SORGHUM BICOLOR GENES INVOLVED IN RAFFINOSE OLIGOSACCHARIDE METABOLISM

A Dissertation

by

MANISH THAKRAN

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	John Mullet
Committee Members,	Keerti Rathore
	Hisashi Koiwa
	Wayne Versaw
Head of Department,	Dirk Hays

May 2021

Major Subject: Molecular and Environmental Plant Sciences

ABSTRACT

Sorghum bicolor is an important lignocellulosic biofuel crop. Its small genome size of ~800 Mbp, reference genome sequence, genetic and genomic resources make it a genetic model for C4 bioenergy grasses. There is a significant natural variation in sugar content in stems of cultivated sorghums. Sweet and biomass sorghums accumulate sugars (mainly sucrose) to 30% of total stem dry weight post-anthesis. These levels rival those of sugarcane but at a fraction of the input as sorghum has low nutritional requirements, high biomass yield, and sustainability features such as drought resistance, high water, and nitrogen use efficiency. The final non-structural carbohydrate content of the stem increases its potential as a biofuel crop.

The analysis of the biochemical and genetic basis of sucrose accumulation in sweet sorghum stems has been a topic of research for more than 20 years. Many gene candidates involved in sugar metabolism, sugar transport, and transcription factors have been analyzed which shed light on the sugar allocation process.

In the present study, genes involved in raffinose family of oligosaccharides (RFOs) metabolism, namely, galactinol synthase (GolS), raffinose synthase (RafS), stachyose synthase (StaS), and alpha-galactosidase (AGA) responsible for the biosynthesis and degradation of RFOs were identified in the sorghum genome. A homology-based approach was used to identify 2 *SbGolS*, 1 *SbRS,* 1 *SbSTS,* and 6 *SbAGA* genes in the sorghum genome. Transcriptome analyses showed that *SbGolS, SbRS,* and *SbAGA* genes exhibited distinct expression profiles in different tissues and developmental stages. The up-regulation of *SbRS* in the leaves and the up-regulation of *SbAGA1* and *SbAGA2* genes

post-anthesis in stems suggested their potential role in RFO synthesis in leaves and hydrolysis in sweet sorghum stems. This study found that genes involved in RFO hydrolysis are induced in sweet sorghum stems post-anthesis when stems are accumulating high levels of sucrose and starch. This is consistent with the hypothesis that RFO is synthesized in leaves, transported to stems, and subsequently hydrolyzed to release sucrose, thereby contributing to the accumulation of high levels of nonstructural carbohydrates in the stems of some bioenergy sorghum genotypes.

DEDICATION

I dedicate this dissertation to my dearest parents and late aunt, who gave me everything I needed to succeed.

ACKNOWLEDGEMENTS

I want to thank my committee chair, Dr. John Mullet, for giving me this opportunity and guidance throughout my Ph.D. I'd also like to thank my committee members, Dr. Rathore, Dr. Koiwa, Dr. Versaw, for their guidance and support throughout this research.

Thanks also go to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University a great experience.

Finally, thanks to my friends and lifting buddies I made at the Texas A&M Rec Center, who encouraged me to be a better version of myself.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor John Mullet of the Department of Biochemistry and Biophysics Department, Professor Keerti Rathore of the Soil and Crop Sciences Department, Professor Hisashi Koiwa of the Horticulture Department, Professor Wayne Versaw of the Department of Biology.

Field studies were planted and maintained by the Rooney lab that was used in research in Chapter III.

Starla Zemelis-Durfee carried out the RNA in-situ hybridization work for Chapter III at Michigan State University as a collaborator. The RNA-seq data was generated by the Joint Genome Institute and analyzed by Dr. Brian McKinley of the Biochemistry and Biophysics Department at Texas A&M University.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

This graduate study was supported by Adkisson Chair in Agricultural biology. This work was also made possible in part by funding from the Department of Energy Great Lakes Bioenergy Center.

NOMENCLATURE

BSA	Bovine Serum Albumin
CDS	Coding Sequence
DAE	Days After Emergence
DTT	Dithiothreitol
EISA	Energy Independence and Security Act of 2007
GRIN	Germplasm Resources Information Network
GWAS	Genome-Wide Association Study
ICRISAT	International Crops Research Institute for the Semi-Arid Tropics
IMAC	Immobilized Metal Affinity Chromatography
IPTG	Isopropyl-β-D-thiogalactopyranoside
MEME	Multiple Expression motifs for Motif Elicitation
MUSCLE	Multiple Sequence Comparison by Log-Expectation
MWCO	Molecular Weight Cut-Off
NJ	Neighbor Joining
PNG	4-nitrophenyl-α-D-galactopyranoside
QTL	Quantitative Trait Locus
RFOs	Raffinose Family of Oligosaccharides
RFS	Renewable Fuel Standards

RPM	Rotations Per Minute
SDS-PAGE	SDS-Polyacrylamide Gel Electrophoresis
SNP	Single Nucleotide Polymorphism
SWEETs	Sugars Will Eventually be Exported Transporters
TPM	Transcripts Per Kilobase Million
USDA	United States Department of Agriculture
WSCs	Water-Soluble Carbohydrates

TABLE OF CONTENTS

Page

ABSTRACT ii

DEDICATION iv

ACKNOWLEDGEMENTS v

CONTRIBUTORS AND FUNDING SOURCES vi

NOMENCLATURE vii

TABLE OF CONTENTS ix

LIST OF FIGURES xi

LIST OF TABLES xiv

1. INTRODUCTION AND LITERATURE REVIEW 1

Sorghum background 1
Renewable energy and biofuels 1
Sorghum as a biofuel crop 3
Carbon partitioning and phloem transport in sorghum 7
Role of sucrose metabolism genes and sucrose transporters 12
Sorghum transgenics 14
Raffinose family of oligosaccharides in sorghum 15

2. CHAPTER II 19

Sequence retrieval of RFO gene sequences 19
Multiple sequence alignments and phylogenetic analysis 22
Physicochemical properties and conserved motif analysis of RFO metabolism proteins 27
Heterologous expression and *in vitro* biochemical characterization of SbAGA1 and SbAGA2 31
Results and discussion 39

3. CHAPTER III 40

Introduction 40

Materials and methods 43
Plant materials and growing conditions 43
Primer design and qPCR 44
RNA-seq expression analysis 45
RNA *in situ* hybridization (RNA-ish) 46
Results 47
Discussion 60

4. CONCLUSION 65

REFERENCES 67

APPENDIX A SUPPLEMENTARY INFORMATION 74

APPENDIX B SUPPLEMENTARY FIGURES 75

APPENDIX C SUPPLEMENTARY TABLES 80

LIST OF FIGURES

Page

Figure 1.1: Renewable Fuel Standards (RFS) mandated under the Energy Independence and Security Act of 2007 (EISA) 2

Figure 1.2 Time course of biomass accumulation during the development of sweet sorghum Della. (a) Duration of Della development stages from germination to floral induction, floral induction and booting to anthesis, and anthesis to grain maturity. (b) Time course of dry weight accumulation (Mg per hectare) in leaves (dark green), leaf sheaths (light green), stems (brown), grain (yellow), and apical tillers (blue) in the field in College Station, 2013. The data shown are the mean ± SE of nine biological replicates. [39].....8

Figure 2.1: Phylogenetic relationships of RFO-synthetic- and selected RFO-hydrolytic enzymes from various taxa. (Names in Supplementary Information S1).....23

Figure 2.2: Selected regions of the protein sequence alignment among raffinose synthases (RS) and alpha galactosidases (AGA).....24

Figure 2.3: Sequence and phylogenetic analysis of SbRS, SbSTS, AtRS, and AtRSTS. **A.** Schematically shows the 783 aa long sequence of AtRS5 (A5g40390) from *A. thaliana* and the 792 aa long sequence of SbRS (Sobic.003G052300) from *S. bicolor*. **B.** Shows a section of a sequence alignment performed with Clustal Omega of RafS and StaS amino acid sequences, which revealed the very high amino acid identity and similarity, except for an 80 amino acid long sequence block insertion **C.** Schematically shows the 876 aa long sequence of AtSTS (At4g01970) from *A. thaliana* and the 904 aa long sequence of SbRS (Sobic.003G052300) from *S. bicolor*. Dashed orange arrows point to the ~80 aa gap in alignment in the raffinose synthases.....26

Figure 2.4**:** Block diagram of conserved motifs in GolS proteins identified in S*orghum bicolor*. Each motif is represented with a box in different colors: motif 1, red; motif 2, cyan; motif 3, green; motif 4, purple; and motif 5, yellow. 28

Figure 2.5**:** Block diagram of conserved motifs in RS, STS, and AGAs proteins identified in *Sorghum bicolor*. Each motif is represented with a box in different color: motif 1, red; motif 2, cyan; motif 3, light green; motif 4, purple; and motif 5, yellow; and motif 6, green 29

Figure 2.6**:** SDS-PAGE of recombinant SbAGA1. Eluate obtained during induction, lysis, and purification washes of the recombinant SbAGA1was analyzed on SDS-PAGE. Lane (M), molecular mass marker; Lane (U), Uninduced control; Lane (I), Induced culture; Lane (L), Lysate; Lane (W 1-3), consecutive washes; Lane (E 1-3), consecutive elutions. 34

Figure 2.7**:** Alpha-galactosidase hydrolysis of a chromogenic substrate PNG. 36

Figure 2.8**:** The effect of different pH values on the alpha-galactosidase enzyme activity of recombinant SbAGA1 and SbAGA2. The activity was measured with 20mM PNG, according to Gao and Schaffer[86]. 37

Figure 2.9**:** The AGA activity shows clear raffinose specificity when raffinose and stachyose are compared (measured at pH 7.0 with 50 mM each of raffinose and stachyose). Data are means ± SE of 4 replicates. 38

Figure 3.1**:** Expression levels of *SbRS*, *SbAGA1*, and *SbAGA2* in sweet sorghum (cultivar Della) stem and leaves, pre, and post-anthesis. A, B, and C show *SbRS*, *SbAGA1*, and *SbAGA2,* respectively. Values are means ± standard error of n = 3 plants. Relative gene expression is shown compared to endogenous *SbUBC* as a normalization control. Relative expression was calculated via the $2^{-\Delta\Delta C_t}$ method relative to the internode section with the gene's highest expression in each group [97]. The fold change in expression between the minimum and maximum values on the y-axis was calculated based on *SbUBC* normalized values according to $\mathrm{FC} = 2^{\Delta C_t(\max) - \Delta C_t(\min)}$. ... 51

Figure 3.2**:** Expression levels of *SbRS*, *SbAGA1*, and *SbAGA2* in sweet sorghum (cultivar Wray) stems and leaves, pre, and post-anthesis. A, B, and C show *SbRS*, *SbAGA1*, and *SbAGA2,* respectively. Values are means ± standard error of n = 3 plants. Relative gene expression is shown compared to endogenous *SbUBC* as a normalization control. Relative expression was calculated via the $2^{-\Delta\Delta C_t}$ method relative to the internode section with the gene's highest expression in each group [97]. Fold change in expression between the minimum and maximum values on the y-axis was calculated based on *SbUBC* normalized values according to $\mathrm{FC} = 2^{\Delta C_t(\max) - \Delta C_t(\min)}$. ... 52

Figure 3.3**:** Expression levels of *SbRS*, *SbAGA1*, and *SbAGA2* in sweet sorghum cultivar Rio leaves and stems post-anthesis. Samples were taken from the field. A, B, and C show *SbRS*, *SbAGA1*, and *SbAGA2,* respectively. Values are means ± standard error of n = 3 plants. Relative gene expression is shown compared to endogenous *SbUBC* as a normalization control. Relative expression was calculated via the $2^{-\Delta\Delta C_t}$ method relative to the internode section with the gene's highest expression in each group [97]. Fold change in expression between the minimum and maximum values on

the y-axis was calculated based on SbUBC normalized values according to $FC = 2^{\Delta C_t(\max) - \Delta C_t(\min)}$. 54

Figure 3.4**:** Expression levels of *SbRS*, *SbAGA1*, and *SbAGA2* in sweet sorghum cultivar Umbrella leaves and stems post-anthesis. Samples were taken from the field. A, B, and C show *SbRS*, *SbAGA1*, and *SbAGA2,* respectively. Values are means ± standard error of n = 3 plants. Relative gene expression is shown compared to endogenous *SbUBC* as a normalization control. Relative expression was calculated via the $2^{-\Delta\Delta C_t}$ method relative to the internode section with the gene's highest expression in each group [97]. The fold change in expression between the minimum and maximum values on the y-axis was calculated based on SbUBC normalized values according to $FC = 2^{\Delta C_t(\max) - \Delta C_t(\min)}$. 55

Figure 3.5**:** Expression levels of *SbGolS1* in sweet sorghum cultivar Della leaves and stems, pre and post-anthesis. 56

Figure 3.6**:** Expression levels of *SbGolS1* in sweet sorghum cultivar Wray leaves and stems, pre and post-anthesis. 57

Figure 3.7**:** RNA in-situ hybridization controls from Della sorghum leaves. A) Negative control: DapB- a bacterial transcript with no detection and pink background. B) Positive control: Leaf tissue with Cytochrome P450 transcript localization in the bundle sheath cells surrounding the leaves' minor veins at 140 DAE. Transcript localization is visible as red precipitate marked with black arrows Magnification using a 40X objective. 58

Figure 3.8**:** RNA in-situ hybridization results from Della sorghum leaves. A) SbRS transcript localization in mesophyll cells surrounding the minor veins of leaves at 140 DAE. B) SbAGA1 transcript localization in the bundle sheath cells surrounding the leaves' minor veins at 140 DAE. C) SbAGA2 transcript localization in the bundle sheath cells surrounding the leaves' minor veins at 140 DAE. Magnification using a 40X objective.....58

Figure 3.9**:** RNA in-situ hybridization results from sorghum leaves. A) SbRS transcript localization in mesophyll cells surrounding the major veins of leaves at 140 DAE. B) SbAGA1 transcript localization in the xylem parenchyma (XP) cells surrounding the leaves' major veins at 140 DAE. Magnification using a 40X objective.....59

LIST OF TABLES

Page

Table 2.1 :Raffinose metabolism genes in *Sorghum bicolor* 21

Table 3.1**:** The expression level of the RFO pathway genes in grain sorghum BTx623 by RNA-seq in various tissues at different developmental stages. The numbers represent the average expression obtained from the TPM (Transcripts Per Kilobase Million) plots. DAE- Days after emergence. 48

Table 3.2**:** The expression level of the RFO pathway genes in sweet Sorghum Della by RNA-seq in stem internodes from floral initiation to grain maturity. The numbers represent the average expression obtained from the TPM (Transcripts Per Kilobase Million) plots. 50

1. INTRODUCTION AND LITERATURE REVIEW

Sorghum background

Renewable energy and biofuels

The Environmental Protection Agency (EPA) administers and enforces the Renewable Fuel Standards (RFS) mandated under the Energy Independence and Security Act of 2007 (EISA), which requires renewable fuel to be blended into transportation fuel in an increasing amount each year. The RFS was amended in 2007 (RFS2); it mandates the blending of 36 billion gallons of renewable fuel by 2022. It also caps conventional corn-starch based ethanol at 16 billion gallons [1] Figure 1.1.

Since corn is an input-intensive crop, the limit on corn-starch-based ethanol as biofuel feedstock ensures food security and availability of arable land resources. In addition to mandating RFS, EISA also demands reduction of greenhouse gas emissions such as CO_2 to mitigate the increase in atmospheric CO_2 due to decades of burning fossil fuels.

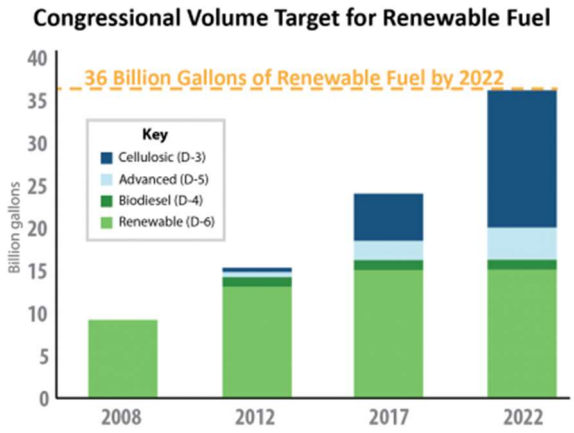

Figure 1.1: Renewable Fuel Standards (RFS) mandated under the Energy Independence and Security Act of 2007 (EISA)

The conventional biofuel made from corn-starch has surpassed its limit of 16 billion gallons in 2018 (~16.091 billion gallons) and 2019 (~15.775 billion gallons). It is no longer allowed to contribute towards the standards set forth by RFS2. Therefore, advanced biofuel feedstocks primarily from lignocellulosic biomass grown on land marginal for food production are being developed to meet the mandate while adhering to current standards.

There is a worldwide push to develop net carbon-negative biofuel crops, leading to increased soil carbon. When coupled with CO_2 mitigation from the conversion process, it could lead to overall net negative CO_2 emission in biofuel's lifecycle. As lignocellulosic bioethanol production becomes more efficient and cost-effective, several bioenergy crops are being investigated to meet the world's energy demand, which is expected to rise significantly. Among the potential candidates, C4 crops such as miscanthus, switchgrass,

and sorghum are promising candidates for bioenergy crops because they have high light, water, and nitrogen use efficiency. Within the C4 crops, sorghum has been advanced as a biofuel crop due to its unique genotypic and phenotypic characteristics. Among other attributes, some sorghum genotypes and hybrids are photoperiod sensitive, allowing long duration of vegetative growth and production requiring minimal fertilizer on marginal lands.

Sorghum as a biofuel crop

Sorghum crops are believed to have originated in present-day Sudan about ~10,000 years ago, from where they spread to Sub-Saharan Africa and parts of India, the Middle East, and Asia[2]. The genetic and phenotypic diversity in sorghum has been driven by its large geographical distribution in Sub-Saharan Africa, India, the Middle East, and other parts of Asia[3]. This geographical spread of sorghum crops has led to distinct sorghum varieties, which can be identified by their floral architecture and seed shape. This spread to various agro-climatic regions, and adaptation gave sorghum a very diverse germplasm to breed and introduce commercially important genes into cultivated sorghums. In the last 150 years in the US, extensive human and natural selection has further led to the diversification of the sorghum germplasm to make it amenable to its unique temperate climate and combine harvesting[4].

Sorghum (*Sorghum bicolor* L. Moench), a member of the Poaceae family, is the third-largest cereal crop grown in the US and fifth (http://www.fao.org/faostat) most important cereal crop in the world [5]. It is a major cereal crop in the semi-arid region of Sub-Sahran Africa and is regarded as a "failsafe" crop in the global agroecosystem[6]. It has also emerged as an important biofuel crop in the US. Sorghum is an excellent C4 crop genetic model mainly due to the fact that sorghum is diploid and primarily an inbreeding species that contains a small genome comprised of ten chromosomes that span ~800Mb. Before the availability of a fully sequenced sorghum genome, researchers and breeders took advantage of genotype-based sequencing to understand the sorghum's genetic diversity[7]. Sorghum was the first C4 grass to be sequenced through traditional Sanger sequencing[5], Next-generation sequencing[8], and long-read sequencing [9] approaches, which provide an excellent genetic resource for crop improvement. The BTx623 reference genome was updated using the Next-generation sequencing methodology, which helped improve the annotation and provided insight into the genome organization[8]. The sweet sorghum Rio has also been fully sequenced to understand differential sugar accumulation in these genotypes [9]. However, there is a high similarity between the sweet sorghum and grain sorghum genotypes and sequences differences did not explain the differences between sweet sorghum genotypes and grain sorghums. The transgenic amenable and repeat-rich sorghum cultivar Tx430 was also sequenced using Oxford Nanopore sequences generated on a MinION sequencer[10]. Several other comprehensive genomic resources are available for sorghum, which include a QTL atlas [11], an expression atlas [12], a transcriptome atlas [8], various association panels[13], and a sorghum SNP

database called SorGSD [14]. Several traits and their underlying QTLs have been identified for sorghum adaptation in different environments and yield improvement [11, 15, 16]. Sorghum has diverse germplasm, which can be accessed through public databases such as GRIN and ICRISAT.

Sorghum is a short-day plant, and it has been adapted to earlier flowering in long days in temperate climates for optimum grain yield. The underlying maturity loci (*Ma*) that control its photoperiod sensitivity were identified to help develop high biomass sorghum hybrids with delayed flowering from early flowering R- and A/B-line inbreds. The genes corresponding to five of the six Maturity loci have been identified. *Ma*1, which significantly influences flowering time, encodes SbPRR37, a pseudo response regulator that inhibits flowering in long days[17]. *Ma*3 encodes phytochrome B (phyB)[18], *Ma*5 encodes phytochrome C (phyC)[19], *Ma*6 encodes Ghd7, a repressor of flowering in long days [20], and *Ma*2 encodes a SMYD protein which has lysine methyltransferase activity[21].

To reduce lodging and make sorghum amenable to mechanized harvesting, four dwarfing loci (*Dw1*, *Dw2*, *Dw3*, and *Dw4*) were identified, and three of them have been cloned and characterized[22, 23]. Among the dwarfing loci, *Dw3* was the first one to be characterized [24]. High biomass sorghums have gained popularity as biofuel crops[25]. Due to its good water-use efficiency, drought, and heat tolerance, sorghum cultivation will likely increase worldwide in the future[26]. To further increase its yield and make it amenable for bioenergy production, successful genetic manipulation of desired traits is critical to producing higher biomass and yields, which is now possible due to advances in

sorghum transgenics. The sorghum transformation efficiency has significantly increased and is no longer restricted to a handful of genotypes[27].

Sorghum, native to the Horn of Africa, is naturally very diverse. It displays tremendous genetic and phenotypic diversity, which can be attributed to its habitat, conventional breeding for various purposes, and its spread throughout the African continent, parts of the Middle East and Asia, India, and Australia. Sorghum can be classified into four different types based on its utility and morphological traits: grain sorghum, sweet sorghum, forage sorghum, and energy sorghum[28].

Among these genotypes, the tall and late-maturing sweet sorghums are of great importance as a lignocellulosic feedstock for biofuel production[26, 29]. These high biomass sorghums accumulate up to ~30% of their fresh weight in sucrose in their stem internodes, which can be mechanically extracted and converted to bioethanol. Because sucrose accumulation in sweet sorghum is an essential attribute that sets it apart from grain sorghums, several studies have been conducted to determine the differences underlying this trait. These studies range from physiological, biochemical, whole genome, transcriptomic, metabolomic[30], microRNA[31], and expression analysis and provide valuable insight into differential sucrose accumulation in sweet sorghum genotypes compared to grain sorghums[9, 31-33].

