM and flowering time is independent of *vey1*. This is further strengthened by the absence of an effect of *vey1* on flowering time in the wild-type siblings in our mapping populations and NIL experiments (Table S3 and Figure S3). Further experiments are required to validate or reject a causal relationship between CCM and flowering time in wild-type siblings. Experiments using populations with greater recombination or allelic diversity, such as in an association panel, should disrupt most fortuitous linkage and would provide a second test of this phenotypic correlation and either argue for or against additional exploration of this relationship. The crosses of *Oy1-N1989/+* to the association panel analyzed for CCM in a previous study (Khangura *et al.* 2019) could be replanted and measured for flowering time of wild-type and *Oy1-N1989/+* mutant sibling pairs. Candidate gene testing of epistatic interactions between *Oy1-N1989, vey1*, and the known flowering time regulators segregating in that population (e.g. *zmmads69, cct10, zcn8, dlf1*, and *vgt1)* could provide some insight. Epistasis, indicating interaction between chlorophyll biosynthetic disruption and developmental determinants of the transition to flowering, would be consistent with photosynthesis acting as part of the autonomous pathway whereas no genetic interaction would be consistent with the slow growth of plant organs in a compromised background resulting in the observed reproductive delays.

Even though the Syn10-DH and IBM-RILs are derived from the same parents, they differ in the method of development and rates of recombination (Liu et al. 2015; Hussain et al. 2007; Lee et al. 2002). Thus, these two populations yield different levels

of resolution for QTL detection. Our previous analysis to fine map *vey1* QTL using CCM values showed Syn10-DH to have a higher mapping resolution than the IBM-RILs population (Khangura *et al.* 2019). We also observed different QTL for flowering time in these populations. QTL analysis in Syn10-DH F_1 population detected the *oof1* QTL on chromosome 2 affecting MT_DTA and MT_DTS as well as the major-effect locus *vey1* (Table S8). The detection of *oof1* was dependent on the presence of the *Oy1-N1989* mutant, but was neither contingent nor displayed any epistatic interactions with *vey1*. Thus, *oof1* appears to be a novel locus of independent mechanism affecting flowering time in the *Oy1-N1989* mutants. This QTL on chromosome 2 was not detected for any other trait in the Syn10-DH and IBM-RILs testcross populations with *Oy1-N1989/+*:B73 in the current study nor was it identified as a modifier of chlorophyll content in our previous study (Khangura *et al.* 2019). The amount of phenotypic variation explained by *vey1* in flowering time was higher in the Syn10-DH experimental material than the IBM-RILs testcross population by ~20% (Tables S7 and S8). Greater variation in flowering time in the IBM-RILs testcross progenies that could not be modeled by marker genotypes may result from higher residual heterozygosity or greater rates of pollen contamination during the single seed descent (SSD) procedure used to generate the IBM-RILs (Lee et al. 2002). Heterozygosity is expected to be negligible in Syn10-DH population because of the DH procedure employed to fix allele segregation during population development (Hussain *et al.* 2007). In addition, we had a moderate increase in sample size in the Syn10-DH (251 lines) as compared to the IBM-RILs (216 lines) that is expected to result in a marginally greater power to detect QTL in the Syn10-DH F_1 populations.

Chlorophyll levels are correlated with the rate of photosynthesis in plants (Huang *et al.* 2009). Some controversy has been reported in mutants of soybean affected in an ortholog of *oy1* with some reports showing little impact on photosynthesis (Sakowska *et al.* 2018) and others clearly demonstrating a linear relationship between chlorophyll variation and photosynthesis (Walker *et al.* 2017). The reasons for conflicting conclusions results from soybean mutants are not fully clear but it warrants some caution in making simplistic interpretations about the impact of chlorophyll deficient mutants on photosynthesis. In our study, net CO_2 assimilation was associated with the severity of the *Oy1-N1989* mutant phenotype. The effect of chlorophyll loss on CO_2 assimilation measurements using a LICOR instrument was somewhat non-linear. A nearly 5-fold reduction in CCM in *Oy1-N1989* NILs carrying a B73 allele at *vey1* resulted in in a 22% reduction in net CO_2 assimilation while the 10-fold reduction in CCM in *Oy1-N1989* NILs carrying a Mo17 allele at *vey1* resulted in a 64% reduction in net photosynthetic rate, compared to their isogenic wild-type siblings (Table A.1). Ultimately, both reductions resulted in less accumulation of free sugars and starch, with a substantially greater reduction in the enhanced mutant NILs (Figure A.7; Tables A.1 and S11).

The allelic interactions of *Oy1-N1989* and wild-type *oy1* alleles are consistent with the inductive role of carbohydrate status on floral transition (Ohto et al. 2001; Seo et al. 2011; Wahl et al. 2013; Minow et al. 2018). Additional work exploring proposed carbohydrate signaling metabolites, such as T6P and organic acids, and analyses of the downstream floral integrators such as the FT orthologs of maize are still needed to link our results to the existing models of floral transition regulation in maize (Minow *et al.*

2018).

One complexity that we observed in our data is that defoliation did not separate photosynthate levels from changes in chlorophyll (Figure A.8). The mechanical removal of leaves resulted in changes to CCM in the newly emerging leaves of defoliated plants. As a result, a mechanical treatment served to highlight the interconnected nature of plant metabolism: large changes to any feature of central metabolism or plant physiology results in large changes to all of central metabolism and plant physiology. In an effort to separate photosynthate and chlorophyll accummulation, one can conceive of alternate experiments such as measuring the time to maturity in different light intensities achieved using neutral shade cloth to produce a gradient of photosynthetically active radiation and sugar accumulation. Similarly, plants could be sprayed with low doses of chemicals such as 3-(3,4-dichlorophenyl)-1,1-dimethylurea (DCMU) that disrupt electron transfer during the light reactions to reduce photosynthetic output of plants without altering light fluence experienced by other photoreceptors. However, all these experiments may suffer from the same confounding of NSC and chlorophyll as the defoliation experiment. As a result, researchers should quantify chlorophyll after these treatments as any change in chlorophyll complicate our ability to uncouple chlorophyll metabolism and sugar accumulation as demonstrated by our defoliation experiment and all the carbohydrate partitioning mutants of maize studied to date.

Two other subunits of magnesium chelatase are encoded by genes that were identified in Arabidopsis as *genomes uncoupled* mutants, *gun4* and *gun5*, with altered retrograde, plastid-to- nuclear, signaling with defects in chlorophyll metabolism (Susek

et al. 1993; Mochizuki *et al.* 2001; Larkin *et al.* 2003). Multiple studies have demonstrated retrograde signaling mutants with defects in chlorophyll metabolism (Hernández-Verdeja and Strand 2018) and circadian rhythm (Jones 2019). Retrograde signaling events are also associated with cell death, chloroplast development, and etiolation but it acts via a number of modifiers of flowering, including phytochromes, phytochrome signaling components, blue light perception via *crytochrome1*, and circadian rhythmicity of key genes that affect flowering time. While retrograde signaling is an attractive alternative model to sugar signaling for the phenomena reported here, mutants in orthologs of *oy1* in both monocots and dicots (Mochizuki *et al.* 2001; Gadjieva *et al.* 2005) have been tested for a *genomes uncoupled* phenotype and they did not perturb retrograde signaling. This makes it unlikely that *Oy1-N1989* and *vey1* are altering flowering time via aberrant retrograde signals. Additional experiments are necessary to clearly separate the effects of chlorophyll metabolite levels from photosynthate levels to independently test their effects on flowering time in maize.