Carbon partitioning and phloem transport in sorghum

Grain sorghum, as the name suggests, was selected for a high yield of grain. In contrast, sweet or biomass sorghum accumulates biomass and large amounts of sucrose in the stem internodes post-anthesis. Sweet sorghums accumulate up to 24-fold more sucrose post floral initiation than in the vegetative phase [34].

Genetic diversity studies were unable to determine when and how the high stem sugar trait appeared in sorghum[35] as sweet sorghum accessions cluster together with grain sorghums. In a genome-wide association study (GWAS) of water-soluble carbohydrates (WSCs), a group claimed that a species-specific tandem duplication in sweet sorghum results in a putative neofunctionalization event involving Sobic.004G01500, which is responsible for higher WSC accumulation in sweet sorghum Rio[36].

Sweet sorghums improved through conventional breeding efforts are rapidly approaching their sucrose accumulation limit in the stem internodes[37, 38]. Therefore, new approaches need to be explored to bypass this bottleneck. Even though continuous efforts are being made to understand the underlying molecular basis for differential carbon partitioning between these sorghum types, a complete understanding of stem sucrose accumulation is lacking. Studies have looked at differences in sucrose metabolism enzymes, phloem loading in leaves, and phloem unloading in the stem. Still, these don't appear to explain differential carbon partitioning in sweet vs. non-sweet genotypes. To

understand these differences, we need to first look at how photoassimilates are transported from the source leaves to the stem sink during sorghum development. In sweet sorghum and sugarcane, sucrose is the primary sugar transported from leaves and accumulated in stem internodes.

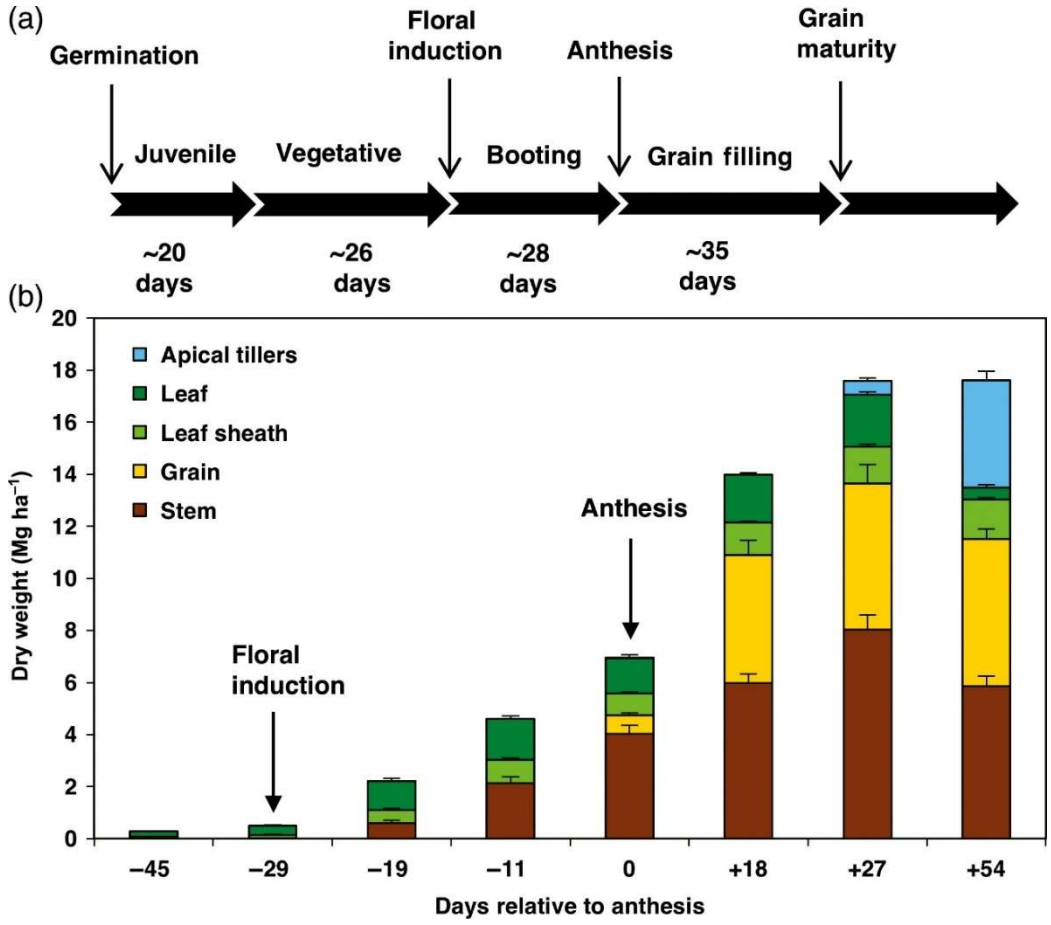

Figure 1.2 Time course of biomass accumulation during the development of sweet sorghum Della. (a) Duration of Della development stages from germination to floral induction, floral induction and booting to anthesis, and anthesis to grain maturity. (b) Time course of dry weight accumulation (Mg per hectare) in leaves (dark green), leaf sheaths (light green), stems (brown), grain (yellow), and apical tillers (blue) in the field in College Station, 2013. The data shown are the mean ± SE of nine biological replicates. [39]

In Figure 1.2, sweet sorghum stem composition data collected over 100 days of sorghum development ranging from floral induction to grain maturity revealed that sweet sorghums could accumulate high sucrose concentrations within their stems (McKinley et al., 2016). The sweet sorghum Della accumulates biomass primarily in its leaves until floral initiation and then in the stem during stem elongation until anthesis, followed by panicles until grain maturity and then in tillers. The primary photosynthate, sucrose in sweet sorghum Della, is used for growth and cell wall biosynthesis in growing vegetative phase stems. Sucrose is mainly accumulated in sweet sorghum stems post floral initiation to support grain filling after anthesis, providing a buffer under conditions when photosynthesis is inhibited by water deficit [39]. Activities that utilize sucrose during stem growth, such as cell wall biosynthesis, and that hydrolyze sucrose, such as a vacuolar invertase, are down-regulated after floral initiation. Decreased consumption and hydrolysis of sucrose in stems is correlated with sucrose and starch accumulation. Sucrose and starch levels in the stem peak between anthesis and grain maturity in sweet sorghums. The accumulation of sucrose in the sweet sorghum Wray was correlated with an up-regulation of Tonoplast Sugar Transporters (TSTs) that transport sucrose into vacuoles of the pith parenchyma cells that surround the phloem/xylem vasculature in stems [40]. The molecular basis for the difference in carbohydrate partitioning between sweet and grain sorghum types is still not completely understood.

Carbohydrate partitioning occurs when photoassimilates are distributed from their sites of synthesis in leaves to their sites of utilization or accumulation. Carbohydrate partitioning

begins with photosynthesis, the assimilation of carbon dioxide in leaf chloroplasts, and the subsequent metabolism of derived sugars. Sucrose accumulation in sweet sorghum stems coincides with the reproductive phase of sorghum growth, but it can also start earlier, depending on the genotype [41]. The first step in phloem loading starts at the source leaves where sucrose is synthesized, and it involves sucrose export from the mesophyll cells to the point of entry into the collection phloem. Sucrose is mobilized from the mesophyll cells to bundle sheath cells and then to the phloem parenchyma through plasmodesmatal connections (symplasmic transport). In C4 grasses such as sorghum, sucrose is unloaded from bundle sheath and phloem parenchyma into the apoplast via SWEET transporters and reloaded into the phloem via SUTs. The last step in phloem loading is facilitated by SbSUT1, which plays a role in apoplasmic phloem loading in source leaves during the vegetative phase and at anthesis in grain and sweet sorghum cultivars [42]. SbSUT1 shares high sequence identity and homology with the ZmSUT1, which, when mutated to a dysfunctional gene, has impaired phloem loading and shows a chlorotic leaf phenotype consistent with carbohydrate hyperaccumulation in leaves[43].

Carbohydrates are produced in leaf mesophyll cells by photosynthesis, and they are transported to the sink tissue via three phloem loading mechanisms. It is important to note that plants may utilize more than one type of phloem loading mechanism simultaneously, even within a single vein [44]. They may dynamically switch from symplasmic to apoplasmic phloem loading [45]. Plants employ two active and one passive phloem loading mechanisms to transfer sucrose from source to sink. Symplasmic loading is the

passive phloem loading mechanism which relies on diffusion along a sucrose gradient maintained primarily by the sink tissue. The photoassimilates directly move between different cell types through the plasmodesmata connections between adjacent cells. Polymer trapping and apoplasmic loading are the two active phloem loading mechanisms. Polymer trapping has not been reported in monocots [44].

The phloem unloading pathway in sweet sorghums and other Andropogoneae grasses is mainly derived by extrapolating studies done in sugarcane, which like sweet sorghum, also stores sucrose high concentrations in its stem[46]. To better understand phloem unloading strategies in sweet sorghums, two pivotal studies have recently looked at phloem unloading in sorghum stems using a combination of symplasmic and apoplasmic dye tracers in sweet sorghums Rio and Wray[34, 40, 47]. These studies correlated the histochemical changes in developing sweet sorghum stems by staining for lignification and suberisation, creating apoplasmic barriers impermeable to sucrose. Depending on the developmental stage and maturity of sweet sorghum stem tissue, sucrose synthesized in the source leaves is translocated via the phloem to the sink tissue either via symplasmic or a mix of both symplasmic and apoplasmic phloem loading mechanisms [44, 47]. A model illustrating apoplasmic and symplasmic sucrose phloem unloading to stem storage parenchyma and showing roles of various sucrose transport proteins (SUTs and SWEETs) at different developmental stages in sweet sorghum stems has been proposed by Milne et al.

The current phloem unloading model [48] supported by tracer studies claims that sucrose unloading can occur via an apoplasmic pathway in meristematic and elongating zones of sorghum internodes. As sorghum internodes transition from an elongating to a mature sorghum internode, phloem unloading switches to symplastic unloading due to the formation of an apoplasmic barrier around the sclerenchyma sheath, which inhibits apoplasmic phloem unloading. However, depending on the genotype the extent of lignification and suberisation in the sclerenchyma sheath and pith parenchyma cells at maturation, apoplasmic unloading can still occur [34], e.g., sorghum cultivar Wray is symplasmically isolated at maturation and has low levels of lignification when compared to cultivar Rio which might help explain why it can transport sucrose to storage parenchyma cells through a radial transfer with the involvement of an apoplasmic step[48].

Role of sucrose metabolism genes and sucrose transporters

Differential expression of genes involved in sucrose metabolism genes or SUT transporters can account for differences in sucrose accumulation between the grain and sweet sorghums. Other potential candidates such as Tonoplast Sucrose Transporters (TSTs), SbSWEETs, Vacuolar Invertases (VINs) were examined to understand their role in carbon partitioning in sorghum [40, 49].

SWEETs are categorized as sucrose efflux uniporters that move sugars across the cell membrane down a concentration gradient, but the specific functions of individual SWEET genes vary. Through transcriptome profiling, it was found that clade III SWEETs are expressed at different times and in different tissues throughout development in sorghum [50]. SWEETs are responsible for export of sucrose from leaf phloem parenchyma into the apoplast. Sucrose in the apoplast is transported via SUTs into the phloem. Some SWEET members may also be involved in sucrose's movement from the phloem into the stem storage sink or export from the stem. A comparative transcriptomics study of sweet sorghum cultivar Rio and grain sorghum BTx623 revealed that two SbSWEETs, *SWEET3–3*, and *SWEET8–2* were absent from the Rio genome. The authors hypothesized that the lack of these SWEET transporters might contribute to higher sucrose accumulation in Rio internodes by reducing sucrose flux from stems to developing grains[9].

Tonoplast Sucrose Transporters (TSTs), also known as Tonoplast Monosaccharide Transporters (TMTs) initially identified and characterized in arabidopsis, are a class of transport proteins which are localized on the tonoplast (vacuolar) membrane where they function as sucrose antiporters to import sucrose into the vacuole [51]. In a study comparing differential expression of the three *SbTSTs*, it was elucidated that in mature leaves, *SbTST1* and *SbTST2* were ~3.5-fold and ~7.4-fold more highly expressed in sweet sorghum when compared to grain sorghum in stem tissues at anthesis [40]. However, in their *SbSUT* study, they got similar fold differences in SUT expression in sweet and grain sorghum (Macia and Wray). Still, they have claimed that the fold

difference is not high enough to account for a 24-fold increase in sucrose content in sweet sorghum stem. In another study comparing the *SbSUT* transcript levels, Milne et al. also showed a similar differential expression in BTx623 and Rio[48].

Vacuolar Invertases (VINs), also known as soluble acid invertase (SAI), hydrolyze sucrose into the hexoses, glucose, and fructose in the vacuole, thereby changing the ratios of these sugars. Soluble acid invertases play an essential role in sucrose accumulation in the sweet sorghum stem, and they are essential candidates for marker-assisted selection in breeding programs[52]. Two vacuolar invertases, *SbVIN1* (Sobic. 004G004800) and *SbVIN2* (Sobic. 006G160700), are expressed in sorghum and associated with stem and grain traits. The expression of *SbVIN2* in the stem is negatively correlated with stem sucrose content[52, 53]. A study comparing the expression of VINs in grain sorghum BTx623 and a sweet sorghum E-Tian determined that variation in *SbVIN1* was associated with grain traits such as size and weight. *SbVIN1* and *SbVIN2* were not associated with sucrose and glucose changes, but *SbVIN1* was associated with Brix, and *SbVIN2* was associated with water-soluble carbohydrates (WSCs)[49].

Sorghum transgenics

Sorghum research has lagged behind other cereal crops due to its recalcitrance to transformation and regeneration [54]. Significant progress has been made in sorghum transgenic technology in the last five years, and transformation efficiency has increased to ~30% using recently developed protocols [55]. A robust sorghum genetic

transformation system has been established that can be used in a broad spectrum of sorghum genotypes [27, 55, 56]. Progress has been made in whole-genome sequencing of sorghum, and a fully annotated genome and a transcriptome atlas are now available[5, 8]. Targeted genome editing using the CRISPR/Cas9 system has also been demonstrated in sorghum and successfully applied to generate stable genome-edited events with targeted modifications [27, 54].

Raffinose family of oligosaccharides in sorghum

Cereal crop yield is reduced by many abiotic stresses, including drought, salinity, and nutrient deficiency that negatively impact plant growth and development, in some instances, dramatically reducing yield and quality [57]. Drought stress affects cereal crops, especially during the reproductive stage, which leads to embryo abortion and yield loss [58]. Plants respond to these adverse conditions by activating complex adaptive mechanisms through differential gene expression, metabolism changes, and morphological adaptations. One of the many tools plants employ to reduce the harmful effects of dehydration and salinity is by synthesizing compatible solutes. These include amino acids, quaternary compounds, amines, and sugars such as trehalose and the raffinose family oligosaccharides (RFOs)[59].

The RFOs such as galactinol and raffinose are water-soluble carbohydrates that accumulate in seeds in response to abiotic stresses in plants [57]. RFOS are α-1, 6-

galactosyl extensions of sucrose and are present in various plant species. They play different roles in abiotic stress tolerance, carbohydrate transport in the phloem, and carbohydrate stores in the seed [60]. Several studies have found that plants accumulated these compatible solutes to enhance abiotic stress tolerance in plants. Arabidopsis plants overexpressing *AtGolS2* showed an increased accumulation of galactinol and raffinose, making the plants more tolerant to oxidative stress and drought stress [61]. Increased raffinose accumulation in rice caused by overexpressing *OsWRKY11*, an abiotic stress-induced transcription factor, showed significant desiccation tolerance [62, 63]. Transgenic poplar overexpressing *AtGolS2* showed higher tolerance to osmotic and salt stresses than non-transgenic plants [64]. Therefore, the RFO biosynthetic genes could improve abiotic stress tolerance in plant species through molecular breeding and genetic engineering. In some species, RFOs also play an essential role in sucrose translocation, and impairment in the RFO biosynthetic genes inhibits phloem loading [65]. However, their role in phloem transport in monocots has largely remained unexplored.

Sucrose serves as the primary sugar that's translocated via the phloem. In addition to sucrose, RFOs, including raffinose and stachyose, are used in some plant species such as cucumbers and melons for long-distance sugar transport.

RFO biosynthesis begins with the formation of galactinol, which is catalyzed by galactinol synthase, a GT8 family glycosyltransferase, using UDP-galactose and myo-inositol as substrates. Then, raffinose and stachyose are synthesized by stepwise addition of

galactosyl units catalyzed by raffinose synthase (RS) and stachyose synthase (STS), respectively. RS transfers a galactosyl unit from galactinol to sucrose and produces raffinose. Then, STS further uses galactinol as a galactosyl donor to raffinose, producing stachyose. These two reactions are reversible (Figure 1.3). RFOs can accumulate in seeds, protecting the embryo against desiccation during seed maturation, and thus play a critical role in prolonging seed longevity[59].

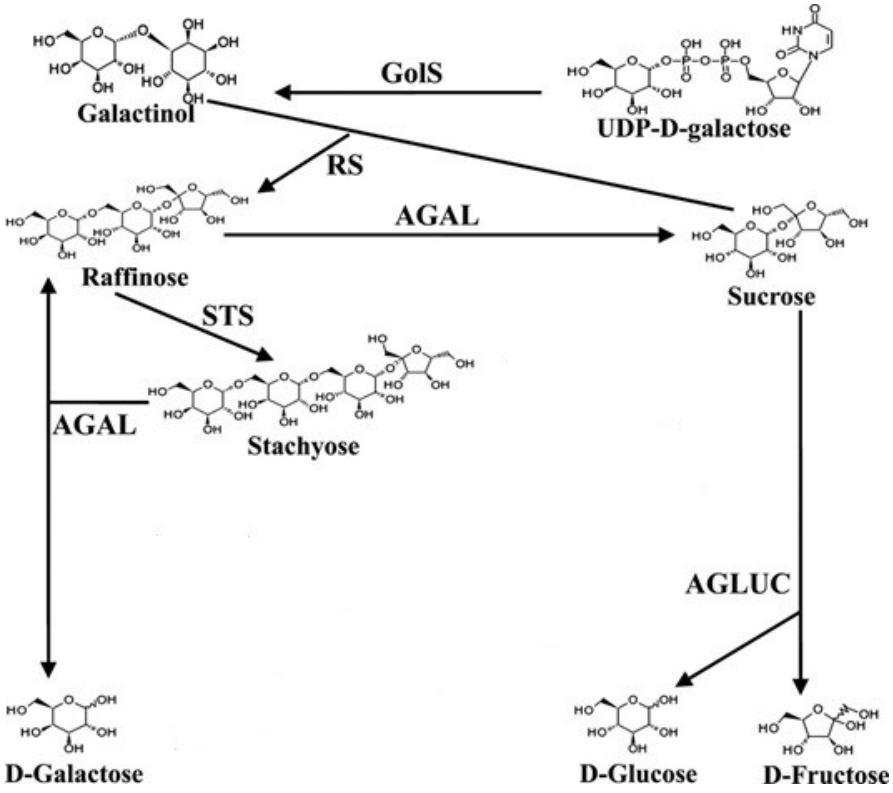

Figure 1.3: Proposed pathway of raffinose metabolism in *sorghum bicolor.* RS- Raffinose Synthase, AGAL- Alpha Galactosidase, STS- Stachyose Synthase, AGLUC- Alpha-Glucosidase.

RFOs increase drought tolerance in maize plants through raffinose synthesis or galactinol hydrolysis [66]. ZmDREB1A and ZmDREB2A - the C-repeat-binding factors (CBF)-which bind to the dehydration responsive element (DRE) in the promoter of a gene and regulate gene expression in response to plant abiotic stress - directly regulate the only raffinose synthase ZmRS in maize controlling chilling tolerance and raffinose accumulation and seed aging respectively [67, 68]. A heat shock responsive element (HSE) AGAAACTTCC (-287) was identified and characterized in Zea mays GolS2 to be independently heat-activated. The same element is found in *SbGolS1* promoter at a similar distance (-277) from the start site, which might provide insight into its regulatory control in sorghum. Another motif called dehydration response binding element (DRE) is also present at the same location in the *SbGolS1* promoter. ZmDREB2A binds to the DRE motif to up-regulate its expression under drought stress[69]. Arabidopsis galactinol synthase 1 (AtGOLS1) negatively regulates seed germination [70]. Galactinol and raffinose protect plants from oxidative damage by scavenging hydroxyl radicals [61]. Galactinol synthase AtGolS2 from Arabidopsis thaliana, when overexpressed in the C4 grass *Brachypodium distachyon*, made it more drought resistant [71].

In this study I wish to identify the RFO metabolism genes in the sorghum genome and explore their expression and potential role in sucrose accumulation in sweet sorghum stem post anthesis.

2. CHAPTER II

IDENTIFICATION OF RFO METABOLISM GENES IN *SORGHUM BICOLOR*

Sequence retrieval of RFO gene sequences

Sorghum genes involved in RFO metabolism were identified by their homology with genes identified in other plant species that encode proteins with validated activity and using phylogenetic analysis. Galactinol synthase (*GolS*) is a member of the Glycosyltransferase family 8. *GolS* encodes a protein containing the protein family domain Pfam PF01501. Thirty-four genes containing PF01501 were identified in the sorghum genome[1] v3.1.1(https://phytozome.jgi.doe.gov) (Supplementary Table S1). To identify sorghum homologs encoding galactinol synthase (*SbGolS*), the amino acid sequences of the 34 sorghum genes containing PF01501 were aligned to GolS proteins from other species using MUSCLE multiple alignment tool [72]. A phylogenetic tree was constructed using MEGA v7.0 [73] using the neighbor-joining (NJ) method with bootstrap values set at 100 replicates (Supplementary Figure S2.1).

Phylogenetic analysis clustered two sorghum genes designated *SbGolS1* and *SbGolS2* with reference *GolS* genes, including three maize genes that encode galactinol synthase (GRMZM2G165919, GRMZM5G872256, GRMZM2G361984)[74, 75].

SbGolS proteins were similar in molecular weight (38.33–39.25 kDa) and amino acid length (338–349 amino acids) to reference GolS proteins, with a slightly acidic pI of 5.55-5.80 (Table 2.1). These values are similar to what has previously been reported for brachypodium and tomato [76], *Camellia sinensis* [77], and *Ammopiptanthus nanus* [78].