We have known that defoliation is practiced in some maize nurseries to stagger flowering time and permit intercrossing of lines with divergent reproductive maturities. We had presumed that this was based on published research. Remarkably, we were not able to find a reference for this practice. Previous work on defoliation in maize has looked at the effect of early and late season defoliation on growth, and yield components but did not report flowering times (Crookston and Hicks 1988; Pearson and Fletcher 2009). As a result, Figure A.8 provides information that was informally shared within the maize genetics research community but not described formally. In addition, we extended this observation from maize to both sorghum and green foxtail,

indicating that defoliation can achieve staggered flowering times in these species as well. This may be a general feature of grasses, but appears to not be universal in angiosperms as early-season defoliation of photoperiod-sensitive strawberries did not affect flowering time (Guttridge 1959).

Sugar export and phloem loading mutants in maize that carry lesions in *tie-dyed1*, *tie-dyed2*, *sucrose export defective1*, *psychedelic*, and *sucrose trasporter1* have all been shown to delay flowering time (Braun *et al.* 2006; Baker and Braun 2008; Ma *et al.* 2008; Slewinski *et al.* 2009; Slewinski and Braun 2010). We also observed declines in sugar levels in *Oy1-N1989* mutants (Figure A.7 and Table S11) and a delay in flowering time (Figure A.1). It is tempting to consider these mutants as a demonstration that sugar signaling can work independent of chlorophyll, but these mutants also display low chlorophyll contents (Braun *et al.* 2006; Baker and Braun 2008; Ma *et al.* 2008; Slewinski *et al.* 2009; Slewinski and Braun 2010). Curiously, *sucrose export defective1* encodes tocopherol cyclase (Porfirova *et al.* 2002; Sattler *et al.* 2003). Tocopherol and chlorophyll share the phytol side chain, and the salvage pathway for phytol side chains from chlorophyll can contribute to tocopherol accumulation in both maize and Arabidopsis (Ischebeck *et al.* 2006; Schelbert *et al.* 2009; Diepenbrock *et al.* 2017). While it is attractive to try and unify these findings, chlorophyll breakdown products were only rate limiting for tocopherol synthesis in senescent tissues of Arabidopsis (Zhang *et al.* 2014). Furthermore, in *Oy1-N1989* mutants the phytol precursors should be abundant as this pool is not being consumed by chlorophyll biosynthesis. Future experiments that more carefully explore these metabolites, for instance, via other mutants that do not simultaneously affect multiple pathways,

especially chlorophyll metabolism, are required.

Accumulation of sugars in the leaves due to sink disruption has been proposed to induce leaf senescence in a variety of angiosperms, including maize and Arabidopsis (Allison and Weinmann 1970; Ceppi *et al.* 1987; Pourtau *et al.* 2006; Sekhon *et al.* 2012, 2019). In maize, leaf senescence is triggered when pollination is prevented, and sucrose accumulates in leaves due to the lack of the sink activity of a pollinated ear (Allison and Weinmann 1970; Ceppi *et al.* 1987). This sucrose- dependent leaf senescence is genotype-dependent, and B73 is particularly susceptible to this phenomenon. Genetic inheritance of the induced senescence phenomenon in maize inbred line B73 was proposed to be under the control of a single dominant locus (Ceppi *et al.* 1987). The leaf senescent phenotypes of the sucrose export mutants demonstrate that excessive photosynthate accumulation can cause tissue to senesce regardless of ear presence (Braun *et al.* 2006; Baker and Braun 2008; Ma *et al.* 2008; Slewinski *et al.* 2009). Induced leaf senescence by sink removal or sugar application shows some overlap of biochemical and molecular mechanisms with natural senescence in plants (summarized in Sekhon *et al*. 2012). A study of gene expression in the B73 inbred of maize identified senescence associated genes that exhibit gene expression changes during pollination-prevented leaf senescence (Sekhon *et al*. 2012). We found that leaf senescence could be prevented or delayed by the suppression of photosynthesis in the *Oy1-N1989* mutant and further modulated by *vey1* variants (Figure A.9 and Table S12). We expect that future gene expression studies in *Oy1-N1989* mutant and sugar export mutants will identify genes consistently impacted by lower and higher chlorophyll levels, presumably including senescence associated genes and genes regulating

carbohydrate metabolism.

Acknowledgements

We would like to thank the National Science Foundation for NSF grant 1444503 awarded to G.S.J. and B.P.D which helped fund this research. We thank the Purdue ACRE and ICSC staff members Mr. Jim Beaty and Mr. Jason Adams for help with planting, and management of the field experiments described in this work. We also thank past Johal lab members for help with planting and data collection. We would like to thank Mr. Mike Gosney for helping with gas-exchange measurements. We gratefully acknowledge solidarity with the operator of the compact track loader who provided a soundtrack and dry workspace by excavating and constructing a French drain on our office walls while we wrote this paper.

References

Abe, M., Y. Kobayashi, S. Yamamoto, Y. Daimon, A. Yamaguchi et al., 2005 FD , a bZIP Protein Mediating Signals from the Floral Pathway Integrator FT at the Shoot Apex. Science 309: 1052–1056.

Allison, J. C. S., and H. Weinmann, 1970 Effect of absence of developing grain on carbohydrate content and senescence of maize leaves. Plant Physiol. 46: 435–436.

Alter, P., S. Bircheneder, L.-Z. Zhou, U. Schlüter, M. Gahrtz et al., 2016 Flowering Time- Regulated Genes in Maize Include the Transcription Factor ZmMADS1. Plant Physiol. 172: 389–404.

Amasino, R. M., and S. D. Michaels, 2010 The Timing of Flowering. Plant Physiol. 154: 516–520.

Baker, R. F., and D. M. Braun, 2008 Tie-dyed2 functions with tie-dyed1 to promote carbohydrate export from maize leaves. Plant Physiol. 146: 1085–1097.

Bolanos, J., and G. O. Edmeades, 1996 The importance of the anthesis-silking interval in breeding for drought tolerance in tropical maize. F. Crop. Res. 48: 65–80.

Bouchet, S., P. Bertin, T. Presterl, P. Jamin, D. Coubriche et al., 2017 Association mapping for phenology and plant architecture in maize shows higher power for developmental traits compared with growth influenced traits. Heredity (Edinb). 118: 249–259.

Braun, D. M., Y. Ma, N. Inada, M. G. Muszynski, and R. F. Baker, 2006 tie-dyed1 Regulates carbohydrate accumulation in maize leaves. Plant Physiol. 142: 1511–22.

Broman, K. W., H. Wu, S. Sen, and G. a. Churchill, 2003 R/qtl: QTL mapping in experimental crosses. Bioinformatics 19: 889–890.

Buckler, E. S., J. B. Holland, P. J. Bradbury, C. B. Acharya, P. J. Brown et al., 2009 The genetic architecture of maize flowering time. Science 325: 714–718.

Ceppi, D., M. Sala, E. Gentinetta, A. Verderio, and M. Motto, 1987 Genotype-Dependent Leaf Senescence in Maize: Inheritance and effects of pollination-prevention. Plant Physiol. 85: 720–725.

Chen, M., J. Chory, and C. Fankhauser, 2004 Light Signal Transduction in Higher Plants. Annu. Rev. Genet. 38: 87–117.

Cho, L. H., R. Pasriga, J. Yoon, J. S. Jeon, and G. An, 2018 Roles of Sugars in Controlling Flowering Time. J. Plant Biol. 61: 121–130.

Cho, L. H., J. Yoon, and G. An, 2017 The control of flowering time by environmental factors. Plant J. 90: 708–719.

Colasanti, J., Z. Yuan, and V. Sundaresan, 1998 The indeterminate gene encodes a zinc finger protein and regulates a leaf-generated signal required for thetransition to flowering in maize. Cell 93: 593–603.

Coles, N. D., M. D. McMullen, P. J. Balint-Kurti, R. C. Pratt, and J. B. Holland, 2010 Genetic control of photoperiod sensitivity in maize revealed by joint multiple population analysis. Genetics 184: 799–812.

Coneva, V., D. Guevara, S. J. Rothstein, and J. Colasanti, 2012 Transcript and metabolite signature of maize source leaves suggests a link between transitory starch to sucrose balance and the autonomous floral transition. J. Exp. Bot. 63: 5079–5092.

Coneva, V., T. Zhu, and J. Colasanti, 2007 Expression differences between normal and indeterminate1 maize suggest downstream targets of ID1, a floral transition regulator in maize. J. Exp. Bot. 58: 3679–3693.

Corbesier, L., P. Lejeune, and G. Bernier, 1998 The role of carbohydrates in the induction of flowering in Arabidopsis thaliana: Comparison between the wild type and a starchless mutant. Planta 206: 131–137.

Corbesier, L., C. Vincent, S. Jang, F. Fornara, Q. Fan et al., 2007 FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. Science 316: 1030–1033.

Crookston, R. K., and D. R. Hicks, 1988 Effect of Early Defoliation on Maize Growth and Yield: an Eleven-year Perspective. Crop Sci. 28: 371–373.

Danilevskaya, O. N., X. Meng, D. A. Selinger, S. Deschamps, P. Hermon et al., 2008 Involvement of the MADS-Box Gene ZMM4 in Floral Induction and Inflorescence Development in Maize. Plant Physiol. 147: 2054–2069.

DePristo, M. A., E. Banks, R. Poplin, K. V Garimella, J. R. Maguire et al., 2011 A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43: 491–498.

Diepenbrock, C. H., C. B. Kandianis, A. E. Lipka, M. Magallanes-Lundback, B. Vaillancourt et al., 2017 Novel Loci Underlie Natural Variation in Vitamin E Levels in Maize Grain. Plant Cell 29: 2374–2392.

Doebley, J., A. Stec, and L. Hubbard, 1997 The evolution of apical dominance in maize. Nature 386: 485–488.

Eichten, S. R., J. M. Foerster, N. de Leon, Y. Kai, C.-T. Yeh et al., 2011 B73-Mo17 near-isogenic lines demonstrate dispersed structural variation in maize. Plant Physiol. 156: 1679–90.

Evans, M. M., and R. S. Poethig, 1995 Gibberellins promote vegetative phase change and reproductive maturity in maize. Plant Physiol. 108: 475–87.

Eveland, A. L., and D. P. Jackson, 2012 Sugars, signalling, and plant development. J. Exp. Bot. 63: 3367–3377.

Gadjieva, R., E. Axelsson, U. Olsson, and M. Hansson, 2005 Analysis of gun phenotype in barley magnesium chelatase and Mg-protoporphyrin IX monomethyl ester cyclase mutants. Plant Physiol. Biochem. 43: 901–908.

Gibson, L. C. D., R. D. Willows, C. G. Kannangara, D. von Wettstein, and C. N. Hunter, 1995 Magnesium-protoporphyrin chelatase of Rhodobacter sphaeroides:

reconstitution of activity by combining the products of the bchH, -I, and -D genes expressed in Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. 92: 1941–1944.

Guo, L., X. Wang, M. Zhao, C. Huang, C. Li et al., 2018 Stepwise cis-Regulatory Changes in ZCN8 Contribute to Maize Flowering-Time Adaptation. Curr. Biol. 28: 3005–3015.e4.

Guttridge, C. G., 1959 Evidence for a Flower Inhibitor and Vegetative Growth Promoter in the Strawberry. Ann. Bot. 23: 351–360.

Hansson, A., C. G. Kannangara, D. V Wettstein, and M. Hansson, 1999 Molecular basis for semidominance of missense mutations in the XANTHA-H (42-kDa) subunit of magnesium chelatase. Proc. Natl. Acad. Sci. U. S. A. 96: 1744–1749.

Hansson, A., R. D. Willows, T. H. Roberts, and M. Hansson, 2002 Three semidominant barley mutants with single amino acid substitutions in the smallest magnesium chelatase subunit form defective AAA+ hexamers. Proc. Natl. Acad. Sci. U. S. A. 99: 13944–13949.

Hernández-Verdeja, T., and Å. Strand, 2018 Retrograde Signals Navigate the Path to Chloroplast Development. Plant Physiol. 176: 967–976.

Huang, M., T. L. Slewinski, R. F. Baker, D. Janick-Buckner, B. Buckner et al., 2009 Camouflage patterning in maize leaves results from a defect in porphobilinogen deaminase. Mol. Plant 2: 773–89.

Huang, C., H. Sun, D. Xu, Q. Chen, Y. Liang et al., 2017 ZmCCT9 enhances maize adaptation to higher latitudes. Proc. Natl. Acad. Sci. 115: E334–E341.

Hung, H.-Y., L. M. Shannon, F. Tian, P. J. Bradbury, C. Chen et al., 2012 ZmCCT and the genetic basis of day-length adaptation underlying the postdomestication spread of maize. Proc. Natl. Acad. Sci. 109: E1913–E1921.

Hussain, T., P. Tausend, G. Graham, and J. Ho, 2007 Registration of IBM2 SYN10 Doubled Haploid Mapping Population of Maize. J. Plant Regist. 1: 81.

Ischebeck, T., A. M. Zbierzak, M. Kanwischer, and P. Dörmann, 2006 A salvage pathway for phytol metabolism in Arabidopsis. J. Biol. Chem. 281: 2470–2477.

Jones, M. A., 2019 Retrograde signalling as an informant of circadian timing. New Phytol. 221: 1749–1753.

Khangura, R. S., S. Marla, B. P. Venkata, N. J. Heller, G. S. Johal et al., 2019 A Very Oil Yellow1 Modifier of the Oil Yellow1-N1989 Allele Uncovers a Cryptic Phenotypic Impact of Cis- regulatory Variation in Maize. G3 9: 375–390.