SbGolS1 and SbGolS*2* also contain a putative serine phosphorylation site at amino acid position 270 (Supplementary Figure S2.2) and a C-terminal hydrophobic pentapeptide (APSAA) that is a common feature of galactinol synthases (Supplementary Figure S2.3)[79] [80].

Table 2.1 :Raffinose metabolism genes in *Sorghum bicolor*

Gene Name	Phytozome Sequence ID	Arabidopsis Orthologue	Chromosome Location	Number of Transcripts	Length (aa)	MW (kDa)	pI	Predicted sub-cellular location
SbGolS1	Sobic.002G423600	AT2G47180	Chr02-77085729-77087411	1	338	38.33	5.80	Cytoplasmic
SbGolS2	Sobic.001G391300	AT1G56600	Chr01-67764680-67766458	1	349	39.25	5.55	Mitochondrial/Cytoplasmic
SbRS	Sobic.003G052300	AT5G40390	Chr03:4721513..4725477	1	792	85.97	5.8	Chloroplast/Cytoplasmic
SbAGA1	Sobic.002G075800	AT3G57520	Chr02-7874437-7878146	1	764	82.33	6.03	Chloroplast
SbAGA2	Sobic.001G044800	AT1G55740	Chr01-3325484-3330705	2	773	83.72	5.85	Cytoplasmic/Chloroplast
SbAGA3	Sobic.007G219900	AT1G55740	Chr07-64802754-64807148	1	754	82.35	6.10	Cytoplasmic
SbAGA4	Sobic.010G057300	AT5G20250	Chr10-4474788-4478602	2	857	91.21	6.51	Chloroplast
SbAGA5	Sobic.010G057400	AT5G20250	Chr10-4487128-4496678	1	738	78.63	6.07	Chloroplast
SbAGA6	Sobic.006G122400	AT3G57520	Chr06-48852708-48856765	1	810	89.29	5.86	Chloroplast/Cytoplasmic
SbSTS	Sobic.005G210100	AT4G01970	Chr05-69689571-69693116	1	904	97.73	5.56	Chloroplast

Multiple sequence alignments and phylogenetic analysis

Raffinose synthase, stachyose synthase, and alpha galactosidases are members of the glycoside hydrolase family 36 that hydrolyze RFO sugars. Proteins in this family contain the Pfam domain PF05691. Nine sorghum genes were identified in the sorghum genome[1] v3.1.1 that encode proteins containing PF05691 (Supplementary Table S2.2). Amino acid sequences of RS, STS, and AGAs from arabidopsis, tomato, maize, melon, rice, and other species were aligned using MUSCLE. A phylogenetic tree was then constructed using MEGA v7.0 [73] using the neighbor-joining (NJ) method with bootstrap values set at 100 replications (Sequence accession numbers in Supplementary information S1) were aligned using MUSCLE [72]. Phylogenetic relationships and functional divergence among RS, STS, and AGA gene family members in sorghum and other plant species were analyzed using multiple sequence alignment of 33 RS, STS, and AGA protein sequences (Figure 2.1). The raffinose synthase from other plant species used to create the phylogenetic tree are available in supplemementray information S2.1.

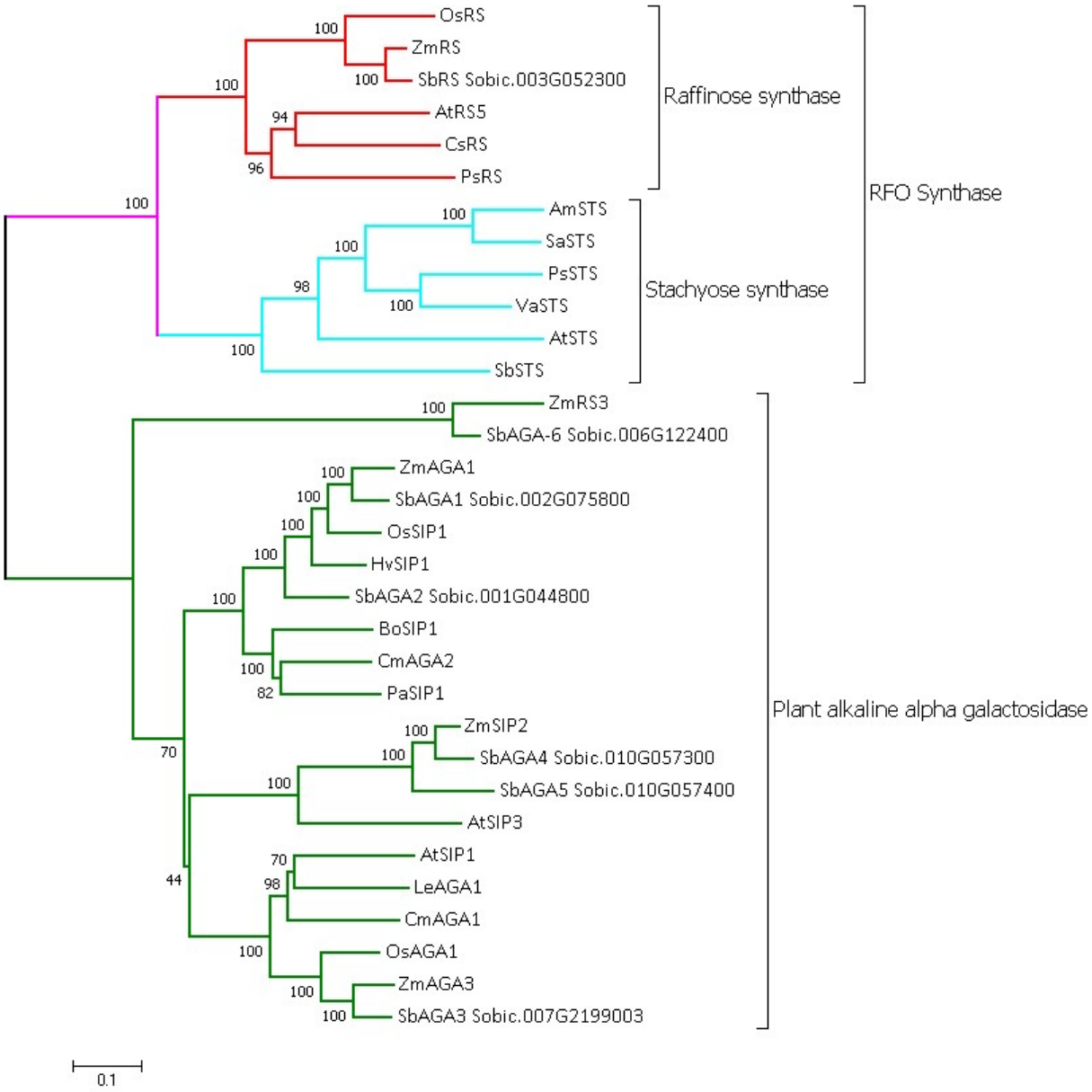

Figure 2.1: Phylogenetic relationships of RFO-synthetic- and selected RFO-hydrolytic enzymes from various taxa. (Names in Supplementary Information S1)

The *SbRS* and *SbSTS* genes clustered with genes from other species previously validated to encode raffinose synthases and the stachyose synthases [81]. The sorghum alpha

galactosidases were named based on their sequence similarity to maize homologs[81]. Neighbor-joining (NJ) bootstrap values are shown on the respective branches. The length of the branches indicates the extent of variation across amino acid positions in protein alignments. Protein sequence alignment showed that SbRS shared 59%, 62%, 59%, 76% and 94% identity with AtRS5, CsRS, PsRS, OsRS and ZmRS respectively (Figure 2.2).

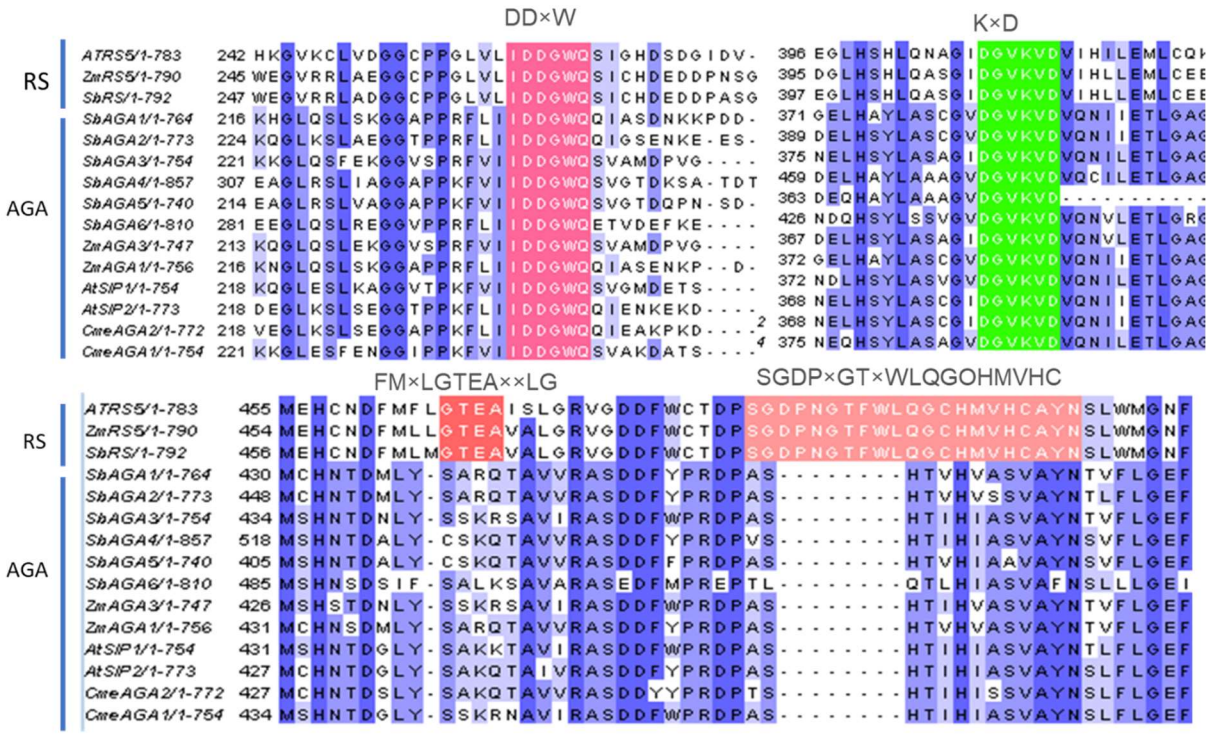

Figure 2.2: Selected regions of the protein sequence alignment among raffinose synthases (RS) and alpha galactosidases (AGA).

Two conserved protein motifs DD×W and K×D were previously identified in alpha galactosidases and raffinose synthases [81]. These motifs are present in SbRS and the 6 SbAGAs identified in this study (Figure 2.2, top). Prior studies also identified two conserved motifs (FM×LGTEA××LG and SGDP×GT×WLQGOHMVHC) specific to

raffinose synthases [82]. The sorghum raffinose synthase (SbRS) identified in this study contains the raffinose synthase specific protein motifs (Figure 2.2, lower). Although both RafS and StaS contain PF05691, a sequence block of about 80 amino acid length characteristic of stachyose synthases is present in SbStaS but not SbRS (Figure 2.3) [83].

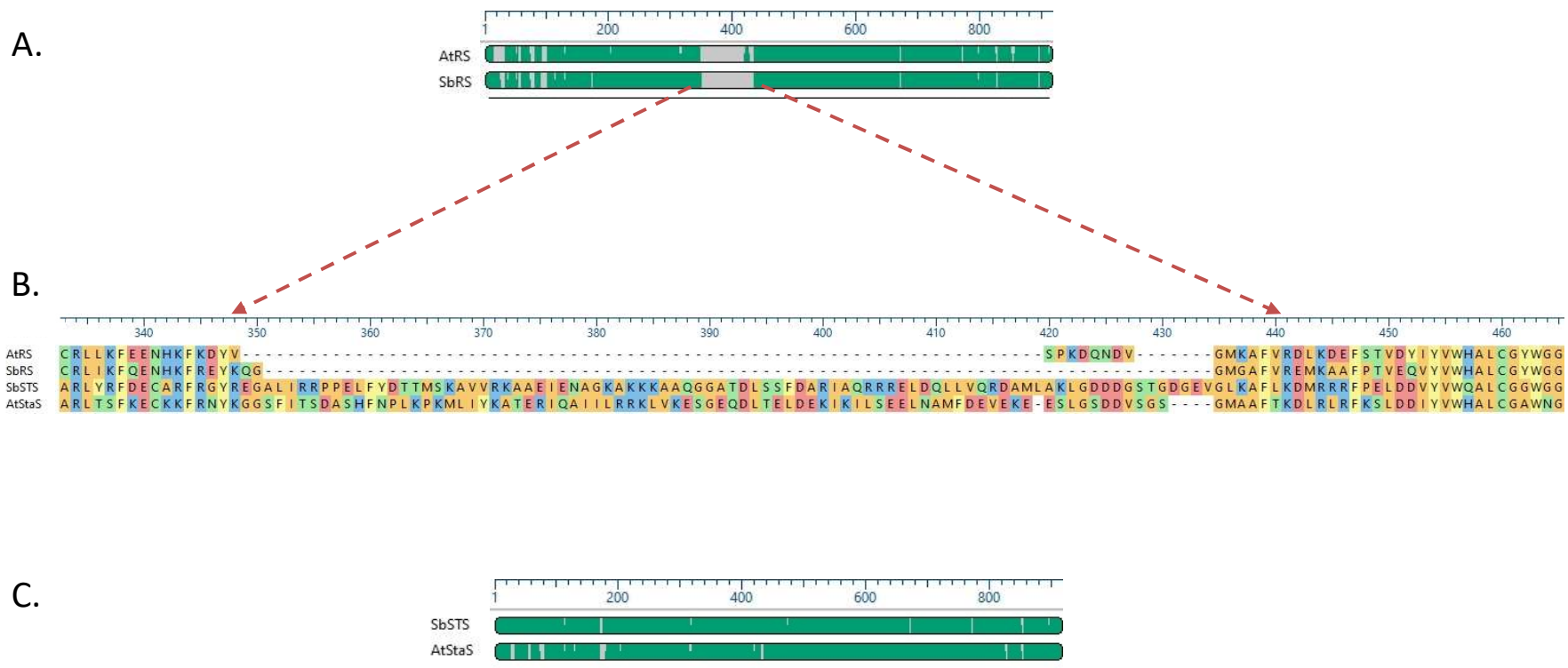

Figure 2.3: Sequence and phylogenetic analysis of SbRS, SbSTS, AtRS, and AtRSTS. **A.** Schematically shows the 783 aa long sequence of AtRS5 (A5g40390) from *A. thaliana* and the 792 aa long sequence of SbRS (Sobic.003G052300) from *S. bicolor*. **B.** Shows a section of a sequence alignment performed with Clustal Omega of RafS and StaS amino acid sequences, which revealed the very high amino acid identity and similarity, except for an 80 amino acid long sequence block insertion **C.** Schematically shows the 876 aa long sequence of AtSTS (At4g01970) from *A. thaliana* and the 904 aa long sequence of SbRS (Sobic.003G052300) from *S. bicolor*. Dashed orange arrows point to the ~80 aa gap in alignment in the raffinose synthases.

Physicochemical properties and conserved motif analysis of RFO metabolism proteins

Sorghum RFO pathway gene information, including accession number, chromosome location, coding sequence (CDS) length, mRNA length, and encoded protein length, was downloaded from Phytozome and summarized in Table 1.1. Predicted protein characteristics such as isoelectric point (pI) and molecular weight (MW) of encoded proteins were calculated using the ProtParam (ProteinParameters) tool (web.expasy.org/protparam). The GOLS, RS, STS, and AGA protein sequences were analyzed with Pfam (http://www.ebi.ac.uk/Tools/pfa/pfamscan/) to confirm the presence of the expected Pfam domains PF01501 (Glyco_transf_8) and PF05691 (Raffinose_syn). The two GOLS proteins had a common Pfam domain, PF01501, a conserved domain of the glycosyltransferase 8 family. The SbRS, SbSTS, and the SbAGAs proteins shared a common Pfam domain PF05691, a conserved domain of the glycoside hydrolase family 36.

Conserved motif analysis was performed on GOLS, RS, STS, and AGA proteins by using MEME tool[3] [84] http://meme-suite.org/tools/meme with following parameters; a maximum number of motifs to find, 5 for GolS and 6 for the rest; minimum width of motif, 6 and a maximum width of motif, 50. Motif analysis for GOLS proteins was used to identify the five most conserved motif sequences. The motif sizes ranged from 25 to 50 amino acids in length and are presented in Figure 2.4. Motifs 1, 2, and 4 of the GOLS proteins were related to Glyco_transf_8 (PF01501) domain structure, motif 3, and 5 and were not related to any domain structure (Supplementary Table 3). A similar motif organization indicates an evolutionary relationship, and motifs 3 and 5 might have different functions.

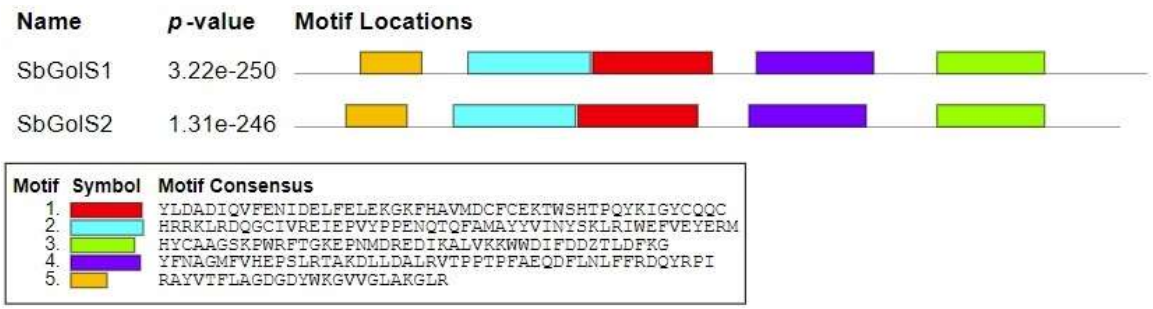

Figure 2.4**:** Block diagram of conserved motifs in GolS proteins identified in S*orghum bicolor*. Each motif is represented with a box in different colors: motif 1, red; motif 2, cyan; motif 3, green; motif 4, purple; and motif 5, yellow.

MEME motif analysis for RS, STS, and AGA proteins was done to identify the most conserved six motif sequences. The motif sizes ranged from 40 to 50 amino acids in length and are presented in Figure 2.5. All motifs of the RS, STS, and AGAs proteins were related to Raffinose_syn (PF05691) domain structure (Supplementary Table 4). SbRS, SbSTS, and AGAs in the same groups have similar motifs, indicating the link between evolutionary and conserved motifs.

Figure 2.5**:** Block diagram of conserved motifs in RS, STS, and AGAs proteins identified in *Sorghum bicolor*. Each motif is represented with a box in different color: motif 1, red; motif 2, cyan; motif 3, light green; motif 4, purple; and motif 5, yellow; and motif 6, green

Two different approaches were used to predict the subcellular localization of the RFO pathway proteins. The first online tool, called CELLO v.2.5[4] (subCELlular LOcalization predictor) http://cello.life.nctu.edu.tw/, uses a homology-based approach to predict subcellular localization (Table 2.1). This analysis predicts that SbGOLS1 is localized to the cytoplasm (for GOLS2, possibly the mitochondria). However, SbRS could be located either in the cytoplasm or the chloroplast. Most of the SbAGA proteins involved in raffinose turnover had predicted localization in chloroplasts, although two could also be located in the cytoplasm. A subcellular localization prediction tool called DeepLoc[5,] which uses deep neural networks relying only on sequence information, was also used to predict protein localization (Supplementary Table 4). Subcellular locations were also predicted by TargetP 1.1[6] server (http://www.cbs.dtu.dk/services/TargetP/), and it strongly predicted that SbAGA4, 5, and 6 were localized in the chloroplast.

Potential sites of post-translational modification in RFO proteins were analyzed using the online NetNGlyc 1.0 Server (http://www.cbs.dtu.dk/services/NetNGlyc/) for N-glycosylation, YinOYang 1.2 server (http://www.cbs.dtu.dk/services/YinOYang/) for O-glycosylation, and NetPhos 3.1 Server (http://www.cbs.dtu.dk/services/NetPhos/) for phosphorylation. All the predictions were made at the default threshold settings. N- or O-glycosylation and phosphorylation sites were predicted to be present in all RFO metabolism proteins.

Heterologous expression and *in vitro* biochemical characterization of SbAGA1 and SbAGA2

Material and methods

After cloning the full-length coding sequence of the SbAGA1 and SbAGA2, to confirm the cloned genes encoded alkaline α-galactosidases, the genes were heterologously expressed in *E. coli* strain BL21. Bacteria transformed with each of the SbAGA genes expressed a protein of the expected MW. AGA1 protein's MW was slightly lower than that of AGA2, and both were approximately 82 kilodaltons, the approximate size of the purified AGA enzyme.

cDNA Isolation and Sequences Analyses of SbAGA1 and SbAGA2

cDNAs of *SbAGA1* and *SbAGA2* were cloned using reverse-transcription polymerase chain reaction (RT- PCR). Total RNA was isolated from the leaves of field-grown Wray sorghum plants. Leaves were harvested and immediately frozen in liquid nitrogen and stored at -80 ˚C. The leaf tissue was then ground in a mortar and pestle and RNA extracted using a Tri-Reagent based method with a Direct-zol RNA Miniprep Plus Kit (Zymoresearch, USA). Five µg of total RNA was used to synthesize single-stranded cDNA using the SuperScript IV kit (Invitrogen, USA) using an oligo dT based method. To prepare recombinant SbAGA1 and SbAGA2 proteins in native form for biochemical characterization, a *SbAGA1* and *SbAGA2* cDNA fragment harboring the full-length coding sequence was PCR amplified with primers

SbAGA1-F: 5'-TTTTGTTTAACTTTAAGAAGGAGATATACATGCACCACCACCACCACCACATGAC-3', SbAGA1-R: 5'-ATGATGGCTGCTGCCTTCACTGGTACCGAGCTCGT-3',

And

SbAGA2-F: 5'ATTTTGTTTAACTTTAAGAAGGAGATATACATGCACCACCACCACCACCACAT-3', SbAGA2-R:5'-ATGATGATGGCTGCTGCCTGGTGTTGTCTTCACTGGTA-3' containing a 6-His sequence on each end. PCR was performed with Q5 High-Fidelity DNA polymerase (New England Biolabs, USA). The PCR product was gel purified and cloned in-frame into pET28b vector (Millipore Sigma, USA) digested with restriction enzyme NcoI (NEB, USA) enzyme through a Hifi assembly kit (NEB, USA). The plasmid maps for vector pET-28 b (+) containing the in-frame his-tagged SbAGA1 and SbAGA2 coding sequences are available in supplementary figure S2.2. After transformation into DH5α, the plasmids were isolated using a miniprep kit and sequenced to confirm sequence fidelity through Sanger sequencing. The sequencing primers used are listed in Table S2.6.

Heterologous Expression of SbAGA1 and SbAGA2

For heterologous expression of recombinant SbAGA1 and SbAGA2 in *E. coli* cells, the expression construct was introduced into *E. coli* XJb (DE3) Autolysis™ (Zymoresearch, USA) using standard *E. coli* transformation procedures. 1L of LB broth supplemented with 50µg/L Kanamycin (Sigma Aldrich, USA) was inoculated with 1 ml of a 2ml saturated culture grown overnight at 37˚C and

grown in a 4L flask with vigorous shaking (250 rpm). After the culture reached an OD600 of 0.6-0.7, it was chilled on ice for 30 minutes and induced by the addition of 1mM Isopropyl-β-D-thiogalactopyranoside (IPTG) and further grown for ~24 hours with vigorous shaking (~250 RPM) at 16°C. A pre-induction control and a 6-hour post-induction control were run on an SDS-PAGE gel. The induced protein can be seen as a thick band against the corresponding uninduced samples in supplementary figure S2.5. Cells were again chilled on ice for 30 min before harvesting. Subsequent purification steps were carried out at room temperature.

The culture was harvested by centrifugation at 5000g for 10 minutes, and the pellet was stored at -80°C until further use. For purification of recombinant SbAGA1 and SbAGA2, 1 gram of wet pellet weight was thawed and resuspended in 5 ml of Lysis buffer (Buffer recipes in the supplementary material) supplemented with a protease inhibitor cocktail (Sigma Aldrich cat# P8849) and Benzonase® nuclease (Sigma Aldrich cat# E1014). The solution was gently shaken for 30 minutes on ice. After lysis, insoluble materials were removed by centrifugation for 20 min at 20,000 g, and the lysate was filtered through a 0.2µm filter and used for Immobilized Metal Affinity Chromatography (IMAC).

For purification of his-tagged recombinant SbAGA1 and SbAGA2, the clarified supernatant was loaded onto a 1 ml HisPur™ Ni-NTA Spin Column (Thermo Scientific, cat# 88225) according to the manufacturer's instructions. The column was washed three times with 2 ml of equilibration buffer, and bound recombinant protein was eluted twice with 1 ml of elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 1mM DTT, and 250 mM imidazole and a pH of 8.0). The two fractions were pooled, and the buffer exchanged and concentrated with a 10,000 MWCO concentrator

(Thermo Scientific, cat# 88527). Purified recombinant SbAGA1 and was analyzed by SDS-PAGE (Figure 2.6) and used for enzyme assays. SbAGA2 was cloned, expressed, and purified similarly. The SDS-PAGE for SbAGA2 is presented in Supplementary Figure S2.5. The eluates were diluted in McIlvaine buffer pH 7.0 to 4 ml with 50% glycerol and 1mM DTT and stored at -20 ºC.

A soluble and functional protein was obtained in both expression experiments. The recombinant protein can be recognized as a single band of about ~80 kDa in SDS-PAGE (Figure 2.6).

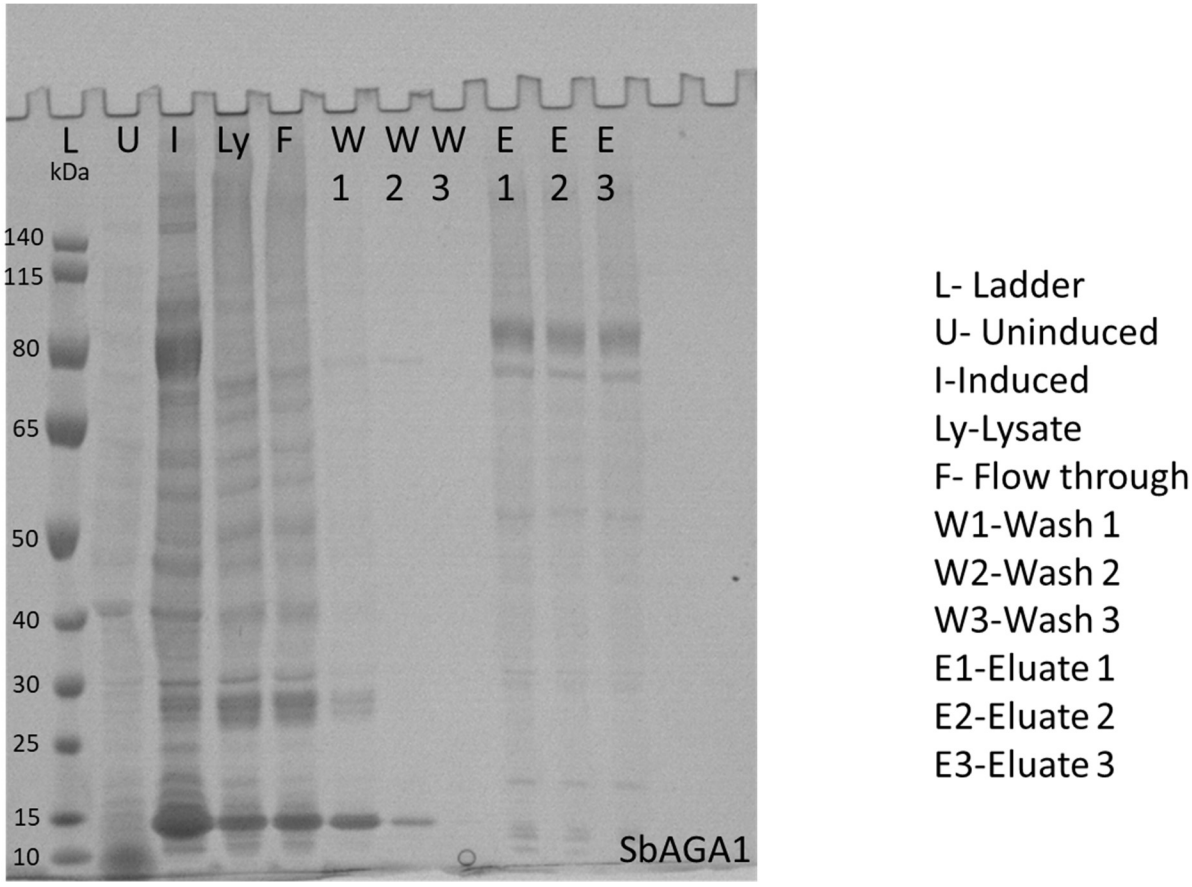

Figure 2.6**:** SDS-PAGE of recombinant SbAGA1 purified using a NiNTA-spin column. Proteins were analyzed using SDS-PAGE collected during purification of AGA1 were obtained following induction (I), lysis (L), and column washes (W) and elution (E) of AGA1 from the column. Lane (L), molecular mass marker. (Invitrogen, SeeBlue™ Plus2 Pre-stained Protein Standard)

Protein quantification

Protein concentration was measured on the NanoDrop® ND-1000 Spectrophotometer using the method of Pierce™ BCA Protein Assay Kit (Thermo Scientific™ cat# 23225) assay with bovine serum albumin (BSA) as a reference standard, which was provided with the kit. Since the eluate showed additional bands on the SDS-PAGE, I performed a secondary quantification using the gel quantification tool of ImageJ software[85]. 2.17 mg of recombinant SbAGA1 and 4.5 mg were purified and used for assays.

AGA characterization

Alpha galactosidase enzyme activity of recombinant SbAGA1 and SbAGA2 was determined in McIlvaine buffer system at pH 7.0 and different pH values ranging from pH 5.0 to pH 9.0; in 0.5 pH increments at 37°C. The enzyme's ability to hydrolyze a chromogenic substrate 4-nitrophenyl-α-D-galactopyranoside (PNG) into p-nitrophenol and galactose was used to detect enzyme activity [86]. The alpha-galactosidase hydrolysis of PNG is presented in Figure 2.7. Alpha-galactosidase hydrolyzes the chromogenic substrate PNG releasing a yellow-colored product, p-Nitrophenol.

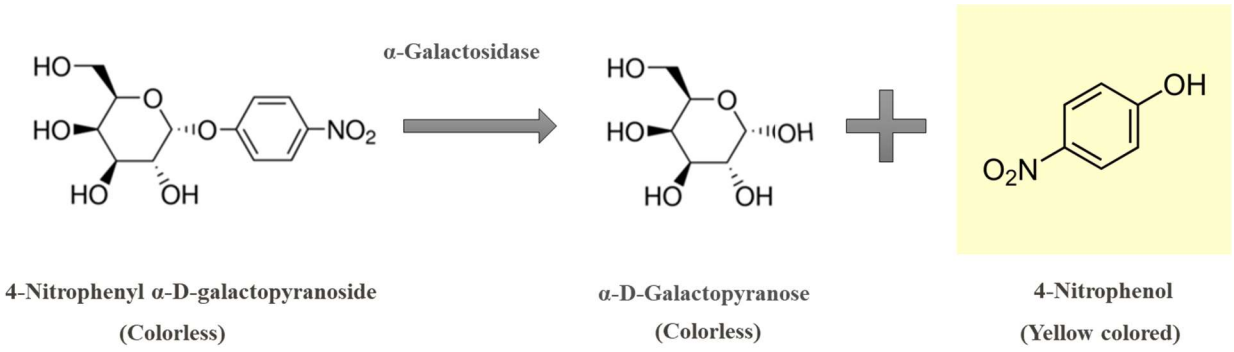

Figure 2.7**:** Alpha-galactosidase hydrolysis of a chromogenic substrate PNG.

The assay for enzyme activity consisted of a 50 µl aliquot of diluted eluate, 50 µl 20 mM PNG (Sigma Aldrich cat# N0877), and 150µl of McIlvaine buffer (pH 7.0). The reaction was incubated for 10 min at 37 °C, stopped by the addition of 1 ml 0.5 M Na_2CO_3, and the amount of product formed determined spectrophotometrically using a plate reader (Tecan Infinite® 200 PRO) at 410 nm. The assay used to determine the pH dependence of recombinant SbAGA1 and SbAGA2 activity consisted of 50 µl aliquot of diluted eluate, 50 µl 20 mM PNG, and 150µl of McIlvaine buffer at different pH values. The reaction was carried out for 10 minutes at 37 ºC in a thermocycler, stopped by the addition of 1 ml 0.5 M Na_2CO_3. The pH optimum is presented in Figure 2.10.

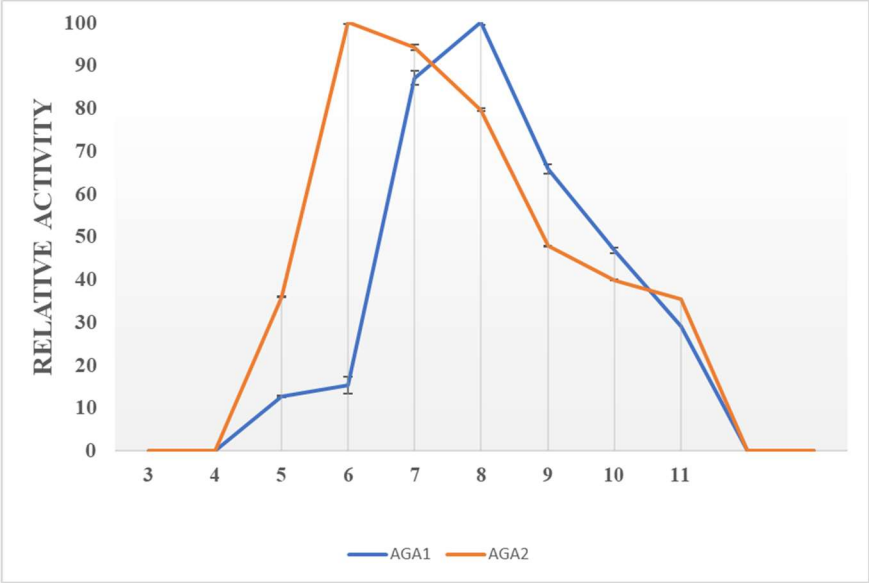

Figure 2.8**:** The effect of different pH values on the alpha-galactosidase enzyme activity of recombinant SbAGA1 and SbAGA2. The activity was measured with 20mM PNG, according to Gao and Schaffer[86].

Incubation of recombinant SbAGA1 and SbAGA2 in buffers with different pH showed that the optimum pH of SbAGA1 is 8.0 (Figure 2.10), and for SbAGA2, the optimum pH is 6.0.

Substrate specificity of SbAGA1 and SbAGA2

AGA enzyme activity was assayed in 0.2 ml PCR tubes using 50mM raffinose and stachyose concentrations in McIlvaine buffer (pH 7.0) with 5mM DTT as previously described [83]. The reactions were incubated in a thermocycler at 37°C for 12 hours and stopped by boiling the tubes at 95 °C for 10 minutes at the end of the reaction. 10µL of this reaction was used to determine the

amount of D-galactose released, which was measured as an index of the enzyme's hydrolysis activity. The amount of enzyme activity using raffinose or stachyose as a substrate was determined by indirectly measuring the D-galactose released from each RFO sugar using a commercially available L-Arabinose/D-Galactose Assay Kit (Megazyme, Ireland, Cat# K-ARGA).

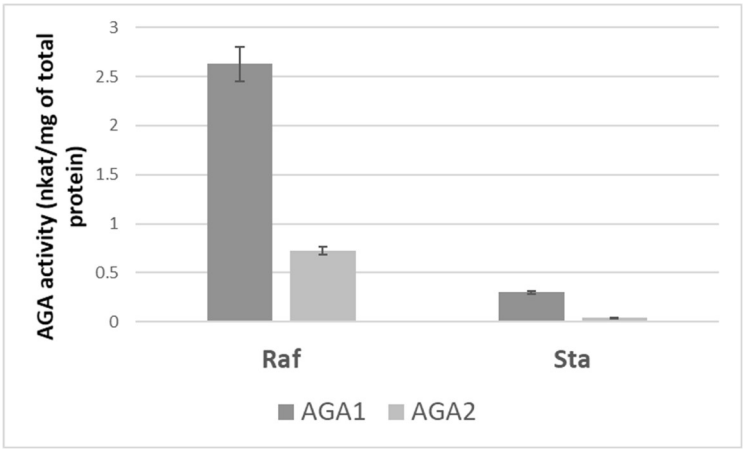

Figure 2.9**:** AGA1 shows greater activity using the substrate raffinose compared to stachyose (measured at pH 7.0 with 50 mM each of raffinose and stachyose). Data are means ± SE of 4 replicates.

Recombinant SbAGA1 and SbAGA2 were found to have a higher activity using raffinose than stachyose as substrate (Figure 2.9).

Results and discussion

In this study, 2 SbGols, 1 SbRS, 1 SbSTS, and 6 SbAGAs were identified in the sorghum genome. SbAGA1 and SbAGA2 were heterologously expressed in *E. coli,* and their enzyme activity was verified following purification. Both enzymes had alpha-galactosidase activity based on hydrolysis of a chromogenic substrate, are able to hydrolyze raffinose and stachyose. On a 4 to 12% gradient Bis-Tris gel (ThermoFisher, cat# NP0322PK2), the apparent molecular mass of the recombinant SbAGA1 and SbAGA2 was ~ 80 kDa, which is close to their estimated mass of ~82 kDa deduced from amino acid sequence data. SbSTS differs from SbRS by ~80 amino acids, which is a a signature of stachyose synthases[83]. In this study, I purified and chemically characterized for the first time two functional recombinant AGAs from *S. bicolor*.

3. CHAPTER III

CHARACTERIZATION OF THE EXPRESSION OF RFO-GENES DURING SORGHUM DEVELOPMENT

Introduction

High biomass yield, low input requirements, and environmental resilience that enables production on marginal lands have made bioenergy sorghum a leading biofuel crop. Unlike corn and soybeans, sweet sorghum can be grown on land not optimal for food crops, and stems accumulate fermentable sugars that are easily extractable in the form of juice[28, 29, 87, 88]. Sweet sorghum is known to accumulate juice that accounts for 78% of its fresh weight biomass. Therefore, high juiciness is an important phenotypic trait selected for in sweet sorghum breeding. The sugar concentration in sorghum is often measured in Brix, where one-degree Brix is equal to 1 g of sugar per 100 g of juice. The sugar concentration in sweet sorghum is estimated to range from 14 to 23%. Even though sweet sorghum sugar yields are lower than those of sugarcane, further improvements are possible through sweet sorghum breeding. A complete understanding of the molecular basis of sucrose accumulation in sweet sorghum stems may help maximize the accumulation of sugars that can be efficiently and economically converted to biofuels. Prior studies have mostly focused on identifying enzymes involved in sucrose metabolism that may play a role in sucrose accumulation post-anthesis in sweet sorghum stems.

Sucrose concentration in the sorghum stem is affected by the rate of transport of sucrose to the stem, the rate of sucrose hydrolysis by invertases (INV) and sucrose synthases (SS), and the rate of sucrose synthesis in stems by sucrose phosphate synthase (SPS). Sucrose synthase (SUC1)

hydrolyzes sucrose producing UDP-glucose, which is utilized as a substrate for cell wall biosynthesis during the vegetative phase of sorghum development. *SUC1* expression in stems is reduced ~10 fold when sweet sorghum reaches anthesis, reducing sucrose utilization in stem internodes post-anthesis [39]. A comparison of stem sugar accumulation and enzyme activities in a sweet sorghum cultivar (SSV74) and a grain sorghum variety (SPV1616) showed that sucrose synthase, sucrose phosphate synthase, and invertase (cytoplasmic and vacuolar) did not contribute significantly to sucrose accumulation differences between the two genotypes [89]. They found that the expression of sucrose synthase (SUC1), two sucrose phosphate synthases (SPS2 and SPS3), and vacuolar invertase (INV3) were lower in sweet sorghum as compared to grain sorghum. This study was pivotal because enzyme activities often cannot be directly correlated with gene expression due to differing protein turnover rates, post-translational modifications such as phosphorylation that affect enzyme activity and oligomerization of gene products [90]. Also, two sorghum vacuolar invertases SbVIN1 and SbVIN2, are differently associated with stem and grain traits in sorghum, but no significant association between the genes and sucrose and glucose accumulation was noted[49].

On the other hand, *SbVIN1* expression in stems of the sweet sorghum Della decreased >10-fold post floral initiation parallel with stem sucrose accumulation (McKinley et al., 2016). SbVIN1 was identified as vacuolar invertase, the compartment where sucrose accumulates in sweet sorghum and sugarcane. Moreover, decreases in invertase in sweet sorghum stems occurred before or in parallel with stem sucrose[39]. Taken together, decreases in SUT1 and VIN1 reduce sucrose turnover, contributing to sucrose accumulation in sweet sorghum stems.

Differential expression of sucrose synthesis genes could not account for the large differences in sucrose accumulation in grain and sweet sorghum stems. Other potential candidates involved in stem sugar accumulation include sugar transporters such as SUTs, Tonoplast Sucrose Transporters (TSTs), and SbSWEETs, were examined to understand their role in carbon partitioning in sorghum [40, 49]. More recent research has focused on identifying sucrose transporters involved in the sequestration of sucrose in stems. A study comparing the sweet sorghum Wray to grain sorghum found that SUTs were expressed at similar levels during stem sucrose accumulation at similar developmental stages [34]. SbSWEETs and SbTSTs were then investigated for their role in sucrose accumulation in Wray stem post-anthesis. The group found that a ~24-fold increase in total stem solutes observed in sweet Sorghum (UNL71-2011) compared with grain sorghum (UNL3016) is correlated with higher expression of SbTST1 and SbTST2 in stem (~2.6- and ~4.4-fold, respectively) [40]. A study also looked at the relationship between plant height and sucrose accumulation and concluded that tall and short stature plants could accumulate high stem sucrose[91].

Since the genetic mechanisms underlying differential sucrose accumulation between sweet and grain sorghums are not fully elucidated, it is critical to examine other pathways that might influence stem sugar accumulation. In the current study, transcriptome analysis of sweet sorghum stem development revealed that genes involved in raffinose metabolism are up-regulated in the stem during the sucrose accumulation phase in the sweet sorghum cultivar Della. In Della, genes encoding *AGAs* that hydrolyze raffinose to sucrose and galactose are induced ~50-fold in sorghum stems post-anthesis during the stem sucrose accumulation phase. Similarly, in another sweet sorghum Rio, the gene Sobic.002G075800 (*SbAGA1*) is expressed at low levels during the

vegetative and up-regulated after floral initiation coincides with the stem sucrose accumulation phase[9]. I propose that raffinose synthesis, transport, and turnover plays a significant role in post-anthesis stem sugar accumulation in sorghum.

Materials and methods

Plant materials and growing conditions

Sorghum genotypes

The sorghum genotypes used in this study included the cultivar Wray [*Sorghum bicolor* (L.) Moench] (Reg. no. 119), sweet sorghum developed from an F2 cross of Brawley, and Rio [92], Rio (Reg. No. 113), sweet sorghum developed from an F2 cross between' Rex' (MN 23) and PI 152959 (MN 1048) [93], and the sweet sorghums Umbrella and Della. Dr. William Rooney provided seeds of the sweet sorghum cultivars Wray, Della, and Umbrella from the Texas A&M Sorghum Breeding Program (College Station, TX).

Field and greenhouse studies

The sweet sorghum genotypes and several other sorghum cultivars were planted at the Texas A&M Research Farm in College Station as part of a diversity panel for developing and designing high-biomass sorghum hybrids. Plot design, fertilization, and crop management have been previously described [39].

For gene expression experiments, Wray and Della were grown in a greenhouse under 14 h long days in 5-gallon pots with greenhouse Metro-Mix 900 (Sungro Horticulture) supplemented with

osmocote and then fertilized biweekly with Peters 20-20-20 general-purpose solution (Peters® Professional). Plants were thinned to one plant per pot and grown at 10-20 cm spacing and watered.