Kozaki, A., S. Hake, and J. Colasanti, 2004 The maize ID1 flowering time regulator is a zinc finger protein with novel DNA binding properties. Nucleic Acids Res. 32: 1710–1720.

Lang, A., 1952 Physiology of Flowering. Ann. Rev. Plant. Physiol. 3: 265–306.

Lang, A., 1957 The effect of gibberellin upon flower formation. Proc. Natl. Acad. Sci. 43: 709– 717.

Larkin, R. M., J. M. Alonso, J. R. Ecker, and J. Chory, 2003 GUN4, a regulator of chlorophyll synthesis and intracellular signaling. Science 299: 902–906.

Leach, K. A., and D. M. Braun, 2016 Soluble Sugar and Starch Extraction and Quantification from Maize (Zea mays) Leaves. Curr. Protoc. Plant Biol. 1: 139–161.

Leach, K. A., T. M. Tran, T. L. Slewinski, R. B. Meeley, and D. M. Braun, 2017 Sucrose transporter2 contributes to maize growth, development, and crop yield. J. Integr. Plant Biol. 59: 390–408.

Lee, M., N. Sharopova, W. D. Beavis, D. Grant, M. Katt et al., 2002 Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. Plant Mol. Biol. 48: 453–61.

Liang, Y., Q. Liu, X. Wang, C. Huang, G. Xu et al., 2019 ZmMADS69 functions as a flowering activator through the ZmRap2.7-ZCN8 regulatory module and contributes to maize flowering time adaptation. New Phytol. 221: 2335–2347.

Liu, H., Y. Niu, P. J. Gonzalez-Portilla, H. Zhou, L. Wang et al., 2015 An ultra-high-density map as a community resource for discerning the genetic basis of quantitative traits in maize. BMC Genomics 16: 1078.

Ma, Y., R. F. Baker, M. Magallanes-Lundback, D. Dellapenna, and D. M. Braun, 2008 Tie-dyed1 and Sucrose export defective1 act independently to promote carbohydrate export from maize leaves. Planta 227: 527–538.

Meng, X., M. G. Muszynski, and O. N. Danilevskaya, 2011 The FT-Like ZCN8 Gene Functions as a Floral Activator and Is Involved in Photoperiod Sensitivity in Maize. Plant Cell 23: 942– 960.

Minow, M. A. A., L. M. Ávila, K. Turner, E. Ponzoni, I. Mascheretti et al., 2018 Distinct gene networks modulate floral induction of autonomous maize and photoperiod-dependent teosinte. J. Exp. Bot. 69: 2937–2952.

Mochizuki, N., J. a Brusslan, R. Larkin, a Nagatani, and J. Chory, 2001 Arabidopsis genomes uncoupled 5 (GUN5) mutant reveals the involvement of Mg-chelatase H subunit in plastid- to-nucleus signal transduction. Proc. Natl. Acad. Sci. U. S. A. 98: 2053–2058.

Moghaddam, M. R. B., and W. Van den Ende, 2013 Sugars, the clock and transition to flowering. Front. Plant Sci. 4: 1–6.

Muszynski, M. G., T. Dam, B. Li, D. M. Shirbroun, Z. Hou et al., 2006 delayed flowering1 Encodes a basic leucine zipper protein that mediates floral inductive signals at the shoot apex in maize. Plant Physiol. 142: 1523–1536.

Mutasa-Göttgens, E., and P. Hedden, 2009 Gibberellin as a factor in floral regulatory networks. J. Exp. Bot. 60: 1979–1989.

Ohto, M.-A., K. Onai, Y. Furukawa, E. Aoki, T. Araki et al., 2001 Effects of Sugar on Vegetative Development and Floral Transition in Arabidopsis. Plant Physiol. 127: 252–261.

Paul, M. J., 2008 Trehalose 6-phosphate: a signal of sucrose status. Biochem. J. 412: 1–2.

Pearson, A., and A. L. Fletcher, 2009 Effect of total defoliation on maize growth and yield. Agron. New Zeal. 39: 1–6.

Porfirova, S., E. Bergmuller, S. Tropf, R. Lemke, and P. Dormann, 2002 Isolation of an Arabidopsis mutant lacking vitamin E and identification of a cyclase essential for all tocopherol biosynthesis. Proc. Natl. Acad. Sci. 99: 12495–12500.

Pourtau, N., R. Jennings, E. Pelzer, J. Pallas, and A. Wingler, 2006 Effect of sugar-induced senescence on gene expression and implications for the regulation of senescence in Arabidopsis. Planta 224: 556–568.

R core Team, 2013 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org/.

Sakowska, K., G. Alberti, L. Genesio, A. Peressotti, G. Delle Vedove et al., 2018 Leaf and canopy photosynthesis of a chlorophyll deficient soybean mutant. Plant Cell Environ. 41: 1427–1437.

Salvi, S., G. Sponza, M. Morgante, D. Tomes, X. Niu et al., 2007 Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. Proc. Natl. Acad. Sci. U. S. A. 104: 11376–11381.

Samach, A., H. Onouchi, S. E. Gold, G. S. Ditta, Z. Schwarz- et al., 2000 Distinct Roles of CONSTANS Target Genes in Reproductive Development of Arabidopsis. Science 288: 1613–1616.

Sattler, S. E., E. B. Cahoon, S. J. Coughlan, and D. Dellapenna, 2003 Characterization of Tocopherol Cyclases from Higher Plants and Cyanobacteria . Evolutionary Implications for Tocopherol Synthesis and Function. Plant Physiol. 132: 2184–2195.

Sawers, R. J. H., J. Viney, P. R. Farmer, R. R. Bussey, G. Olsefski et al., 2006 The maize Oil yellow1 (Oy1) gene encodes the I subunit of magnesium chelatase. Plant Mol. Biol. 60: 95– 106.

Schelbert, S., S. Aubry, B. Burla, B. Agne, F. Kessler et al., 2009 Pheophytin Pheophorbide Hydrolase (Pheophytinase) Is Involved in Chlorophyll Breakdown during Leaf Senescence in Arabidopsis. Plant Cell 21: 767–785.

Searle, N. E., 1965 Physiology of Flowering. Ann. Rev. Plant. Physiol. 16: 97–118.

Sekhon, R. S., K. L. Childs, N. Santoro, C. E. Foster, C. R. Buell et al., 2012 Transcriptional and Metabolic Analysis of Senescence Induced by Preventing Pollination in Maize. Plant Physiol. 159: 1730–1744.

Sekhon, R. S., C. Saski, R. Kumar, B. Flinn, F. Luo et al., 2019 Integrated Genome-Scale Analysis Identifies Novel Genes and Networks Underlying Senescence in Maize. Plant Cell 31: 1968- 1989.

Sen, T. Z., L. C. Harper, M. L. Schaeffer, C. M. Andorf, T. E. Seigfried et al., 2010 Choosing a genome browser for a Model Organism Database: surveying the Maize community. Database 2010: 1–9.