Tissue harvest

Leaf and stem internode tissue sections were collected at 11 am at ten days before anthesis and forty days after anthesis. They were flash-frozen in liquid nitrogen and then kept at -80 °C until RNA extraction. Leaf tissue was prepared for RNA extraction by grinding tissue to a fine powder using a pre-chilled mortar and pestle in liquid nitrogen and storing ground tissue in 50-ml conical tubes at -80 °C until extraction. Internode tissue was first coarsely ground in a coffee grinder and then in a pre-chilled mortar and pestle in liquid nitrogen and stored in 50-ml conical tubes at -80 °C until extraction. The internode tissue was further pulverized in TRI-reagent containing grinding tubes for efficient extraction. Total RNA was then extracted from each sample using the Direct-Zol RNA Miniprep Plus Kit (Zymo Research) according to the manufacturer's instructions using TRI-Reagent and an on-column DNase treatment step. RNA concentration was quantified using a NanoDrop™ One/One^{C} Microvolume UV-Vis Spectrophotometer (Thermo Scientific). 1µg of total RNA was used to synthesize complementary DNA (cDNA) using the SuperScript™ IV First-Strand Synthesis System for qRT-PCR (ThermoFisher Scientific cat# 18091050) according to the kit's instructions.

Primer design and qPCR

qPCR primers for GolS, RS, and AGAs were designed using PrimerQuest Tool (Integrated DNA Technologies) across splice junctions (Supplementary Table S3.1). Their specificity was tested

by dissociation curve analysis and gel electrophoresis of qRT-PCR products. The primers were designed across exon junctions to amplify individual SbAGA gene family members and generate products between 100–150 bp in size for each SbAGA gene. For each reaction, 10 ng of cDNA was mixed with 5µl PowerUp™ SYBR™ Green Master Mix (ThermoFisher Scientific cat# A25742) and 0.2 µM of both the forward and reverse primers for a final volume of 10 µl. The samples were run on a The CFX384 Touch Real-Time PCR Detection System (Bio-Rad, California). The amplification program was: 95 °C for 30 s followed by 40 cycles of 95 °C for 5 s, 57 °C for 30 s, and a final temperature increment of 0.5 °C for 5 s from 65 °C to 95 °C.

Quantitative reverse-transcription polymerase chain reaction (qRT-PCR) analysis was carried out as described [94]. Relative expression was calculated using the comparative cycle threshold (Ct) method. Raw Ct values for each sample were normalized to Ct values of a previously validated reference gene SbUBC (Sobic.001G526600) [21]. ΔΔCt values were calculated relative to the sample with the highest expression (lowest Ct value). Relative expression values were calculated with the $2^{-\Delta\Delta Ct}$ method.

RNA-seq expression analysis

A sorghum gene expression database comprised of RNA-seq profiles was used to identify RFO pathway genes that were differentially expressed in leaf and stem tissues. RNA-seq data from the grain sorghum BTx623 published in the sorghum RNA Atlas (McCormick et al., 2018) was used to survey variation in expression of the RFO-genes in organs/tissues and during plant development. Transcripts Per Kilobase Million (TPM) values of RFO-genes that are differentially expressed in

various sorghum tissues of BTx623 during different developmental stages are presented in Table 3.1.

RNA *in situ* hybridization (RNA-ish)

Plant tissue for RNA *in situ* hybridization was obtained from sweet sorghum Della grown in the greenhouse in College Station in 2018. Internode and leaf sections were harvested 40 days after grain maturity when the expression of the target RFO genes was expected to be high in Della leaves and stems. Internode sections were harvested from the middle of an internode that was part of a phytomer with a green leaf/leaf sheath. Leaf sections were harvested from the same phytomer. In the case of internodes, 0.5 cm sections were harvested from the middle of the internode, and in the case of leaves, 1 cm cross-sections were harvested from the middle of the length of the leaf blade. Plant samples were placed under a vacuum to infuse 10% formalin fixative buffered saline solution into the tissue for 24 hours at room temperature. Samples were then washed with 1X phosphate-buffered saline and dehydrated through a series of graded ethanol washes, cleared with xylene, and embedded in paraffin wax according to established protocols[95]. Embedded samples were stored and shipped at 4°C to our collaborators at Michigan State University for RNA-ish.

RNA probes were designed and synthesized by Advanced Cell Diagnostics, Inc. (ACD) using their custom probe design service. In general, several double Z probe pairs are produced with high specificity hybridizing to the target RNA. When bound in pairs, it allows for an amplifier signal to be generated, which is then observed as a red chromogenic precipitate dot.

Embedded sorghum material was thinly sectioned (6μm thick) using a Leica Microtome. Sections were placed on Fisherbrand Superfrost Plus slides (Cat# 12-550-15), put on a hot plate at

42°C for a few minutes, and then further dried overnight at room temperature. Slides were processed as described by Wang et al. [96]. Specifically, RNAscope 2.5 HD Detection Reagent – the RED kit was used (Cat# 322360). Slides were sealed with EcoMount (Biocare Medical, Cat# EM897L). Images of plant sections were obtained using a bright-field through a Zeiss Axio Imager.M2 microscope.

Results

Genes involved in RFO metabolism (*GolS, RS, STS, AGAs*) were identified in the sorghum genome based on homology and phylogenetic analysis in chapter II. In this study, the expression of RFO pathway genes was investigated, starting with the analysis of expression in various organs during the development of the grain sorghum BTx623 (Table 3.1).

Table 3.1**:** The expression level of the RFO pathway genes in grain sorghum BTx623 by RNA-seq in various tissues at different developmental stages. The numbers represent the average expression obtained from the TPM (Transcripts Per Kilobase Million) plots. DAE- Days after emergence.

	Juvenile roots	Juvenile Leaf Blade	Leaf at anthesis	Leaf at grain maturity	Internode at anthesis	Internode at grain maturity	Panicle at floral initiation	Dry Seed at grain maturity	Mature seed imbibed for 24 hours
SbGolS1	6	345	199	279	11	27	318	487	15
SbGolS2	0	163	20	33	0	0	6	11	2
SbRS	0	56	19	11	0	0	4	8	0
SbSTS	0	6	1	1	0	0	1	3	2
SbAGA1	0	3	3	6	2	40	661	1466	496
SbAGA2.1	0	3	11	3	1	13	0	0	0
SbAGA2.2	0	2	12	3	3	27	0	3	1
SbAGA3	52	92	3	5	1	3	0	12	38
SbAGA4.1	5	5	6	3	1	0	0	0	1
SbAGA4.2	8	10	5	8	3	3	2	2	11
SbAGA5	1	3	0	1	0	1	0	0	1
SbAGA6	604	68	44	24	148	119	45	7	216
DAE	8	8	65	96	65	96	44	96	96

Transcriptome data from BTx623 were used to characterize the expression of *SbGolS* and *SbRS*, and *SbAGA* genes at different developmental stages and organs (Table 3.1). Galactinol synthase acts on UDP-galactose and myo-inositol to form galactinol, considered the rate-limiting and the committing step in RFO biosynthesis [75]. In BTx623, *SbGolS1* is highly expressed in leaves of juvenile and adult phase plants from the vegetative phase through grain maturity (Table 3.1). *SbGolS1* is also highly expressed in developing panicles and mature grain. *SbGolS2* was highly expressed only in leaf blades of juvenile leaves. *SbRS* is expressed at relatively high levels in leaf

blades of juvenile plants and expressed but at somewhat lower levels in leaves of adult plants through anthesis and grain maturity. *SbSTS* expression was relatively low in most tissues. *SbAGAs* showed a wide range of expression patterns. *SbAGA1* was expressed at the highest level among these genes. *SbAGA1* was highly expressed in developing panicles (shortly after floral initiation), dry seed, and imbibing seed. *SbAGA2* was most highly expressed in internodes at grain maturity. *SbAGA3* was differentially expressed in juvenile tissues, including imbibed seed. *SbAGA4* and *SbAGA5* showed relatively low expression. In contrast, *SbAGA6* was highly expressed in the roots of juvenile plants and in imbibed seed, which typically activates root growth during the imbibition phase. *SbAGA6* also showed relatively high levels of expression in internodes at anthesis.

RFO pathway gene expression in Della stems

In a prior study, RNA-seq data was collected from Della stems from floral initiation through post-grain maturity when stem sucrose and starch increase significantly [39]. Analysis of RFO-gene expression during this phase of development showed that *SbRS* and *SbSTS* expression was consistently low in stems, but *SbGolS1* expression increased 20-fold between anthesis and post-grain maturity (Table 3.2). Similarly, *SbAGA1* expression increased ~20-fold between floral initiation and anthesis in Della stems and 2-fold more by grain maturity (~40-fold overall). *SbAGA2* expression also increased ~8-fold between floral initiation and anthesis and continued to increase through grain maturity (~21-fold overall). The ~20-40-fold increase in *SbAGA1* and *SbAGA2* expression in stem internodes post floral initiation suggests hydrolysis of raffinose in sweet sorghum stems may contribute to the storage or retrieval of sugars from sorghum stems during this phase of development.

Table 3.2**:** The expression level of the RFO pathway genes in sweet Sorghum Della by RNA-seq in stem internodes from floral initiation to grain maturity. The numbers represent the average expression obtained from the TPM (Transcripts Per Kilobase Million) plots.

	Days relative to anthesis							
RFO genes	-45	-29	-19	-11	0	18	27	54
SbGolS1	24	31	10	5	7	35	76	162
SbGolS2	2	2	1	1	1	1	3	4
SbRS	4	1	3	3	1	5	1	8
SbSTS	0	1	1	3	1	1	1	1
SbAGA1	20	34	63	182	745	1012	1463	1181
SbAGA2.1	0	10	5	53	89	94	222	387
SbAGA2.2	3	11	10	30	70	177	176	93
SbAGA3	13	1	1	2	2	3	3	6
SbAGA4.1	20	3	5	17	4	0	7	3
SbAGA4.2	19	7	17	8	4	5	8	6
SbAGA5	4	1	3	2	3	3	7	4
SbAGA6	75	24	18	19	22	20	19	24

Expression of *SbRS*, *SbAGA1*, and *SbAGA2* in Della pre- and post-anthesis

RNA-seq data showed relatively high expression of *SbRS* in leaves of BTx623 and that expression of *SbAGA1/2* in Della stems increased post floral initiation when sucrose accumulates in stems. This suggests the hypothesis that a portion of the raffinose synthesized in leaves is transported to stems for hydrolysis into sucrose and galactose. This hypothesis was examined further by examining the expression of *SbRS, SbAGA1,* and *SbAGA2* in leaves and stems of Della pre- and post-anthesis using qRT-PCR. The results in Figure 3.1 show that expression of *SbRS* was >32-fold higher in leaves compared to stems and that expression was maintained at similar levels in

leaves pre- and post-anthesis (Fig 3.1, top panel). In contrast, *SbAGA1* and *SbAGA2* are differentially expressed at higher levels in stems compared to leaves pre-anthesis (7-fold), and that expression increases significantly in stems post-anthesis (Fig 3.1, lower two panels).

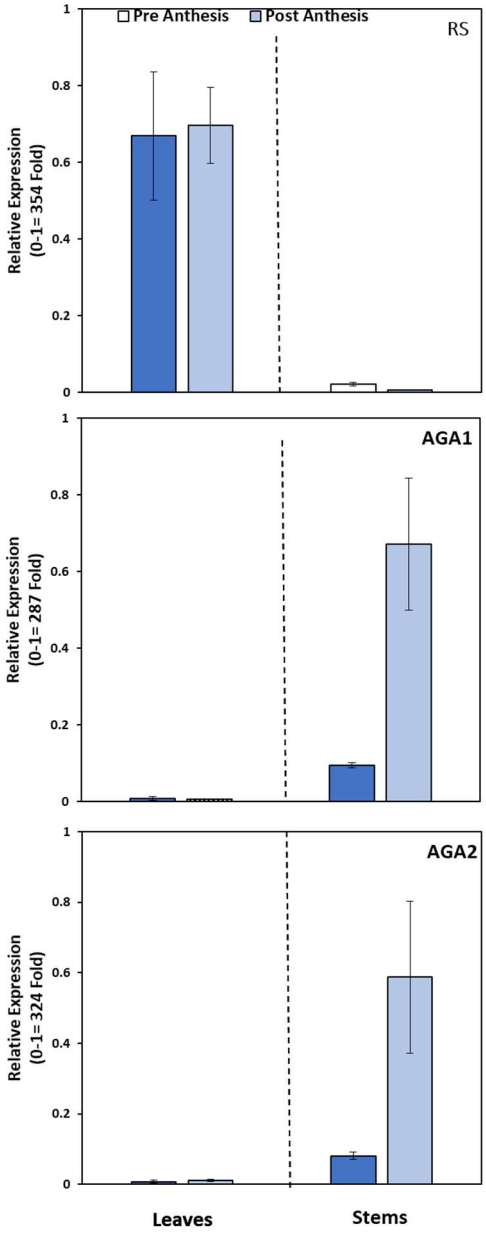

Figure 3.1**:** Expression levels of *SbRS*, *SbAGA1*, and *SbAGA2* in sweet sorghum (cultivar Della) stem and leaves, pre, and post-anthesis. A, B, and C show *SbRS*, *SbAGA1*, and *SbAGA2,* respectively. Values are means ± standard error of n = 3 plants. Relative gene expression is shown compared to endogenous *SbUBC* as a normalization control. Relative expression was

calculated via the $2^{-\Delta\Delta C_t}$ method relative to the internode section with the gene's highest expression in each group [97]. The fold change in expression between the minimum and maximum values on the y-axis was calculated based on *SbUBC* normalized values according to $\mathrm{FC} = 2^{\Delta C_t(\max) - \Delta C_t(\min)}$.

To understand if a similar pattern was exhibited in other sweet sorghum cultivars, I analyzed the expression of *SbRS* and *SbAGA1* and *SbAGA2* genes in the sweet sorghum Wray grown under greenhouse conditions.

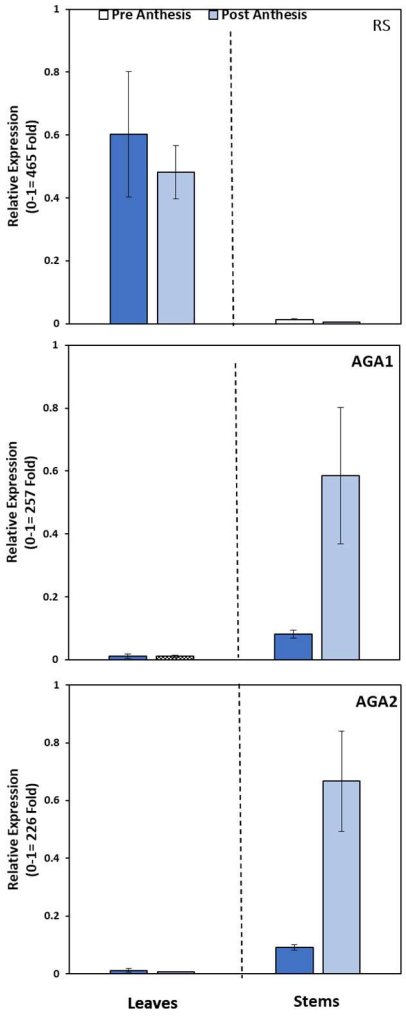

Figure 3.2**:** Expression levels of *SbRS*, *SbAGA1*, and *SbAGA2* in sweet sorghum (cultivar Wray) stems and leaves, pre, and post-anthesis. A, B, and C show *SbRS*, *SbAGA1*, and *SbAGA2,* respectively. Values are means ± standard error of n = 3 plants. Relative gene expression is shown compared to endogenous *SbUBC* as a normalization control. Relative expression was

calculated via the $2^{-\Delta\Delta C_t}$ method relative to the internode section with the gene's highest expression in each group [97]. Fold change in expression between the minimum and maximum values on the y-axis was calculated based on *SbUBC* normalized values according to $\mathrm{FC} = 2^{\Delta C_t(\max) - \Delta C_t(\min)}$.

The expression of *SbRS*, *SbAGA1*, and *SbAGA2* was similar in leaves of Wray pre- and post-anthesis. In Wray preanthesis, *SbRS* expression was ~44-fold higher in the leaves relative to stem internodes. In contrast, expression of *SbAGA1* and *SbAGA2* was ~7.2 fold and ~7.3-fold higher in stems relative to the leaves pre-anthesis. Post-anthesis, *SbRS* expression was ~111-fold higher in the leaves relative to Wray stem internodes, and *SbAGA1* and *SbAGA2* expression were ~56-fold and ~97-fold higher in stems relative to leaves.

To further extend the analysis to other sweet sorghum cultivars, I analyzed the relative expression of *SbRS*, *SbAGA1,* and *SbAGA2* in sweet sorghum cultivars Rio and Umbrella sorghum grown under field conditions post-anthesis when sucrose accumulates in the stems (Fig 3.3, 3.4).

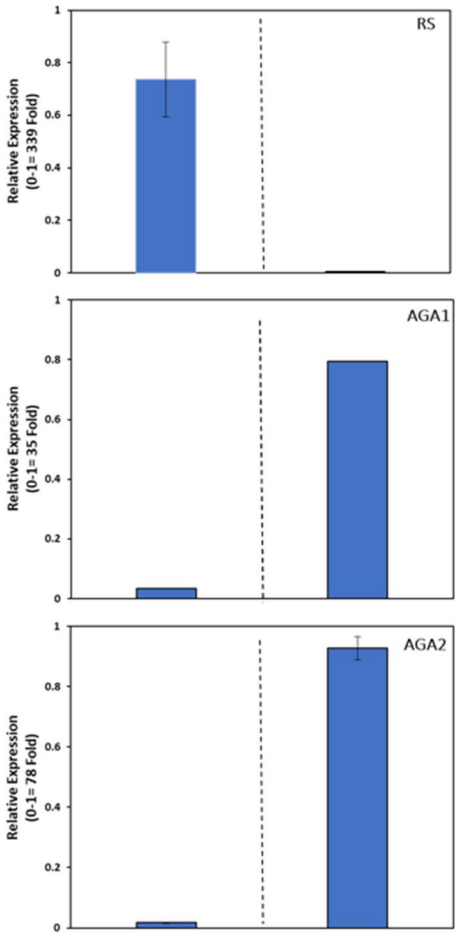

Figure 3.3**:** Expression levels of *SbRS*, *SbAGA1*, and *SbAGA2* in sweet sorghum cultivar Rio leaves and stems post-anthesis. Samples were taken from the field. A, B, and C show *SbRS*, *SbAGA1*, and *SbAGA2,* respectively. Values are means ± standard error of n = 3 plants. Relative gene expression is shown compared to endogenous *SbUBC* as a normalization control. Relative expression was calculated via the $2^{-\Delta\Delta C_t}$ method relative to the internode section with the gene's highest expression in each group [97]. Fold change in expression between the minimum and maximum values on the y-axis was calculated based on SbUBC normalized values according to $\mathrm{FC} = 2^{\Delta C_t(\max) - \Delta C_t(\min)}$.

In Rio, *SbRS* was expressed at ~173-fold higher levels in the leaves relative to stem internodes. *SbAGA1* and *SbAGA2* were expressed ~24-fold and ~55-fold higher in stems relative to the leaves, respectively.

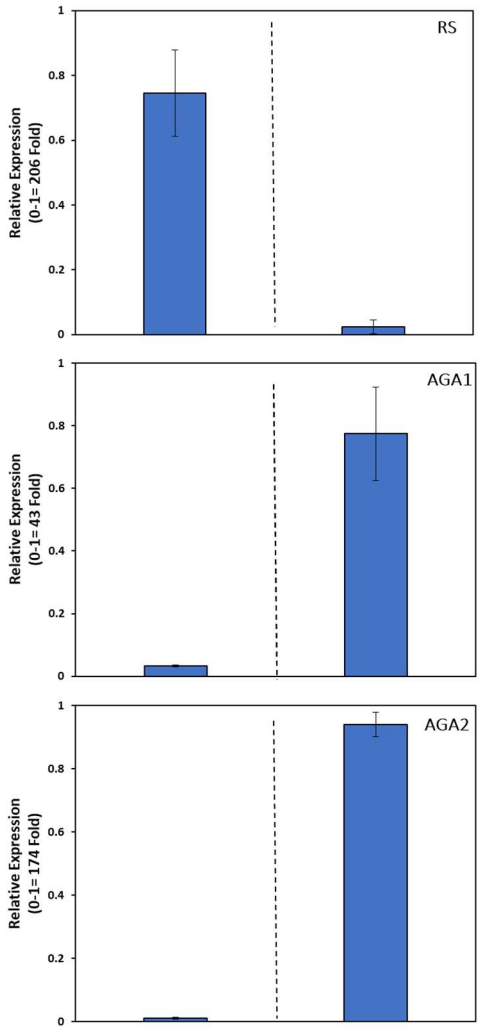

Figure 3.4**:** Expression levels of *SbRS*, *SbAGA1*, and *SbAGA2* in sweet sorghum cultivar Umbrella leaves and stems post-anthesis. Samples were taken from the field. A, B, and C show *SbRS*, *SbAGA1*, and *SbAGA2,* respectively. Values are means ± standard error of n = 3 plants. Relative gene expression is shown compared to endogenous *SbUBC* as a normalization control. Relative expression was calculated via the $2^{-\Delta\Delta C_t}$ method relative to the internode section with the gene's highest expression in each group [97]. The fold change in expression between the minimum and maximum values on the y-axis was calculated based on SbUBC normalized values according to $\mathrm{FC} = 2^{\Delta C_t(\max) - \Delta C_t(\min)}$.

Post-anthesis samples were also collected from field-grown Umbrella after grain maturity. *SbRS* expression was ~30-fold higher in the leaves relative to stem internodes. *SbAGA1* and *SbAGA2* expression were ~23-fold and ~92-fold higher expression in stems relative to the leaves, respectively, in this genotype.

Data in Table 3.2 shows that *SbGolS1* expression increases in the leaves post-anthesis compared to expression in the stem. The step catalyzed by galactinol synthases is the rate-limiting step in RFO synthesis. Therefore, I checked the expression of SbGolS1 in the stems and leaves of sweet sorghum cultivars Della and Wray pre and post-anthesis.

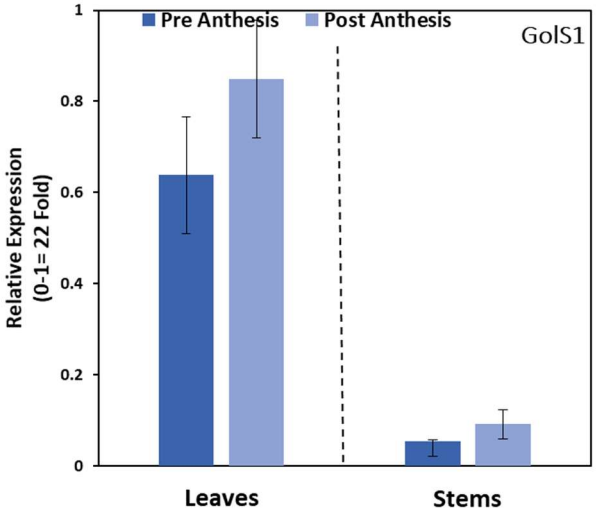

Figure 3.5**:** Expression levels of *SbGolS1*in sweet sorghum cultivar Della leaves and stems, pre and post-anthesis.