Seo, P. J., J. Ryu, S. K. Kang, and C. M. Park, 2011 Modulation of sugar metabolism by an INDETERMINATE DOMAIN transcription factor contributes to photoperiodic flowering in Arabidopsis. Plant J. 65: 418–429.

Slewinski, T. L., and D. M. Braun, 2010 The Psychedelic genes of maize redundantly promote carbohydrate export from leaves. Genetics 185: 221–232.

Slewinski, T. L., A. Garg, G. S. Johal, and D. M. Braun, 2010 Maize SUT1 functions in phloem loading. Plant Signal. Behav. 5: 1–4.

Slewinski, T. L., R. Meeley, and D. M. Braun, 2009 Sucrose transporter1 functions in phloem loading in maize leaves. J. Exp. Bot. 60: 881–892.

Steinhoff, J., W. Liu, J. C. Reif, G. Della Porta, N. Ranc et al., 2012 Detection of QTL for flowering time in multiple families of elite maize. Theor. Appl. Genet. 125:.

Susek, R. E., F. M. Ausubel, and J. Chory, 1993 Signal Transduction Mutants of Arabidopsis Uncouple Nuclear CAB and RBCS Gene Expression from Chloroplast Development. Cell 74:787–799.

Swarts, K., R. M. Gutaker, B. Benz, M. Blake, R. Bukowski et al., 2017 Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America. Science 357: 512–515.

Vega, S. H., M. Sauer, J. A. J. Orkwiszewski, and S. Poethig, 2002 The early phase change gene in Maize. Plant Cell 14: 133–147.

Wahl, V., J. Ponnu, A. Schlereth, S. Arrivault, T. Langenecker et al., 2013 Regulation of flowering by Trehalose-6-phosphate signalling in Arabidopsis thaliana. Science 339: 704–708.

Walker, B. J., D. T. Drewry, R. A. Slattery, A. VanLoocke, Y. B. Cho et al., 2017 Chlorophyll can be reduced in crop canopies with little penalty to photosynthesis. Plant Physiol. 176: 1215– 1232.

Walker, C. J., and J. D. Weinstein, 1991 In vitro assay of the chlorophyll biosynthetic enzyme Mg- chelatase: resolution of the activity into soluble and membrane-bound fractions. Proc. Natl. Acad. Sci. U.S.A. 88: 5789–5793.

Wang, R. L., A. Stec, J. Hey, L. Lukens, and J. Doebley, 1999 The limits of selection during maize domestication. Nature 398: 236–239.

Wigge, P. A., M. C. Kim, K. E. Jaeger, W. Busch, M. Schmid et al., 2005 Integration of Spatial and Temporal Information During Floral Induction in Arabidopsis. Science 309: 1056–1059.

Yoo, S. K., K. S. Chung, J. Kim, J. H. Lee, S. M. Hong et al., 2005 CONSTANS Activates SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 through FLOWERING LOCUS T to Promote Flowering in Arabidopsis. Plant Physiol. 139: 770–778.

Zeevaart, J. A. D., 1962 Physiology of Flowering. Science 137: 723–731.

Zhang, W., T. Liu, G. Ren, S. Hörtensteiner, Y. Zhou et al., 2014 Chlorophyll Degradation: The Tocopherol Biosynthesis-Related Phytol Hydrolase in Arabidopsis Seeds Is Still Missing. Plant Physiol. 166: 70–79.

Zhang, L., H. Yu, S. Lin, and Y. Gao, 2016 Molecular Characterization of FT and FD Homologs from Eriobotrya deflexa Nakai forma koshunensis. Front. Plant Sci. 7: 1–10.

TABLE

Table A. 1

The chlorophyll index, gas-exchange, and maximum quantum yield of the mutant (*Oy1- N1989/+*) and wild-type (+/+) siblings in four independent B73-like NILs x *Oy1-N1989/+*:B73 F_1 progenies.

Genotype[1]	***vey1*-status**	**CCM**[1]	**A**[2]	**g_s**[3]	**C_i**[4]	**E**[5]	**F_v/F_m**[6]
Oy1-N1989/+:b094	*vey1*^{Mo17}	1.5±0.04^a	4.9±1.55^a	0.2±0.07	331.3±5.01^a	4.3±0.99	0.16±0.03^a
Oy1-N1989/+:b189	*vey1*^{Mo17}	1.6±0.14^a	6.3±0.85^a	0.2±0.06	319.9±24.93^a	4.7±0.92	0.16±0.04^a
Oy1-N1989/+:b135	*vey1*^{B73}	4.9±0.27^b	27.9±2.47^b	0.3±0.01	172.4±13.36^b	5.8±0.49	0.43±0.01^b
Oy1-N1989/+:b185	*vey1*^{B73}	4.4±0.29^b	23.5±1.66^b	0.3±0.03	202.1±27.97^b	5.5±0.26	0.42±0.02^b
+/+:b094	*vey1*^{Mo17}	19.7±3.14	34.4±3.97	0.4±0.12	176.1±25.61	6.9±0.84	0.66±0.04
+/+:b189	*vey1*^{Mo17}	14.4±0.66	27.8±0.91	0.3±0.03	166.4±7.39	5.5±0.41	0.70±0.05
+/+:b135	*vey1*^{B73}	19.2±3.47	32.6±3.32	0.3±0.05	162.0±13.11	6.4±0.52	0.72±0.03
+/+:b185	*vey1*^{B73}	22.4±1.84	33.2±2.49	0.4±0.06	166.6±13.76	6.5±0.53	0.72±0.03

The data are provided as mean ± standard deviations of three experimental replications. Each replication consisted of three independent plants measurements for mutants and two for wild-type siblings. The superscript connecting letter report between each genotype group (mutant or wild-type) indicates statistical significance determined using ANOVA with post-hoc analysis using Tukey's HSD at $p<0.01$. No statistically significant difference was found between the wild-type siblings.

[1]Chlorophyll index measured using CCM-200 plus; [2]Net CO_2 assimilation rate (µmol CO_2 m^{-2} s^{-1}); [3]Stomatal conductance (mol H_2O m^{-2} s^{-1}); [4]Substomatal CO_2 concentration (µmol CO_2 mol air^{-1}); [5]Transpiration rate (mmol m^{-2} s^{-1}); [6]Maximum quantum yield of PSII (Fv/Fm)