SbGolS1 expression was ~12-fold higher in the leaves relative to stem internodes in cultivar Della pre- and post-anthesis (Figure 3.5).

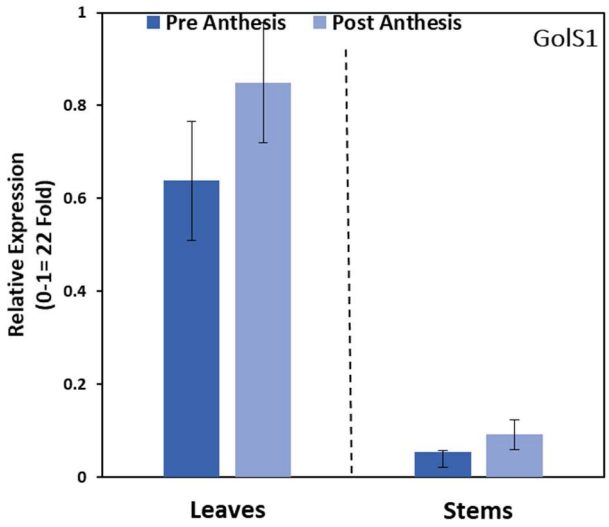

Figure 3.6**:** Expression levels of *SbGolS1* in sweet sorghum cultivar Wray leaves and stems, pre and post-anthesis.

SbGolS1 expression was ~9-fold higher in the leaves relative to stem internodes in cultivar Wray (Figure 3.6).

In situ analysis of SbRS, SbAGA1, and SbAGA2 expression in leaves and stems

The role of RFOs in leaves and stems was further investigated by determining where genes in the RFO pathway were expressed using RNA *in situ* hybridization (RNA-ish). This technique was performed in collaboration with Dr. Starla Zemelis-Durfee (Michigan State University) using an RNA-scope platform to determine the cellular location of transcripts that encode RS, AGA1, and AGA2 in mature leaf blade tissue. Leaf tissue samples were collected at ~40 days after anthesis at 11 am since this is approximately when peak RS and AGA1 expression occurs.

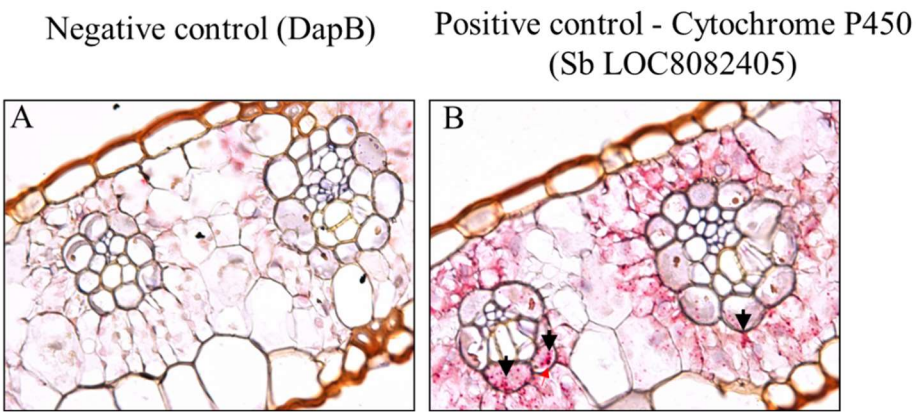

Figure 3.7: RNA in-situ hybridization controls from Della sorghum leaves. A) Negative control: DapB- a bacterial transcript with no detection and pink background. B) Positive control: Leaf tissue with Cytochrome P450 transcript localization in the bundle sheath cells surrounding the leaves' minor veins at 140 DAE. Transcript localization is visible as red precipitate marked with black arrows Magnification using a 40X objective.

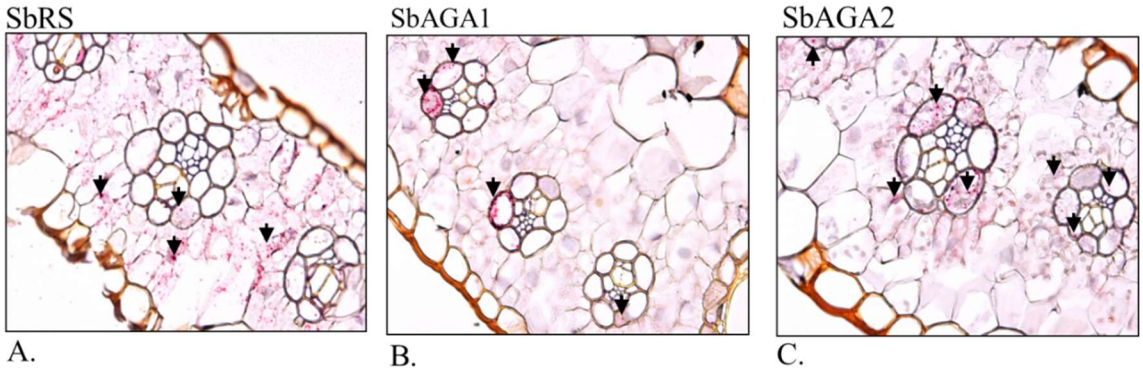

Figure 3.8: RNA in-situ hybridization results from Della sorghum leaves. A) SbRS transcript localization in mesophyll cells surrounding the minor veins of leaves at 140 DAE. B) SbAGA1 transcript localization in the bundle sheath cells surrounding the leaves' minor veins at 140 DAE. C) SbAGA2 transcript localization in the bundle sheath cells surrounding the leaves' minor veins at 140 DAE. Magnification using a 40X objective.

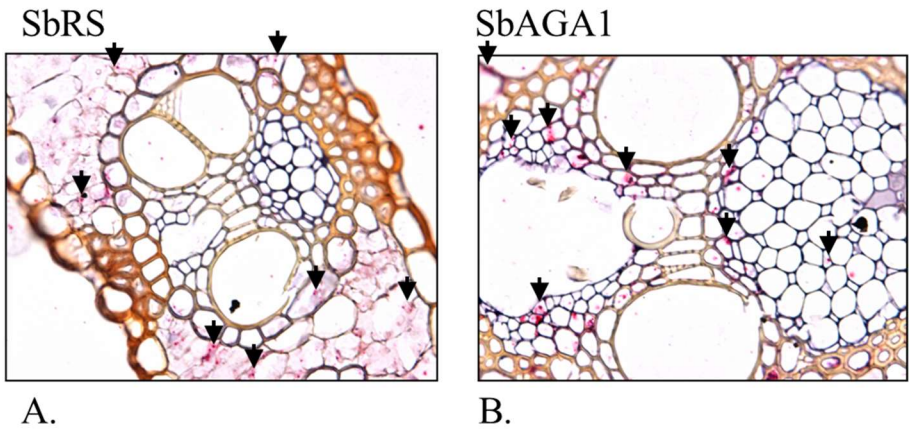

Figure 3.9**:** RNA in-situ hybridization results from sorghum leaves. A) SbRS transcript localization in mesophyll cells surrounding the major veins of leaves at 140 DAE. B) SbAGA1 transcript localization in the xylem parenchyma (XP) cells surrounding the leaves' major veins at 140 DAE. Magnification using a 40X objective.

Positive and negative controls for RNA-ish are shown in Figure 3.7. The transcript localization is visible as a red precipitate in the form of red dots against a pink background marked with black arrows. Negative controls have a pink background that lack red dots. In leaves, *SbRS* transcripts were detected primarily in mesophyll cells, a site of sucrose synthesis in sorghum leaves (Figure 3.8). A low abundance of *SbRS* transcripts was also observed in bundle sheath cells surrounding the large and small vascular bundles. It was observed that *SbAGA1* and *SbAGA2* are expressed primarily in the bundle sheath cells of the small and large veins of sorghum leaves. *SbAGA1* was also expressed in the xylem (XP) and phloem parenchyma cells in large leaf veins. This expression pattern suggests that a portion of the sucrose produced in the mesophyll cells of mature sorghum leaves is converted into raffinose in the mesophyll cells and transported into bundle sheath and xylem/phloem parenchyma cells for hydrolysis, releasing sucrose. Sucrose

could be exported from these cells via SWEET transporters [50] and taken up by phloem cells via SUTs[48].

Discussion

Expression of genes in the RFO pathway during development

Galactinol synthase acts on UDP-galactose and myo-inositol to form galactinol, considered the rate-limiting and committed step in RFO biosynthesis [75]. In grain sorghum BTx623, *SbGolS1* is highly expressed in leaf blades of juvenile leaves, in leaves at floral initiation, anthesis, and grain maturity (Table 3.1). *SbGolS2* was only highly expressed in leaf blades of juvenile leaves, consistent with expression data from different isoforms of galactinol synthases in plants. *SbGolS1* is highly expressed in dry, mature seeds and might play a critical role in maintaining seed longevity, desiccation tolerance, and germination capability as it does in maize and rapeseed [75, 98].

Since RFOs are accumulated in plants in response to abiotic and biotic stresses, the galactinol synthase gene is a potential breeding candidate for improving stress tolerance in plants. GolS have been identified in a number of plant species on a genome-wide scale[76, 99] and molecularly characterized through heterologous expression in *E. coli*[100]. Overexpression of an arabidopsis galactinol synthase (*AtGolS2*) gene in rice increased its grain yield while also providing relief from drought stress [63]. These observations have made galactinol synthases and other RFOs an important target for abiotic stress tolerance in crop species. Also, the high

expression of *SbGolS1* and *SbGolS2* in the leaf sheath and leaf blades at anthesis and maturation suggests that raffinose synthesized in leaves is exported to the growing seeds.

SbSTS expression was relatively low in most tissues. *SbAGAs* (1-6) showed a wide range of expression patterns, with *SbAGA1* showing the highest expression levels among these genes. *SbAGA1* was highly expressed in internodes during grain filling, in nascent spikelets, and seed at maturity. In contrast, *SbAGA6* was expressed at high levels in roots.

Even though raffinose has been detected in mature and desiccated sorghum seeds[101, 102], minimal to no *SbRS* expression is observed in the developing panicles and maturing seeds, which suggests that raffinose and stachyose are transported to the developing seed. Raffinose synthase also increases seed vigor in maize and improves seedling growth[82, 103].

Potential involvement of the RFO pathway in source-sink sugar transport

A prior RNA-seq study explored biomass partitioning and sucrose accumulation in sweet sorghum Della[39]. In that study, internode samples were collected over 100 days from floral initiation through post grain maturity. This period spans a phase of development when stem sucrose and starch increase significantly. Analysis of RFO-gene expression during this development phase showed that *SbRS* and *SbSTS* expression remained low in stems, but *SbAGA1* and *SbAGA2* expression increased significantly in parallel with stem sucrose accumulation (Figure 3.1, 3.2).

The expression of *SbRS, SbAGA1,* and *SbAGA2* remains relatively constant in the leaves of sweet sorghum genotypes pre-anthesis and post-anthesis. Expression of *SbRS*, *SbGolS1,* and *SbGolS1* in leaves indicates that raffinose synthesis occurs in leaves. The expression of *SbAGA1* and *SbAGA2* increases in stems post-anthesis, potentially increasing hydrolysis of RFO sugars in stems during the phase of stem sucrose accumulation. Raffinose synthesized in leaves could be transported to stems where it is hydrolyzed, thereby increasing sucrose levels. This sucrose transport mechanism is activated during the accumulation of sucrose/starch post-anthesis in stems of sweet sorghum, potentially helping to account for the very high sucrose levels in stems of these plants. Thus, these data suggest that differential expression of *SbAGA1* and *SbAGA2* genes may play an important role in sugar accumulation in sweet sorghum stems post-anthesis.

The pool of RFO metabolites could also increase photosynthetic activity by reducing leaf sucrose levels that could inhibit photosynthesis. Galactinol and raffinose are also thought to protect plants from reactive oxygen species during high rates of photosynthesis.[61]

SbAGA1 was highly expressed in Della stems during the grain filling phase. This expression pattern points to a role for RFO turnover in stems during the reproductive phase when stem sugar levels often increase significantly in sweet sorghums Della [39]. This result is consistent with an observation from a recent whole-genome sequencing study of sweet sorghum Rio which shows high sequence similarity to grain sorghum BTx623 and is an appropriate choice to explore differential sucrose accumulation in sorghum. In the whole-genome study, they noted that *SbAGA1* (Sobic.002G075800) is downregulated during the vegetative phase but is significantly up-regulated, starting at floral initiation, which points to its role in increasing sucrose

metabolism and storage in sweet sorghum stem[9]. Three cDNAs for alkaline α-galactosidases were identified in *Zea mays,* and two of them were cloned (*ZmAGA1* and *ZmAGA3*), both of which were expressed in seeds and callus tissue when exposed to abiotic stresses. It was suggested that the alkaline α-galactosidases are solely responsible for RFO breakdown in germinating maize seeds, while acidic galactosidases would have other functions [104].

Based on these preliminary findings, I decided to explore the potential role of sorghum *AGAs* in sucrose turnover in sorghum stems. I first looked at expression levels of *SbRS*, *SbAGA1*, and *SbAGA2* in the leaves and stems of two other sweet sorghum genotypes, Wray and Rio, pre-anthesis and post-anthesis in the field and under greenhouse conditions. In leaves of sweet sorghum Della and Wray, pre-anthesis and post-anthesis *SbRS* expression remained relatively constant, but the ratio of expression between the leaves and stems significantly increased as the plant transitioned from pre-anthesis to post-anthesis. Expression levels of *SbAGA1* and *SbAGA2* significantly increased post-anthesis in the stem, which coincides with sucrose accumulation in the stem. Along with hydrolyzing raffinose, AGAs are also capable of hydrolyzing α-galactosides such as galactinol. A study comparing the transcriptome and metabolome of a sweet sorghum Rio during the sucrose accumulation phase across five-time points noticed that as sucrose concentration began to increase in the stem, galactinol and myo-inositol levels decreased [30].

As an NADP-malic enzyme type of C4 plant, sorghum exhibits Kranz anatomy in the leaf blade[105]. The carbon fixing mechanism is split between mesophyll cells (M), where PEPCase initially fixes carbon dioxide in the form of a C4 organic acid, which is subsequently transported into bundle sheath (BS) cells that surround the vasculature where CO_2 is released and refixed by

RUBISCO and subsequently converted into sucrose. Mesophyll cells and the bundle sheath cells share dense plasmodesmatal (PD) connections to exchange photosynthetic metabolites. This anatomy and the accompanying increase in CO_2 concentration in cells containing RUBISCO reduces photorespiration and enables C4 plants to achieve more efficient carbon dioxide fixation than C3 plants[106, 107]. The pattern of *SbRS, SbAGA1,* and *SbAGA2* transcript localization in the leaves suggests their potential role in phloem loading, which will require further investigation using transgenic approaches to downregulate or upregulate the key genes.

4. CONCLUSION

Globally, sorghum is the fifth largest cereal crop in terms of production and planted area. What makes it even more unique is the variety of terminal uses it is cultivated for in various parts of the world. It is a staple food crop in semi-arid regions of Africa, whereas in the USA and Australia, it is a major feed crop and cultivated for bioenergy purposes[108]. In the last decades, most sugar yield increases have been made by increasing biomass through conventional breeding methods[28]. With the improvements in whole genome sequencing and transformation technologies, sorghum is rapidly emerging as a potential C4 biofuel crop. Even though sorghum has minimal input requirements, it lags behind grain yield compared to other cereal crops. The difference is even more drastic when comparing developed vs developing countries. However, due to its high stem sugar content it has emerged as a promising target for both sugar based and lignocellulosic biofuels production.

Sweet sorghum stems account for approximately 80% of its harvested biomass, which makes stem composition an important target for optimization and most of the stem biomass is accumulated in the form of soluble sugars and cell walls. Soluble sugars measured in Brix can account from 14% to 23% in sweet sorghum stems which can be directly fermented to ethanol. High stem sucrose yield is a useful trait for biofuels production because the conversion of simple sugars to biofuels and bio-products is low cost and highly efficient[109]. According to USDA, the ratio of energy invested to energy obtained during biofuel extraction from sweet sorghum is estimated at 1:8[110]. Ethanol fermented from sweet sorghum juice has a high-octane value, is low sulfur and biological content[111]. The fermentable carbohydrate content is also a key

determinant of the economic and environmental feasibility of renewable biofuel production[112]. Current sweet sorghum varieties, producing comparatively low sugar content (around 500 mmol/L), urgently require plant breeders to improve sugar accumulation in stems for biofuel[28]. In this study I've identified *SbAGAs* as potential targets for sorghum genetic engineering to potentially regulate stem sugar content.

In a recent study total sugar content of a sweet sorghum cultivar Rio was increased from 480 mM to 750 mM by introducing a stem pith parenchyma cell vacuole targeted sucrose isomerase under the control of a stem specific promoter[113]. Similarly, SbAGAs in the stem can be used as a potential target for genetic engineering to improve stem sugar content. Future work will be required to provide a better understanding of *SbAGAs* involvement in sucrose accumulation in stem post anthesis.

The research described in this thesis contributes to a greater understanding of differential sucrose accumulation in sweet sorghum. Integration of the RNA-seq and RNA-ISH data revealed that high *SbRS* expression in the leaves and high *SbAGA1* and *SbAGA2* expression in sorghum internode plays an important role in sucrose accumulation post-anthesis. Overall, identifying these genes is an important step toward understanding the role of RFOs in controlling sucrose accumulation in sweet sorghum. This study also provides the starting point for validation of these gene functions and modification of stem composition.

REFERENCES

1. Schnepf, R.D. and B.D. Yacobucci, *Renewable fuel standard (RFS): overview and issues*. Vol. 40155. 2010: Congressional Research Service Washington, DC.
2. Smith, C.W. and R.A. Frederiksen, *Sorghum: Origin, history, technology, and production*. Vol. 2. 2000: John Wiley & Sons.
3. Harlan, J.R. and J.M. de Wet, *A simplified classification of cultivated sorghum 1.* Crop science, 1972. **12**(2): p. 172-176.
4. Quinby, J., *The genetics of sorghum improvement.* Journal of Heredity, 1975. **66**(2): p. 56-62.
5. Paterson, A.H., et al., *The Sorghum bicolor genome and the diversification of grasses.* Nature, 2009. **457**(7229): p. 551-556.
6. Paterson, A.H., *Genomics of sorghum.* International Journal of Plant Genomics, 2008. **2008**.
7. Morris, G.P., et al., *Population genomic and genome-wide association studies of agroclimatic traits in sorghum.* Proceedings of the National Academy of Sciences, 2013. **110**(2): p. 453-458.
8. McCormick, R.F., et al., *The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization.* The Plant Journal, 2017.
9. Cooper, E.A., et al., *A new reference genome for Sorghum bicolor reveals high levels of sequence similarity between sweet and grain genotypes: implications for the genetics of sugar metabolism.* BMC genomics, 2019. **20**(1): p. 420.
10. Deschamps, S., et al., *A chromosome-scale assembly of the sorghum genome using nanopore sequencing and optical mapping.* Nature communications, 2018. **9**(1): p. 1-10.
11. Mace, E., et al., *The Sorghum QTL Atlas: a powerful tool for trait dissection, comparative genomics and crop improvement.* Theoretical and applied genetics, 2019. **132**(3): p. 751-766.
12. Shakoor, N., et al., *A Sorghum bicolorexpression atlas reveals dynamic genotype-specific expression profiles for vegetative tissues of grain, sweet and bioenergy sorghums.* BMC plant biology, 2014. **14**(1): p. 35.
13. Brenton, Z.W., et al., *A genomic resource for the development, improvement, and exploitation of sorghum for bioenergy.* Genetics, 2016. **204**(1): p. 21-33.
14. Luo, H., et al., *SorGSD: a sorghum genome SNP database.* Biotechnology for biofuels, 2016. **9**(1): p. 6.
15. Boyles, R.E., Z.W. Brenton, and S. Kresovich, *Genetic and genomic resources of sorghum to connect genotype with phenotype in contrasting environments.* The Plant Journal, 2019. **97**(1): p. 19-39.
16. Mace, E. and D. Jordan, *Location of major effect genes in sorghum (Sorghum bicolor (L.) Moench).* Theoretical and Applied Genetics, 2010. **121**(7): p. 1339-1356.

17. Murphy, R.L., et al., *Coincident light and clock regulation of pseudoresponse regulator protein 37 (PRR37) controls photoperiodic flowering in sorghum.* Proceedings of the National Academy of Sciences, 2011. **108**(39): p. 16469-16474.
18. Childs, K.L., et al., *The sorghum photoperiod sensitivity gene, Ma3, encodes a phytochrome B.* Plant Physiology, 1997. **113**(2): p. 611-619.
19. Yang, S., et al., *Sorghum phytochrome B inhibits flowering in long days by activating expression of SbPRR37 and SbGHD7, repressors of SbEHD1, SbCN8 and SbCN12.* PloS one, 2014. **9**(8): p. e105352.
20. Murphy, R.L., et al., *Ghd7 (Ma6) represses sorghum flowering in long days: Ghd7 alleles enhance biomass accumulation and grain production.* The Plant Genome, 2014. **7**(2): p. 1-10.
21. Casto, A.L., et al., *Maturity2, a novel regulator of flowering time in Sorghum bicolor, increases expression of SbPRR37 and SbCO in long days delaying flowering.* Plos one, 2019. **14**(4): p. e0212154.
22. Quinby, J. and R. Karper, *Inheritance of Height in Sorghum 1.* Agronomy Journal, 1954. **46**(5): p. 211-216.
23. Hilley, J.L., et al., *Sorghum Dw2 encodes a protein kinase regulator of stem internode length.* Scientific reports, 2017. **7**(1): p. 1-13.
24. Multani, D.S., et al., *Loss of an MDR transporter in compact stalks of maize br2 and sorghum dw3 mutants.* Science, 2003. **302**(5642): p. 81-84.
25. Calviño, M. and J. Messing, *Sweet sorghum as a model system for bioenergy crops.* Current Opinion in Biotechnology, 2012. **23**(3): p. 323-329.
26. Mullet, J., et al., *Energy sorghum—a genetic model for the design of C4 grass bioenergy crops.* Journal of experimental botany, 2014. **65**(13): p. 3479-3489.
27. Che, P., et al., *Developing a flexible, High‐efficiency Agrobacterium‐mediated Sorghum Transformation System with Broad Application.* Plant biotechnology journal, 2018.
28. Mathur, S., et al., *Sweet sorghum as biofuel feedstock: recent advances and available resources.* Biotechnology for biofuels, 2017. **10**(1): p. 146.
29. Anami, S.E., et al., *Sweet sorghum ideotypes: genetic improvement of the biofuel syndrome.* Food and Energy Security, 2015. **4**(3): p. 159-177.
30. Li, Y., et al., *Transcriptome and metabolome reveal distinct carbon allocation patterns during internode sugar accumulation in different sorghum genotypes.* Plant biotechnology journal, 2019. **17**(2): p. 472-487.
31. Yu, H., et al., *Identification of differentially expressed microRNA in the stems and leaves during sugar accumulation in sweet sorghum.* Gene, 2015. **571**(2): p. 221-230.
32. Zheng, L.-Y., et al., *Genome-wide patterns of genetic variation in sweet and grain sorghum (Sorghum bicolor).* Genome biology, 2011. **12**(11): p. R114.
33. Jiang, S.-Y., et al., *Genetic variation and expression diversity between grain and sweet sorghum lines.* BMC genomics, 2013. **14**(1): p. 18.
34. Bihmidine, S., et al., *Sucrose accumulation in sweet sorghum stems occurs by apoplasmic phloem unloading and does not involve differential Sucrose transporter expression.* BMC plant biology, 2015. **15**(1): p. 186.

35. Ritter, K.B., et al., *An assessment of the genetic relationship between sweet and grain sorghums, within Sorghum bicolor ssp. bicolor (L.) Moench, using AFLP markers.* Euphytica, 2007. **157**(1-2): p. 161-176.
36. Brenton, Z.W., et al., *Species-specific duplication event associated with elevated levels of nonstructural carbohydrates in Sorghum bicolor.* G3: Genes, Genomes, Genetics, 2020. **10**(5): p. 1511-1520.
37. Murray, S.C., et al., *Genetic improvement of sorghum as a biofuel feedstock: II. QTL for stem and leaf structural carbohydrates.* Crop science, 2008. **48**(6): p. 2180-2193.
38. Murray, S.C., et al., *Sweet sorghum genetic diversity and association mapping for brix and height.* The Plant Genome, 2009. **2**(1).
39. McKinley, B., et al., *Dynamics of biomass partitioning, stem gene expression, cell wall biosynthesis, and sucrose accumulation during development of Sorghum bicolor.* The Plant journal, 2016. **88**(4): p. 662-680.
40. Bihmidine, S., et al., *Tonoplast Sugar Transporters (SbTSTs) putatively control sucrose accumulation in sweet sorghum stems.* Plant signaling & behavior, 2016. **11**(1): p. e1117721.
41. Hoffmann‐Thoma, G., et al., *Sucrose accumulation in sweet sorghum stem internodes in relation to growth.* Physiologia Plantarum, 1996. **97**(2): p. 277-284.
42. Milne, R.J., et al., *Are sucrose transporter expression profiles linked with patterns of biomass partitioning in Sorghum phenotypes?* Frontiers in plant science, 2013. **4**: p. 223.
43. Slewinski, T.L., R. Meeley, and D.M. Braun, *Sucrose transporter1 functions in phloem loading in maize leaves.* Journal of Experimental Botany, 2009. **60**(3): p. 881-892.
44. Braun, D.M., L. Wang, and Y.-L. Ruan, *Understanding and manipulating sucrose phloem loading, unloading, metabolism, and signalling to enhance crop yield and food security.* Journal of Experimental Botany, 2013. **65**(7): p. 1713-1735.
45. Baker, R.F., et al., *Sucrose transporter ZmSut1 expression and localization uncover new insights into sucrose phloem loading.* Plant physiology, 2016: p. pp. 00884.2016.
46. Walsh, K.B., R.C. Sky, and S.M. Brown, *The anatomy of the pathway of sucrose unloading within the sugarcane stalk.* Functional Plant Biology, 2005. **32**(4): p. 367-374.
47. Milne, R.J., et al., *Sucrose transporter localization and function in phloem loading and unloading.* Plant physiology, 2016: p. pp. 01594.2016.
48. Milne, R.J., et al., *Sucrose transporter localization and function in phloem unloading in developing stems.* Plant physiology, 2017. **173**(2): p. 1330-1341.
49. Chi, Y., et al., *Vacuolar invertase genes SbVIN1 and SbVIN2 are differently associated with stem and grain traits in sorghum (Sorghum bicolor).* The Crop Journal, 2020. **8**(2): p. 299-312.
50. Mizuno, H., S. Kasuga, and H. Kawahigashi, *The sorghum SWEET gene family: stem sucrose accumulation as revealed through transcriptome profiling.* Biotechnology for biofuels, 2016. **9**(1): p. 127.
51. Wormit, A., et al., *Molecular identification and physiological characterization of a novel monosaccharide transporter from Arabidopsis involved in vacuolar sugar transport.* The Plant Cell, 2006. **18**(12): p. 3476-3490.
52. Yang, L., et al., *Correlation analysis between the key enzymes activities and sugar content in sweet sorghum (Sorghum bicolor L. Moench) stems at physiological maturity stage.* Australian Journal of Crop Science, 2013. **7**(1): p. 84.

53. Morey, S.R., et al., *Evaluation of performance of sorghum varieties grown in Tokyo for sugar accumulation and its correlation with vacuolar invertase genes SbInv1 and SbInv2.* Plant Production Science, 2018. **21**(4): p. 328-338.
54. Jiang, W., et al., *Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice.* Nucleic acids research, 2013. **41**(20): p. e188-e188.
55. Wu, E., et al., *Optimized Agrobacterium-mediated sorghum transformation protocol and molecular data of transgenic sorghum plants.* In Vitro Cellular & Developmental Biology-Plant, 2014. **50**(1): p. 9-18.
56. Belide, S., et al., *Robust genetic transformation of sorghum (Sorghum bicolor L.) using differentiating embryogenic callus induced from immature embryos.* Plant Methods, 2017. **13**(1): p. 109.
57. Mahajan, S. and N. Tuteja, *Cold, salinity and drought stresses: an overview.* Archives of biochemistry and biophysics, 2005. **444**(2): p. 139-158.
58. Kakumanu, A., et al., *Effects of drought on gene expression in maize reproductive and leaf meristem tissue revealed by RNA-Seq.* Plant Physiology, 2012. **160**(2): p. 846-867.
59. Downie, B., et al., *Expression of a GALACTINOL SYNTHASE gene in tomato seeds is up-regulated before maturation desiccation and again after imbibition whenever radicle protrusion is prevented.* Plant Physiology, 2003. **131**(3): p. 1347-1359.
60. Sengupta, S., et al., *Significance of galactinol and raffinose family oligosaccharide synthesis in plants.* Frontiers in plant science, 2015. **6**: p. 656.
61. Nishizawa, A., Y. Yabuta, and S. Shigeoka, *Galactinol and raffinose constitute a novel function to protect plants from oxidative damage.* Plant physiology, 2008. **147**(3): p. 1251-1263.
62. Wu, X., et al., *Accumulation of raffinose in rice seedlings overexpressing OsWRKY11 in relation to desiccation tolerance.* Plant biotechnology, 2009. **26**(4): p. 431-434.
63. Selvaraj, M.G., et al., *Overexpression of an Arabidopsis thaliana galactinol synthase gene improves drought tolerance in transgenic rice and increased grain yield in the field.* Plant biotechnology journal, 2017.
64. Yu, X., et al., *Enhancement of abiotic stress tolerance in poplar by overexpression of key Arabidopsis stress response genes, AtSRK2C and AtGolS2.* Molecular Breeding, 2017. **37**(5): p. 57.
65. Lü, J., et al., *Suppression of cucumber stachyose synthase gene (CsSTS) inhibits phloem loading and reduces low temperature stress tolerance.* Plant molecular biology, 2017. **95**(1-2): p. 1-15.
66. Li, T., et al., *Raffinose synthase enhances drought tolerance through raffinose synthesis or galactinol hydrolysis in maize and Arabidopsis plants.* Journal of Biological Chemistry, 2020. **295**(23): p. 8064-8077.
67. Han, Q., et al., *ZmDREB2A regulates ZmGH3. 2 and ZmRAFS, shifting metabolism towards seed aging tolerance over seedling growth.* The Plant Journal, 2020.
68. Han, Q., et al., *ZmDREB1A Regulates RAFFINOSE SYNTHASE Controlling Raffinose Accumulation and Plant Chilling Stress Tolerance in Maize.* Plant and Cell Physiology, 2020. **61**(2): p. 331-341.

69. Gu, L., et al., *ZmGOLS2, a target of transcription factor ZmDREB2A, offers similar protection against abiotic stress as ZmDREB2A.* Plant molecular biology, 2016. **90**(1-2): p. 157-170.
70. Jang, J.-H., et al., *Arabidopsis galactinol synthases 1 (AtGOLS1) negatively regulates seed germination.* Plant Science, 2018. **267**: p. 94-101.
71. Himuro, Y., et al., *Arabidopsis galactinol synthase AtGolS2 improves drought tolerance in the monocot model Brachypodium distachyon.* Journal of plant physiology, 2014. **171**(13): p. 1127-1131.
72. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput.* Nucleic acids research, 2004. **32**(5): p. 1792-1797.
73. Kumar, S., G. Stecher, and K. Tamura, *MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets.* Molecular biology and evolution, 2016. **33**(7): p. 1870-1874.
74. Zhao, T.Y., et al., *Expression of the maize GALACTINOL SYNTHASE gene family:(II) Kernel abscission, environmental stress and myo‐inositol influences accumulation of transcript in developing seeds and callus cells.* Physiologia Plantarum, 2004. **121**(4): p. 647-655.
75. Zhao, T.Y., et al., *Expression of the maize GALACTINOL SYNTHASE gene family:(I) Expression of two different genes during seed development and germination.* Physiologia Plantarum, 2004. **121**(4): p. 634-646.
76. Filiz, E., I.I. Ozyigit, and R. Vatansever, *Genome-wide identification of galactinol synthase (GolS) genes in Solanum lycopersicum and Brachypodium distachyon.* Computational biology and chemistry, 2015. **58**: p. 149-157.
77. Zhou, Y., et al., *Molecular cloning and characterization of galactinol synthases in Camellia sinensis with different responses to biotic and abiotic stressors.* Journal of Agricultural and Food Chemistry, 2017. **65**(13): p. 2751-2759.
78. Liu, Y., et al., *Molecular cloning and expression of an encoding galactinol synthase gene (AnGolS1) in seedling of Ammopiptanthus nanus.* Scientific reports, 2016. **6**: p. 36113.
79. dos Santos, T.B. and L.G.E. Vieira, *Involvement of the galactinol synthase gene in abiotic and biotic stress responses: A review on current knowledge.* Plant Gene, 2020: p. 100258.
80. Sprenger, N. and F. Keller, *Allocation of raffinose family oligosaccharides to transport and storage pools in Ajuga reptans: the roles of two distinct galactinol synthases.* The Plant Journal, 2000. **21**(3): p. 249-258.
81. Zhou, M.-L., et al., *Genome-wide identification of genes involved in raffinose metabolism in Maize.* Glycobiology, 2012. **22**(12): p. 1775-1785.
82. Li, T., et al., *Regulation of Seed Vigor by Manipulation of Raffinose Family Oligosaccharides in Maize and Arabidopsis thaliana.* Molecular plant, 2017. **10**(12): p. 1540-1555.
83. Gangl, R., R. Behmüller, and R. Tenhaken, *Molecular cloning of AtRS4, a seed specific multifunctional RFO synthase/galactosylhydrolase in Arabidopsis thaliana.* Frontiers in plant science, 2015. **6**: p. 789.
84. Bailey, T.L., et al., *MEME SUITE: tools for motif discovery and searching.* Nucleic acids research, 2009. **37**(suppl_2): p. W202-W208.

85. Schneider, C.A., W.S. Rasband, and K.W. Eliceiri, *NIH Image to ImageJ: 25 years of image analysis.* Nature methods, 2012. **9**(7): p. 671-675.
86. Gao, Z. and A.A. Schaffer, *A novel alkaline α-galactosidase from melon fruit with a substrate preference for raffinose.* Plant physiology, 1999. **119**(3): p. 979-988.
87. Wang, L., et al., *Fermentation of sweet sorghum derived sugars to butyric acid at high titer and productivity by a moderate thermophile Clostridium thermobutyricum at 50° C.* Bioresource technology, 2015. **198**: p. 533-539.
88. Li, B.-Z., et al., *Process optimization to convert forage and sweet sorghum bagasse to ethanol based on ammonia fiber expansion (AFEX) pretreatment.* Bioresource technology, 2010. **101**(4): p. 1285-1292.
89. Qazi, H.A., S. Paranjpe, and S. Bhargava, *Stem sugar accumulation in sweet sorghum–activity and expression of sucrose metabolizing enzymes and sucrose transporters.* Journal of plant physiology, 2012. **169**(6): p. 605-613.
90. Duncan, K.A., S.C. Hardin, and S.C. Huber, *The three maize sucrose synthase isoforms differ in distribution, localization, and phosphorylation.* Plant and Cell Physiology, 2006. **47**(7): p. 959-971.
91. Shukla, S., et al., *The relationship between plant height and sugar accumulation in the stems of sweet sorghum (Sorghum bicolor (L.) Moench).* Field Crops Research, 2017. **203**: p. 181-191.
92. Broadhead, D., K. Freeman, and N. Zummo, *Registration of Wray Sweet Sorghum1 (Reg. No. 119).* Crop Science, 1981. **21**(6): p. 987-987.
93. Broadhead, D., *Registration of Rio Sweet Sorghum1 (Reg. No. 113).* Crop Science, 1972. **12**(5): p. 716-716.
94. Casto, A.L., et al., *Sorghum stem aerenchyma formation is regulated by SbNAC_D during internode development.* Plant Direct, 2018. **2**(11): p. e00085.
95. Karlgren, A., et al., *Non-radioactive in situ hybridization protocol applicable for Norway spruce and a range of plant species.* JoVE (Journal of Visualized Experiments), 2009(26): p. e1205.
96. Wang, F., et al., *RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues.* The Journal of molecular diagnostics, 2012. **14**(1): p. 22-29.
97. Pfaffl, M.W., *A new mathematical model for relative quantification in real-time RT–PCR.* Nucleic acids research, 2001. **29**(9): p. e45-e45.
98. Li, X., et al., *Expression of a GALACTINOL SYNTHASE gene is positively associated with desiccation tolerance of Brassica napus seeds during development.* Journal of plant physiology, 2011. **168**(15): p. 1761-1770.
99. Fan, Y., et al., *Genome-wide identification, evolutionary and expression analyses of the GALACTINOL SYNTHASE gene family in rapeseed and tobacco.* International journal of molecular sciences, 2017. **18**(12): p. 2768.
100. Zhou, T., R. Zhang, and S. Guo, *Molecular cloning and characterization of GhGolS1, a novel gene encoding galactinol synthase from cotton (Gossypium hirsutum).* Plant Molecular Biology Reporter, 2012. **30**(3): p. 699-709.
101. Watson, S. and Y. Hirata, *Carbohydrates in grain sorghum kernels.* Sorghum Newsl., 1960. **3**: p. 7-.
102. Anglani, C., *Sorghum carbohydrates–A review.* Plant foods for human nutrition, 1998. **52**(1): p. 77-83.

103. Brenac, P., et al., *Raffinose accumulation related to desiccation tolerance during maize (Zea mays L.) seed development and maturation.* Journal of Plant Physiology, 1997. **150**(4): p. 481-488.
104. Zhao, T.-Y., et al., *An alkaline [alpha]-galactosidase transcript is present in maize seeds and cultured embryo cells, and accumulates during stress.* Seed Science Research, 2006. **16**(2): p. 107.
105. Dengler, N.G., et al., *Quantitative leaf anatomy of C3 and C4 grasses (Poaceae): bundle sheath and mesophyll surface area relationships.* Annals of Botany, 1994. **73**(3): p. 241-255.
106. Döring, F., et al., *Most photorespiratory genes are preferentially expressed in the bundle sheath cells of the C4 grass Sorghum bicolor.* Journal of Experimental Botany, 2016. **67**(10): p. 3053-3064.
107. Sedelnikova, O.V., T.E. Hughes, and J.A. Langdale, *Understanding the genetic basis of C4 Kranz anatomy with a view to engineering C3 crops.* Annual review of genetics, 2018. **52**: p. 249-270.
108. Taylor, J.R., T.J. Schober, and S.R. Bean, *Novel food and non-food uses for sorghum and millets.* Journal of cereal science, 2006. **44**(3): p. 252-271.
109. Regassa, T.H. and C.S. Wortmann, *Sweet sorghum as a bioenergy crop: literature review.* Biomass and Bioenergy, 2014. **64**: p. 348-355.
110. Billings, M., *Biomass sorghum and sweet sorghum data gathering report.* W&A Crop Insurance. USDA-RMA, CTOR: Jaime Padget, Missouri Watts and Associates, Inc, 2015.
111. Zhou, A. and E. Thomson, *The development of biofuels in Asia.* Applied Energy, 2009. **86**: p. S11-S20.
112. Long, H., et al., *Biomass resources and their bioenergy potential estimation: A review.* Renewable and Sustainable Energy Reviews, 2013. **26**: p. 344-352.
113. Liu, G., et al., *Stem vacuole-targetted sucrose isomerase enhances sugar content in sorghum.* Biotechnology for biofuels, 2021. **14**(1): p. 1-14.

APPENDIX A

SUPPLEMENTARY INFORMATION

S1: Accession numbers for phylogenetic analysis and protein alignment for sequences used to construct the tree in Figure 2.1 and protein alignments in Figure 2.2 are as follows:

CsRS (*Cucumis sativus*; E15707); AtRS5 (*Arabidopsis thaliana*; NP_198855); PsRS (*Pisum sativum*; CAD20127); OsRS (*Oryza sativa*; XP_015621501); PsSTS (*Pisum sativum*; CAC38094); VaSTS (*Vigna angularis*; CAB64363); AmSTS (*Alonsoa meridionalis*; CAD31704); SaSTS (*Stachys affinis*; CAC86963); AtSTS (*Arabidopsis thaliana*; NP_192106); AtSIP1 (*Arabidopsis thaliana*; NP_175970); LeAGA1 (*Lycopersicon esculentum*; AAN32954); CmAGA1 (*Cucumis melo*; AAM75139); ZmAGA3 (*Zea mays*; AAQ07253); OsAGA1 (*Oryza sativa*; XP_483143); ZmAGA1 (*Zea mays*; AAQ07251); OsSIP1 (*Oryza sativa*; XP_477103); HvSIP1 (*Hordeum vulgare*; Q40077); CmAGA2 (*Cucumis melo*; AAM75140); PaSIP1 (*Persea americana*;CAB77245); AtSIP2 (*Arabidopsis thaliana*; NP_191311); BoSIP1 (*Brassica oleracea*; CAA55893); ZmSIP2 (*Zea mays*; AAQ07252); AtSIP3 (*Arabidopsis thaliana*;NP_001190347); SsSIP1 (*Sulfolobus solfataricus*; AAK43227); ZmRS7 (*Zea mays*; XP_008669826); ZmRS2 (*Zea mays*; ONM02661); ZmRS3 (*Zea mays*; XP_008665643)

APPENDIX B

SUPPLEMENTARY FIGURES

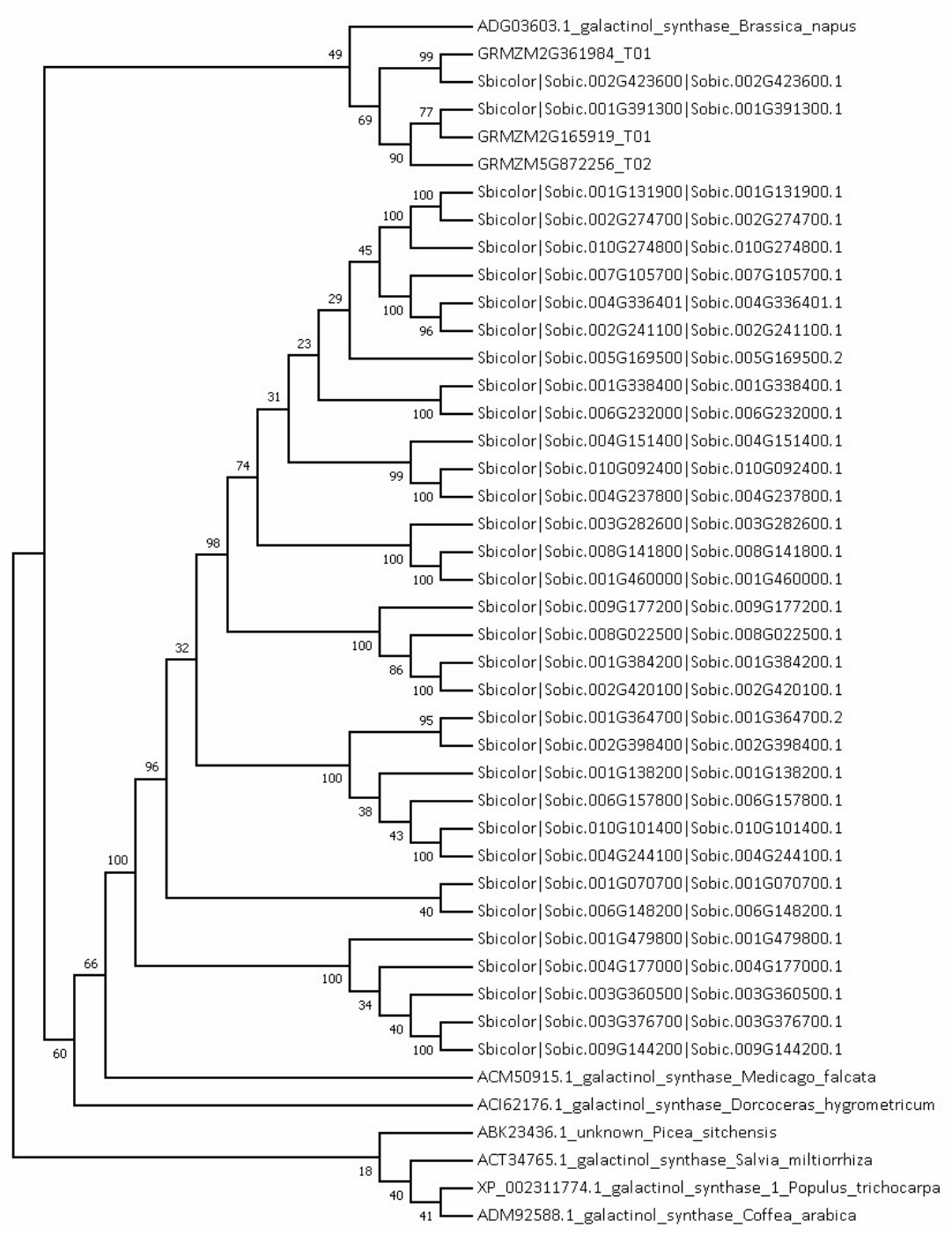

Figure S2.1: Phylogenetic tree of putative *S. bicolor* galactinol synthases and previously identified galactinol synthases from other species.