FIGURES

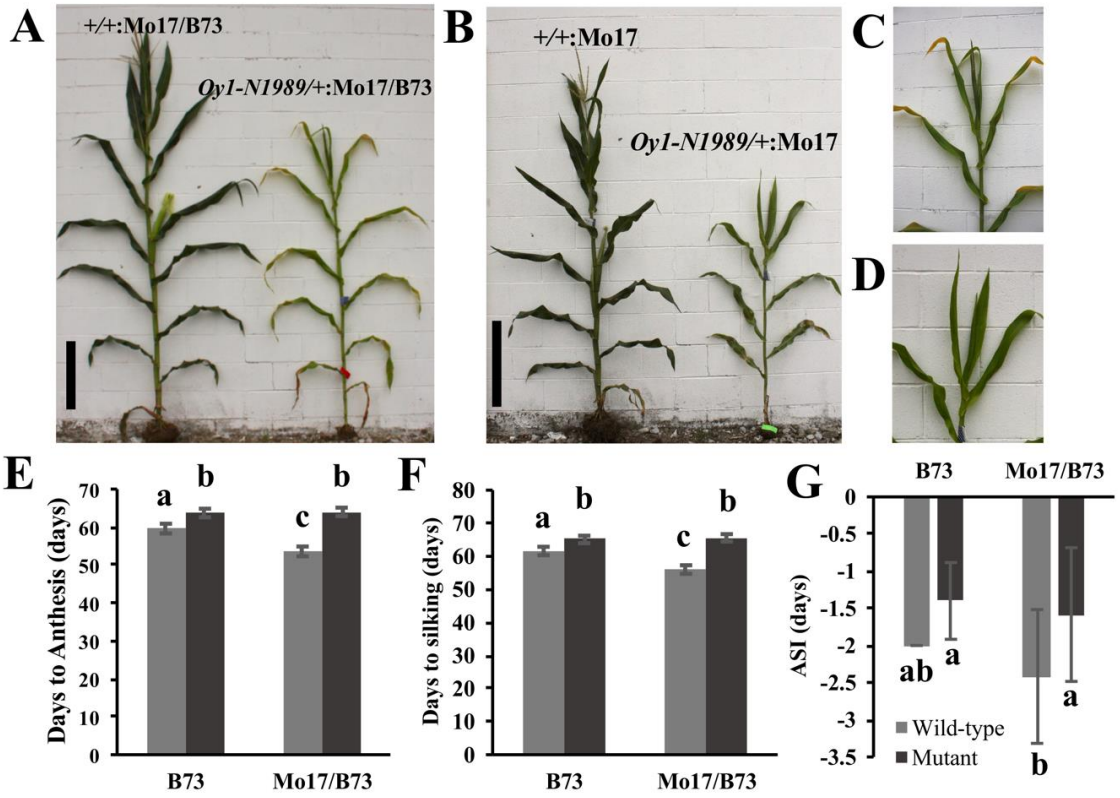


Figure A. 1

The reproductive maturity is delayed by *Oy1-N1989* allele in maize.

The representative (A) wild-type (left) and mutant (right) sibling from Mo17 x *Oy1-N1989/+*:B73 cross (reproduced from Figure 1b in Khangura *et. al*. 2019), black scale bar = 50 cm. The representative (B) wild-type (left) and mutant (right) sibling from Mo17 x *Oy1-N1989/+*:Mo17 cross (BC7). The close-up view of the emerging tassel of mutant siblings (C) in panel A, and (D) panel B. The distribution of (E) days to anthesis, (F) days to silking, and (G) anthesis-silking interval (ASI) in the wild-type and mutant siblings in B73 and Mo17 x B73 (Mo17/B73) hybrid genetic backgrounds.

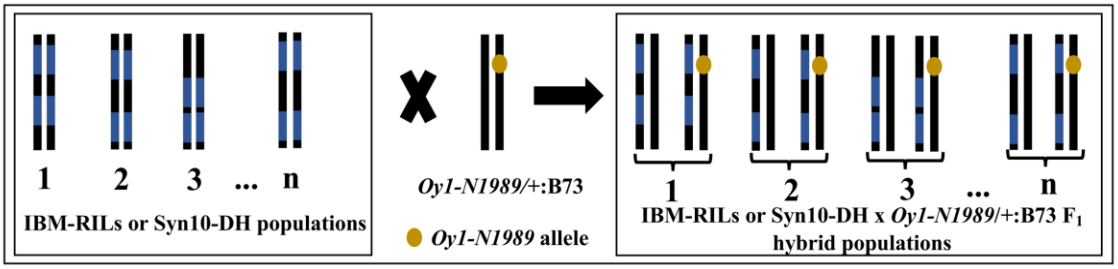


Figure A. 2

The schematic of crossing strategy used to map *Oy1-N1989* enhancer/suppressor loci using IBM-RILs (n=216) and Syn10-DH (n=251) populations.

Black and blue colors indicate B73 and Mo17 genotypes. Chromosome 10 of the heterozygous pollen-parent *Oy1-N1989/+*:B73 is shown with a golden spot indicating *Oy1-N1989* mutant allele. The resulting F1 progenies from these crosses are depicted with state of chromosome 10 for each F1 testcross showing segregation of wild-type and mutant (with the golden spot) siblings.

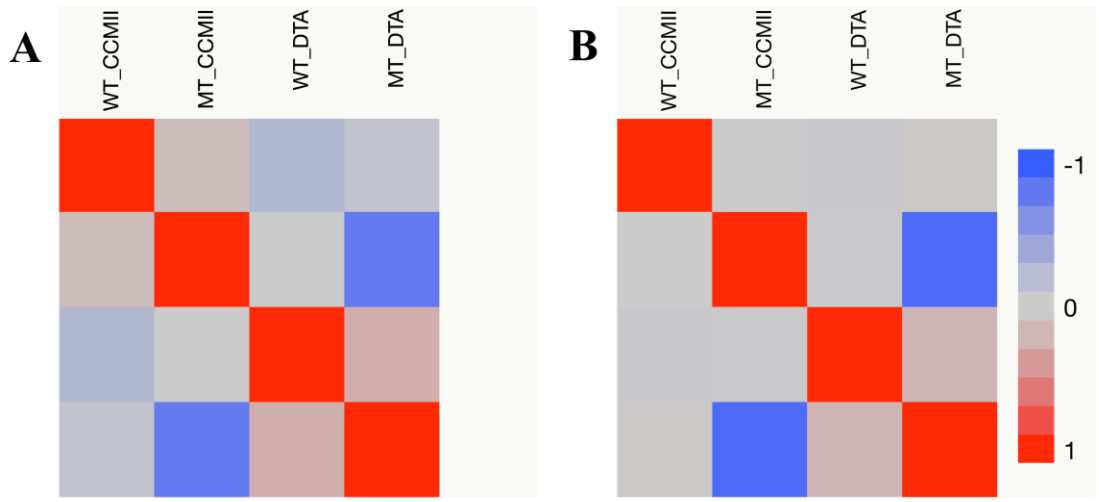


Figure A. 3

The pairwise correlation matrix heatmap between wild-type CCMII (WT_CCMII), mutant CCMII (MT_CCMII), wild-type days to anthesis (WT_DTA), and mutant days to anthesis (MT_DTA) in (A) IBM-RILs x *Oy1-N1989/+*:B73, and (B) Syn10-DH x *Oy1-N1989/+*:B73 F1 test cross populations.

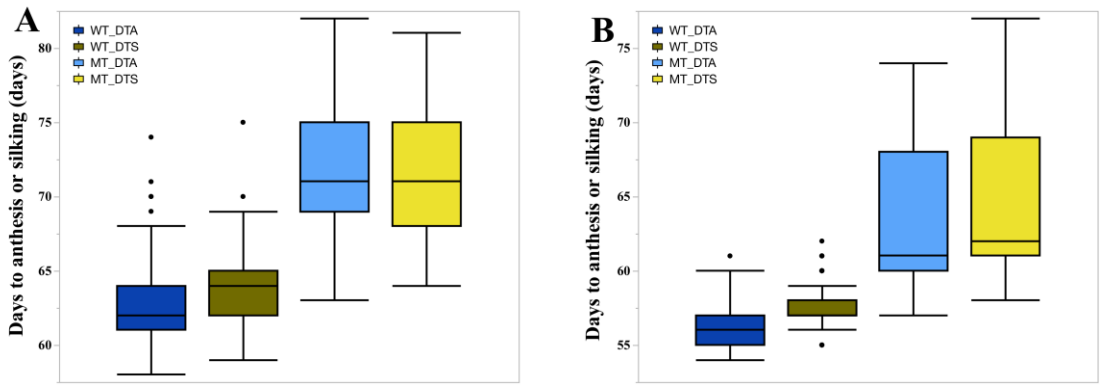


Figure A. 4

The distribution of days to flower (anthesis and silking) in the wild-type and mutant siblings in (A) IBM-RILs x ***Oy1-N1989/+*:B73, and (B) Syn10-DH** x ***Oy1-N1989/+*:B73 F1 test cross populations.**

Abbreviations: wild-type days to anthesis (WT_DTA), wild-type days to silking (WT_DTS), mutant days to anthesis (MT_DTA), and mutant days to silking (DTS).

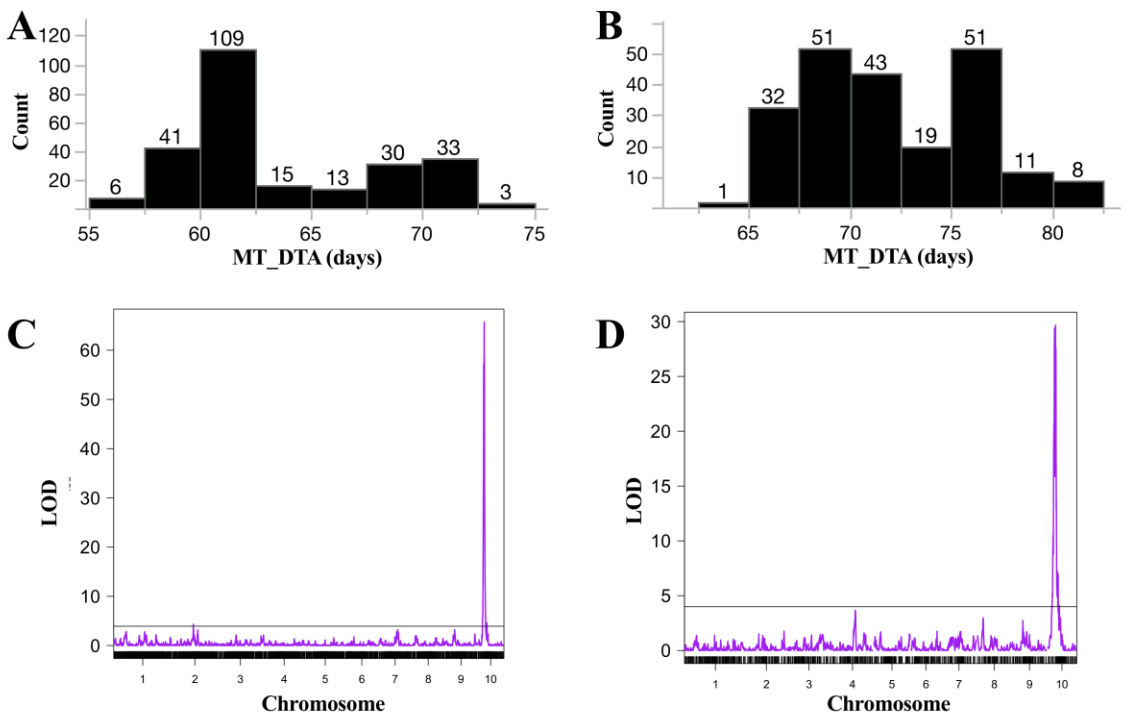


Figure A. 5

The distribution of days to anthesis of the mutant siblings (MT_DTA) in (A) Syn10-DH x ***Oy1-N1989/+*:B73, and (B) IBM-RILs** x ***Oy1-N1989/+*:B73 F1 population.**

Genome-wide QTL plot of MT_DTA in (C) Syn10-DH, and (D) IBM-RILs F1 population. The x-axis in (C) and (D) indicates the chromosome number, and y-axis indicates the logarithm of odds (LOD) of tested markers. Black horizontal bar indicates permutation testing based threshold for QTL detection.

Figure A. 6

Effect of *vey1* locus from B73 and Mo17 on reproductive maturity in reciprocal isogenic backgrounds.

Representative plants showing (A) Delayed maturity of the heterozygous mutants in isogenic Mo17 (BC7 generation) inbred background, compared to the wild-type siblings. Mutant and wild-type siblings in b030 (B73-like NIL with *vey1B73*) shows early flowering while the isogenic B73 introgression of the *vey1Mo17* allele in b094 NIL exhibits delayed flowering. (B) Early maturity in the heterozygous mutant in m097 (Mo17-like NIL with *vey1B73*) compared to the mutant in B73 x Mo17 F1 hybrid background. Measuring stick on the left in both panels is 243 cm.

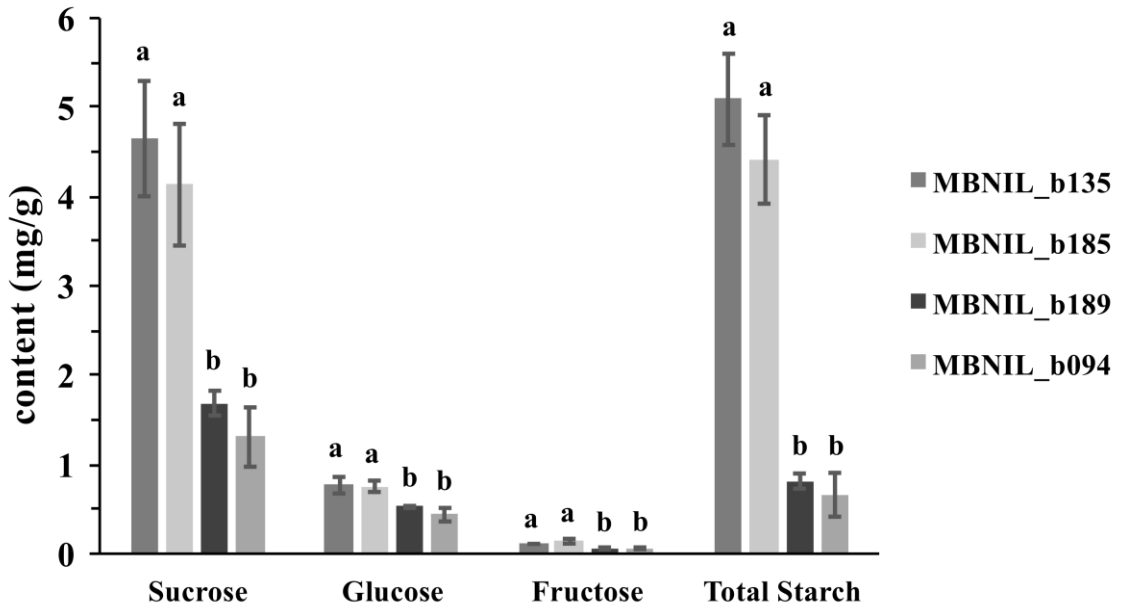


Figure A. 7

The effect of alleles at *vey1* on leaf soluble sugars and starch content in *Oy1-N1989/+* heterozygotes in the isogenic B73 background.