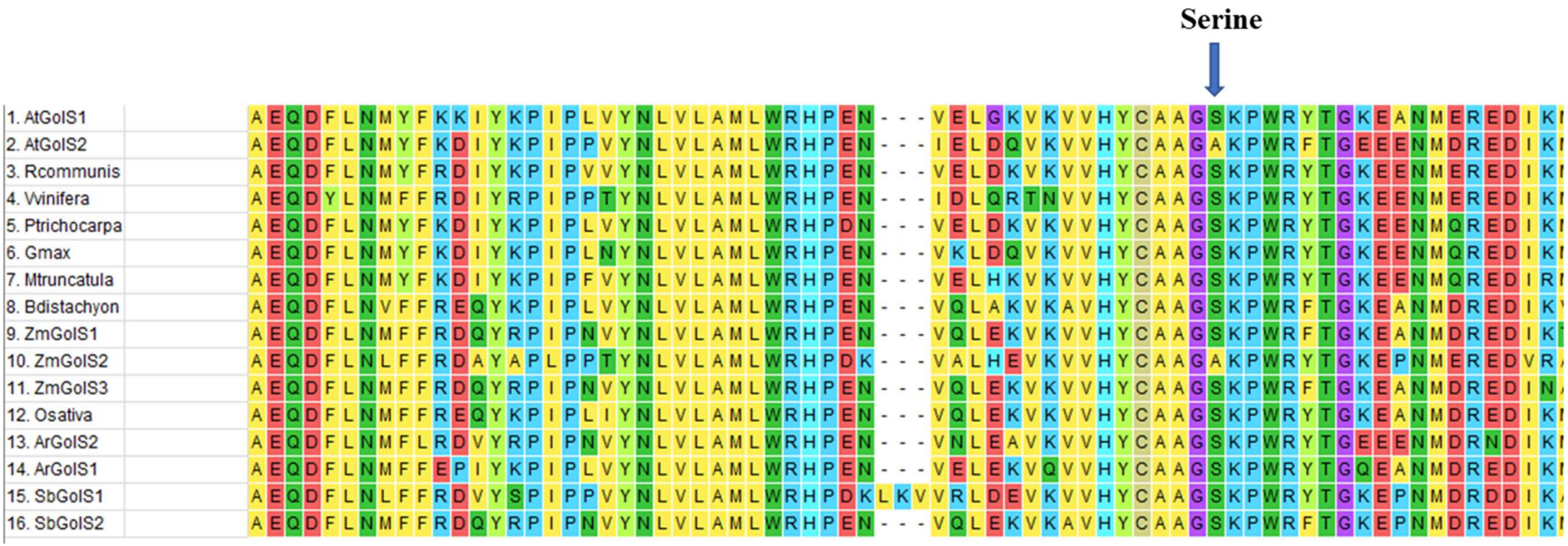

Figure S2.3: A putative serine phosphorylation site at amino acid position 270 in SbGolS1 and SbGolS2.

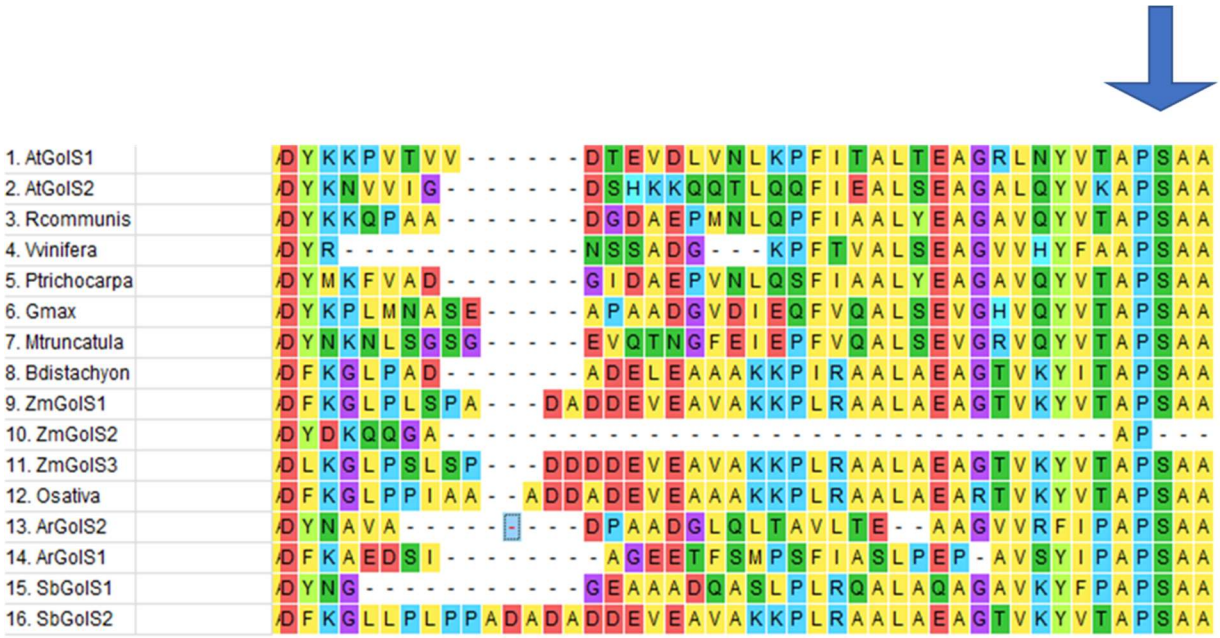

Figure S2.3: C-terminal hydrophobic pentapeptide (APSAA) is a common feature of galactinol synthases in SbGolS1 and SbGolS2.

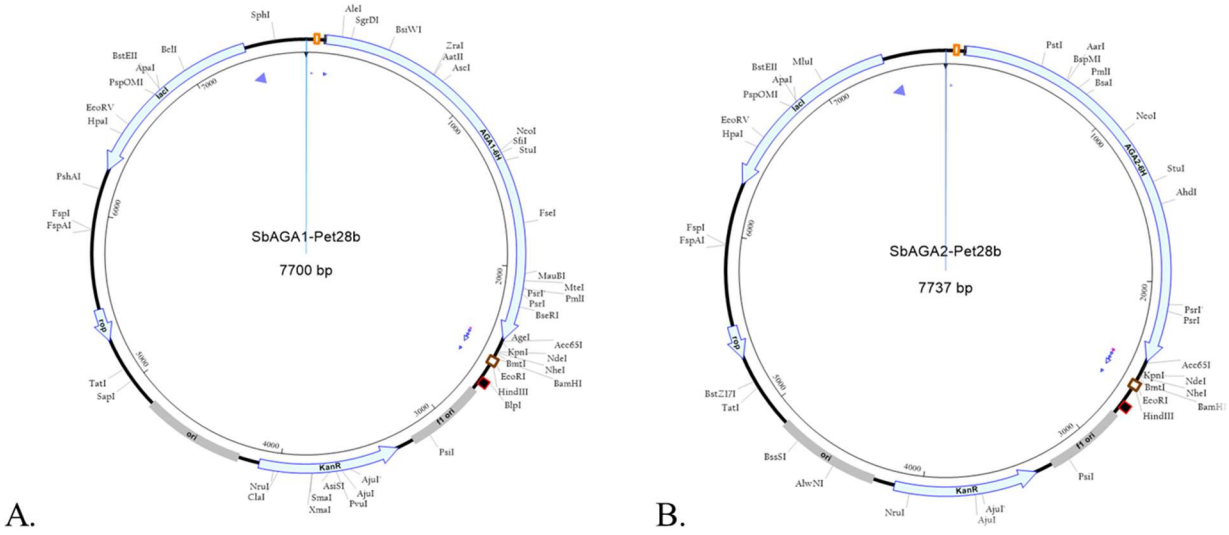

Figure S2.4: Plasmid maps for vector pET-28 b (+) containing the in-frame his-tagged SbAGA1 and SbAGA2 coding sequences. A) SbAGA1-pET-28 b. B)SbAGA2-pET-28 b.

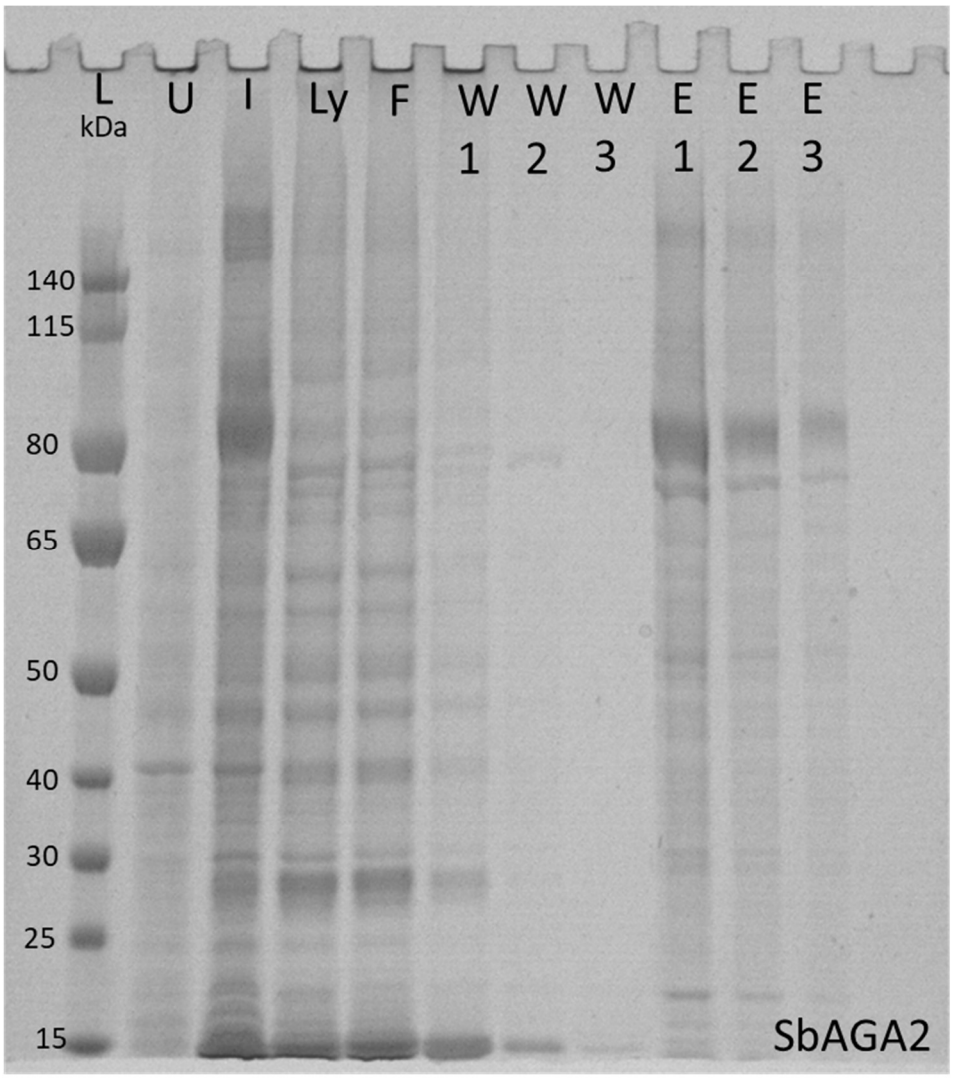

Figure S2.5: SDS-PAGE of SbAGA2 IMAC purification. Lane (L), molecular mass marker (Invitrogen, SeeBlue™ Plus2 Pre-stained Protein Standard)

APPENDIX C

SUPPLEMENTARY TABLES

Table S2.1: Putative galactinol synthases in sorghum bicolor identified through PFAM probing of the sorghum genome

	Locus Name	Transcript Name	PFAM	PFAM_DEF
1	Sobic.009G144200	Sobic.009G144200.1	PF01501	PF01501 - Glycosyl transferase family 8
2	Sobic.004G244100	Sobic.004G244100.1	PF01501	PF01501 - Glycosyl transferase family 8
3	Sobic.004G336401	Sobic.004G336401.1	PF01501	PF01501 - Glycosyl transferase family 8
4	Sobic.004G237800	Sobic.004G237800.1	PF01501	PF01501 - Glycosyl transferase family 8
5	Sobic.004G177000	Sobic.004G177000.1	PF01501	PF01501 - Glycosyl transferase family 8
6	Sobic.008G022500	Sobic.008G022500.1	PF01501	PF01501 - Glycosyl transferase family 8
7	Sobic.001G131900	Sobic.001G131900.1	PF01501	PF01501 - Glycosyl transferase family 8
8	Sobic.010G274800	Sobic.010G274800.1	PF01501	PF01501 - Glycosyl transferase family 8
9	Sobic.001G070700	Sobic.001G070700.1	PF01501	PF01501 - Glycosyl transferase family 8
10	Sobic.002G398400	Sobic.002G398400.1	PF01501	PF01501 - Glycosyl transferase family 8
11	Sobic.010G101400	Sobic.010G101400.1	PF01501	PF01501 - Glycosyl transferase family 8
12	Sobic.001G338400	Sobic.001G338400.1	PF01501	PF01501 - Glycosyl transferase family 8
13	Sobic.002G420100	Sobic.002G420100.1	PF01501	PF01501 - Glycosyl transferase family 8
14	Sobic.006G148200	Sobic.006G148200.1	PF01501	PF01501 - Glycosyl transferase family 8
15	Sobic.001G384200	Sobic.001G384200.1	PF01501	PF01501 - Glycosyl transferase family 8
16	Sobic.003G360500	Sobic.003G360500.1	PF01501	PF01501 - Glycosyl transferase family 8
17	Sobic.004G151400	Sobic.004G151400.1	PF01501	PF01501 - Glycosyl transferase family 8
18	Sobic.005G169500	Sobic.005G169500.2	PF01501	PF01501 - Glycosyl transferase family 8
19	Sobic.008G141800	Sobic.008G141800.1	PF01501	PF01501 - Glycosyl transferase family 8
20	Sobic.009G177200	Sobic.009G177200.1	PF01501	PF01501 - Glycosyl transferase family 8
21	Sobic.010G092400	Sobic.010G092400.1	PF01501	PF01501 - Glycosyl transferase family 8
22	Sobic.001G460000	Sobic.001G460000.1	PF01501	PF01501 - Glycosyl transferase family 8
23	Sobic.002G423600	Sobic.002G423600.1	PF01501	PF01501 - Glycosyl transferase family 8
24	Sobic.003G282600	Sobic.003G282600.1	PF01501	PF01501 - Glycosyl transferase family 8
25	Sobic.007G105700	Sobic.007G105700.1	PF01501	PF01501 - Glycosyl transferase family 8
26	Sobic.002G241100	Sobic.002G241100.1	PF01501	PF01501 - Glycosyl transferase family 8
27	Sobic.002G274700	Sobic.002G274700.1	PF01501	PF01501 - Glycosyl transferase family 8
28	Sobic.001G479800	Sobic.001G479800.1	PF01501	PF01501 - Glycosyl transferase family 8
29	Sobic.001G391300	Sobic.001G391300.1	PF01501	PF01501 - Glycosyl transferase family 8
30	Sobic.001G364700	Sobic.001G364700.2	PF01501	PF01501 - Glycosyl transferase family 8
31	Sobic.006G232000	Sobic.006G232000.1	PF01501	PF01501 - Glycosyl transferase family 8
32	Sobic.001G138200	Sobic.001G138200.1	PF01501	PF01501 - Glycosyl transferase family 8
33	Sobic.003G376700	Sobic.003G376700.1	PF01501	PF01501 - Glycosyl transferase family 8
34	Sobic.006G157800	Sobic.006G157800.1	PF01501	PF01501 - Glycosyl transferase family 8

Table S2.2: Putative raffinose synthases and alpha galactosidases in sorghum bicolor identified through PFAM probing of the sorghum genome

	Locus Name	Transcript Name	PFAM	PFAM_DEF
1	Sobic.005G210100	Sobic.005G210100.1	PF05691	Raffinose synthase
2	Sobic.010G057400	Sobic.010G057400.1	PF05691	Raffinose synthase
3	Sobic.007G219900	Sobic.007G219900.1	PF05691	Raffinose synthase
4	Sobic.002G075800	Sobic.002G075800.1	PF05691	Raffinose synthase
5	Sobic.010G057300	Sobic.010G057300.1	PF05691	Raffinose synthase
6	Sobic.003G052300	Sobic.003G052300.1	PF05691	Raffinose synthase
7	Sobic.001G044800	Sobic.001G044800.1	PF05691	Raffinose synthase
8	Sobic.006G122400	Sobic.006G122400.1	PF05691	Raffinose synthase
9	Sobic.001G019000	Sobic.001G019000.1	PF05691	Raffinose synthase

Table S2.3: Subcellular location prediction of RFO metabolism genes in S. bicolor using DeepLoc and CELLO

	Gene Name	Gene ID	DeepLoc Prediction	CELLO Prediction
1	SbGolS1	Sobic.001G391300	Cytoplasm, Soluble	Cytoplasmic
2	SbGolS2	Sobic.002G423600	Cytoplasm, Soluble	Cytoplasmic
3	SbRS	Sobic.003G052300	Cytoplasm, Soluble	Chloroplast
4	SbSTS	Sobic.005G210100	Plastid, Soluble	Chloroplast
5	SbAGA1	Sobic.002G075800	Cytoplasm, Soluble	Chloroplast
6	SbAGA2	Sobic.001G044800	Cytoplasm, Soluble	Cytoplasmic
7	SbAGA3	Sobic.007G219900	Cytoplasm, Soluble	Cytoplasmic
8	SbAGA4	Sobic.010G057300	Plastid, Soluble	Chloroplast
9	SbAGA5	Sobic.010G057400	Cytoplasm, Soluble	Chloroplast
10	SbAGA6	Sobic.006G122400	Plastid, Soluble	Chloroplast

Table S2.4: Most conserved five motifs of GolS proteins in *S. bicolor* using MEME motif finder tool

Motif	Width	Sequence	Pfam Domain
1	49	YLDADIQVFENIDELFELEKGKFHAVMDCFCEKTWSHTPQYKIGYCQQC	Glyco_transf_8 (PF01501)
2	50	HRRKLRDQGCIVREIEPVYPPENQTQFAMAYYVINYSKLRIWEFVEYERM	Glyco_transf_8 (PF01501)
3	44	HYCAAGSKPWRFTGKEPNMDREDIKALVKKWWDIFDDZTLDFKG	Not found
4	48	YFNAGMFVHEPSLRTAKDLLDALRVTPPTPFAEQDFLNLFFRDQYRPI	Glyco_transf_8 (PF01501)
5	25	RAYVTFLAGDGDYWKGVVGLAKGLR	Not found

Table S2.5: Most conserved six motifs of RS, STS, and AGA proteins in *S. bicolor* using MEME motif finder tool

Motif	Width	Sequence	Pfam Domain
1	50	AYNSLFLGEFMQPDWDMFHSLHPAAEYHGAARAISGGPVYVSDKPGNHDF	Raffinose_syn (PF05691)
2	50	PSIVDWFGWCTWDAFYTDVTPEGVEEGLQSLAEGGVPPRFLIIDDGWQSV	Raffinose_syn (PF05691)
3	50	PDGSILRARLPGRPTRDCLFTDPARDGKSLLKIWNLNKFTGVIGVFNCQG	Raffinose_syn (PF05691)
4	50	ASVARNFPDNGLISCMSHNTDALYSAKQTAVVRASDDFYPRDPASHTVHI	Raffinose_syn (PF05691)
5	50	VHPDKVYEFYNELHSYLASAGVDGVKVDVQNILETLGAGHGGRVALTRKY	Raffinose_syn (PF05691)
6	40	GLLRDVRFMSLFRFKLWWMTQRMGASGRDVPLETQFLLLE	Raffinose_syn (PF05691)

Table S2.6: Sequencing primers for SbAGA1 and SbAGA2 cDNA clones in pET-28b vector

SbAGA1	
AGA1F-1	CGCCTCCGACGACTTCTA
AGA1R-1	GGGTGCAAGCTATGGAACAT
AGA1F-2	GACATGTTCCATAGCTTGCAC
AGA1R-2	AGCTCGAAGTTGTGGTTCC
AGA1F-3	CCCCTTCGACGCCGTCACCG
AGA1R-3	TAGAAGTCCCGCGCCCGCC
SbAGA12	
AGA2F-1	AACCCTTTCGACACCATCAC
AGA2R-1	GTTGGACTCCTCCTTGTTCTC
AGA2F-2	CCTGCTCAAGATCTGGAACGTG
AGA2R-2	CTAGACATCGATCTCCAGGTCCC
AGA2F-3	CGCGACGGCACCAGCCTGCT
AGA2R-3	TCTCCAGCTCGTCGCGGTCGT
AGA2F-4	TCGTCAGGCTGCCC
AGA2R-4	GCCTCCGCGAGCAGTACGC

Table S3.1: qPCR primers for SbGolS1, SbRS, SbAGA1 and SbAGA2

Gene ID	**Gene name**	**Forward primer (5' → 3')**	**Reverse primer (5' → 3')**
Sobic.001G526600	SbUBC	CATGCTGCACATTCGCATAG	AGAGACATGGTCCACAAGAAC
Sobic.001G391300	SbGolS1	AATCTGGGAGTTCGTGGAGTA	CAGTCCATCACCGCGTAAA
Sobic.003G052300	SbRS	CGTCGACAAGATCGTCAACAA	ACAGCATCTCCAGCAAGTG
Sobic.002G075800	SbAGA1	CGCCTCCGACGACTTCTA	GGGTGCAAGCTATGGAACAT
Sobic.001G044800	SbAGA2	AAGGTCTCAAGAGCTTGGC	GTTGGACTCCTCCTTGTTCTC