B73-like NILs b135 and b185 have *vey1B73* introgression, whereas b189 and b094 have *vey1Mo17* introgression. The values of different sugars and starch are reported as mg/g of fresh weight.

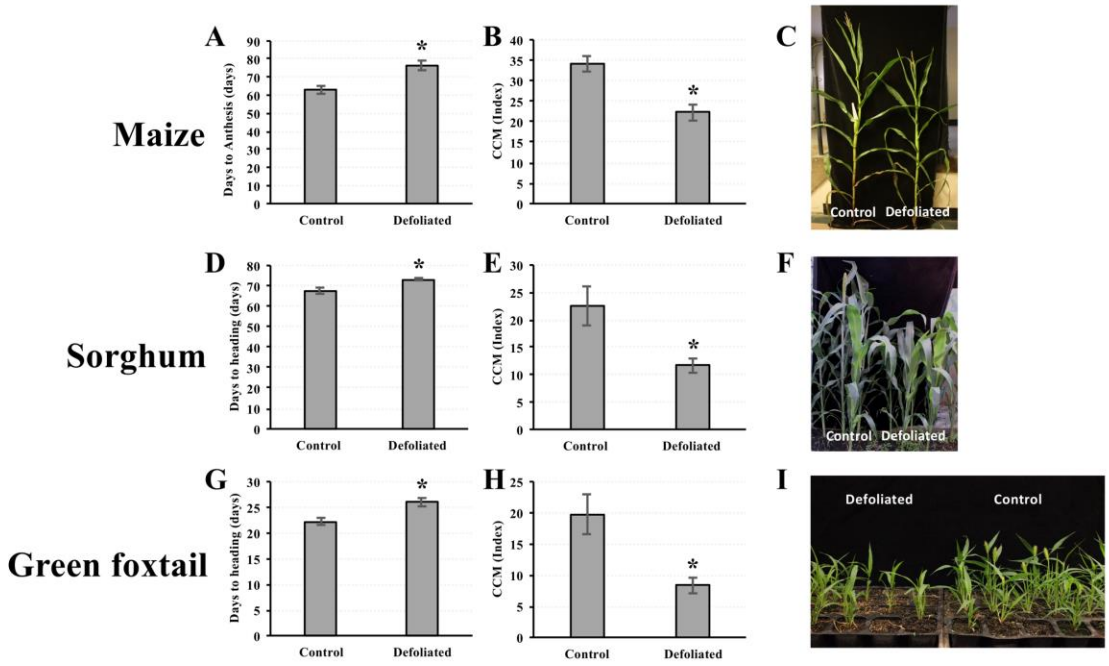


Figure A. 8

The effect of defoliation on reproductive maturity and chlorophyll in the newly emerged leaves of (A-C) maize, (D-F) sorghum, and (G-I) green foxtail.

The asterisk indicates significant difference between treatment means using student's t-test at $p<0.05$.

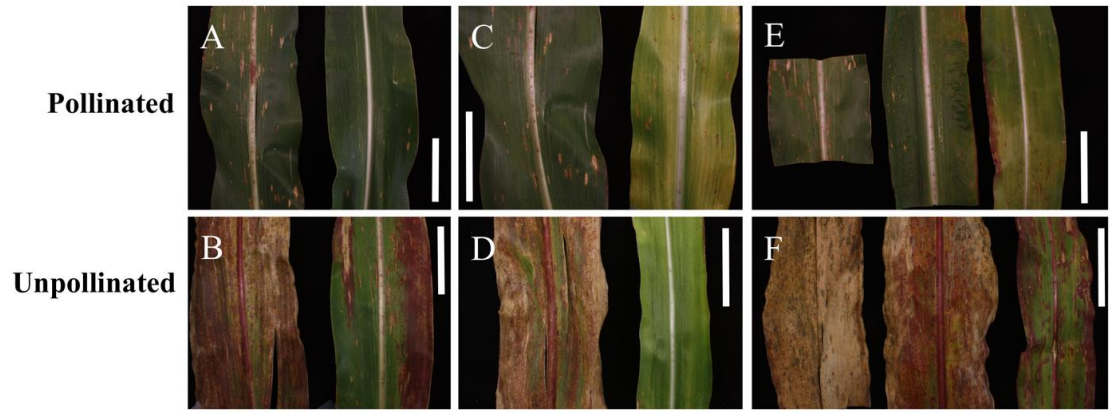


Figure A. 9

The effect of *vey1* on senescence induced by pollination prevention.

Pictures of the representative primary ear leaf derived from plants at 30 and 42 days-after-anthesis (DAA) either with open pollinated (top panels: A, C, and E) or unpollinated (bottom panels: B, D, and F) ears. The representative primary ear leaf from (A-B) B73 wild-type (left) and mutant (right) sibling at 30DAA, (C-D) wild-type (left) and mutant (right) sibling from B73-like NIL-b107 (homozygous *vey1Mo17*) ´ *Oy1-N1989/+*:B73 cross at 30 DAA, (E-F) wild-type sibling (B73-like NIL-b107), mutant B73-like NIL-b135 with *vey1B73* introgression (middle), and mutant B73-like NIL-b107 sibling (right) from crosses with *Oy1-N1989/+*:B73 at 42 DAA. The scale bar in each figure is 6.35 cm.

Supplemental Materials:

Supplemental material available at Figshare: https://doi.org/10.25387/g3.9985415.

VITA

Singha Raj Dhungana was born April 4th, 1989, in Kathmandu, Nepal to Murari Prasad Dhungana and Ambika Devi Dhungana. Upon graduating from high school in Kathmandu, Nepal, Singha came to the USA to pursue higher education. He obtained his Associate Degree in Science from Northern Virginia Community College in 2012. He then attended Truman State University and joined the research group of Dr. Diane Janick-Buckner and Dr. Brent Buckner to study the genetic basis of maize leaf development. Singha graduated from Truman State University with a B.S. in Biology in 2014. After graduation, he worked for a biotech startup company in St. Louis, Missouri until 2016. He joined the Division of Biological Sciences graduate program at University of Missouri-Columbia and joined the lab of Dr. David Braun to understand the genetic control of carbohydrate partitioning in maize. He also collaborated with researchers from other institutions working with sorghum and genomic data on sugarcane. He completed his doctorate degree in Biological Sciences in May 2022, during which he was the first author of one review paper and one research article, and co-author of two research articles.