SHRIVELED SEED (*SVD*), A P1B-TYPE HMA5 TRANSPORTER ATPASE, IS INVOLVED IN COPPER DETOXIFICATION TO CONTROL SORGHUM GROWTH AND DEVELOPMENT

By:

Dimiru Tilahun Tadesse

Bachelor of Science in Plant Science
Bahir Dar University
Bahir Dar, Ethiopia
2011

Master of Science in Genetics /Biology
Wollo University
Dessie, Ethiopia
2014

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements
for the Degree of
DOCTOR OF PHILOSOPHY
December 2020

***SHRIVELED SEED* (*SVD*), A P1B-TYPE HMA5 TRANSPORTER ATPASE, IS INVOLVED IN COPPER DETOXIFICATION TO CONTROL SORGHUM GROWTH AND DEVELOPMENT**

Dissertation approved:

Dr. Million Tadege

_____

Dissertation Advisor

Dr. Liuling Yan

_____

Committee Member

Dr. Michael P Anderson

_____

Committee Member

Dr. Randy D Allen

_____

Outside Committee Member

ACKNOWLEDGEMENTS

First and foremost, I would like to thank the almighty God for giving me the strength and courage to overcome the adversity and hardships throughout my study and to safely pass through this totally precarious pandemic which challenges the entire globe.

This is a good chance for me to express my heartfelt gratitude to my advisor Dr. Million Tadege, who gave me this opportunity to pursue my goals. I am highly grateful for his enormous patience, continuous guidance and evaluation, insightful comments, and encouragement all the way through my study. This would not be possible without his diligent efforts to make me better every time. I also would like to thank my committee members Dr. Randy Allen, Dr. Michael Anderson and Dr. Liuling Yan for their vital comments and suggestions.

I am very honored to take this chance to thank my mother, Mrs. Meseret Birhanie for her unconditional love through my entire life. At various stages of my life, I feel her effectual roles in myself, and they would never go away. I muscularly believe this is the result of her unreserved fortitude and prayer. I also wanted to express my thanks to my sisters; Mrs. Tibebish Tilahun, Mrs. Genet Tilahun, Mrs. Seble Tilahun and Mrs. Baynak Tilahun, and my brothers; Dr. Birkneh Tilahun, Asnake Tilahun and Samuel Tilahun, my sister in laws; Mrs. Freshwork Ayalew, and Mrs. Hanna Andualem for their massive support in all aspects of my life. I also thank you Dr. Worku Legesse, Mr. Cheru Legesse, and all their families for their phenomenal guidance throughout my study. I strongly cherish their wonderful inputs throughout my study.

I am excited to express my sincere gratefulness to my wife Mrs. Senait Tekle for her amazing love and dedicated support in my life. My extended gratitude goes to her family; Insp. Solomon Tekle, Dr. Yohannes Tekle, Mrs. Eden Tekle, Mrs. Yordanos Tekle, Ms. Ruta Tekle, Mrs. Eyerus Tekle, Mr. Belete Tekle and all their loved ones. This would

not have been possible without their support and prayers. Thank you my children Nati, Saron and Eyuel for giving me a remarkable love.

I greatly appreciate many sholars who contributed their great efforts in advising and helping me get here. I would like to acknowledge my M.Sc. advisors Dr. Faris Hailu and Dr. Eleni Shiferaw, and my colleagues; Dr. Hui Wang, Dr. Tezera Watira, Dr. Mussa Adal, Dr. Kelemewerk Geleta, Dr. Seid Hussein, Mr. Beyene Mamo, Dr. Tadele Tadesse, Dr. Habtamu Ayalew, Dr. Shweta Kulkarni, Dr. Marjan Behzadirad, Dr. Jianghua Chen, Dr. Naichong Chen, Dr. Yuqui Li, and Dr. Guan Cong for their support through my study.

My greatest thanks to Mrs. Yetemwork Estifanos, Mrs. Serkalem Laki, Mr. Birhanie Zebemichael, Amy Riley, Mr. John Skinner, and Mr. Marty Waters who were supporting me in different situations. I am deeply indebted to each one of you in some way or the other.

DEDICATION

This work is dedicated to the memory of my father, the late Tilahun Tadesse, and my sister, the late Serkalem Tilahun.

Name: DIMIRU TILAHUN TADESSE

Date of Degree: DECEMBER 2020

Title of Study: *SHRIVELED SEED* (*SVD*), A P1B-TYPE HMA5 TRANSPORTER ATPASE, IS INVOLVED IN COPPER DETOXIFICATION TO CONTROL SORGHUM GROWTH AND DEVELOPMENT

Major Field: CROP SCIENCE

Abstract:

The world population is projected to reach 9.6 billion by 2050, putting strong pressure on food security and agricultural productivity. Understanding plant developmental programs will aid in improving productivity by manipulating genetic pathways and identifying molecular markers that can be used in breeding programs for marker assisted selection. This project generally aims to examine environmentally regulated developmental mutants with the goal of identifying key developmental regulators that integrate environmental signals and understanding the mechanisms of their actions. Specifically, in this study I examined a heavy metal transporter and its impact on sorghum growth and development. Using a forward genetics approach employing a sorghum deletion mutant population, a total of 1,200 M*2* fast neutron irradiated mutant lines were screened in the field to identify mutants. As a result, I found a severely affected developmental mutant named *shriveled seeds* (*svd*). The *svd* mutants display growth defects throughout development from germination to seed filling and appear to be particularly sensitive to heavy metals including copper and manganese. Whole genome sequencing and bioinformatics analysis showed there were seven deletions in *svd* mutant. One of these, the locus *Sobic.006G173800*, was found to cause the *svd* phenotype. Sequence analysis indicated that *SVD* encodes a heavy metal transporter P1B-type ATPase with 8 transmembrane domains, an ATP binding domain and metal binding domains. SVD is primarily localized in the cell membrane and based on the results from yeast 2 hybrid and BiFC analyses, SVD interacts with other metal binding proteins (metallochaperones) SbATX1 and SbFRN3 to extrude copper out of the cell. My gene expression study of *svd* mutants showed there are 951 differentially expressed genes including transporters, kinases, kinesins and histone related genes. A total of 38 genes were selected from the DEG genes based on the fold-change value and their co-expression level with SVD in Phytozome v13.0 for further analysis. This study was focused on understanding the mechanism of *SVD* function and its impact on sorghum development and maturation.

TABLE OF CONTENTS

Chapter **Page**

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

Figure **Page**

CHAPTER I

1. GENERAL INTRODUCTION

1.1. Preface

Sorghum [*Sorghum bicolor* (L.) Moench] is a multipurpose cereal crop grown primarily for food, feed, and fuel worldwide. It is especially suited to arid and semi-arid regions. Due to its efficient C4 photosynthesis and relatively small diploid genome (~730Mb), sorghum is a model for functional genomics studies of sugarcane, *Miscanthus*, maize and other C4 grasses with complex genomes. Its relative resilience to abiotic stresses also makes it an attractive target for crop improvement in suboptimal environments and it could serve as source of important agronomic traits for genetic manipulation. However, despite the publication of a full genome sequence over a decade ago (Paterson et al., 2009), sorghum functional genomics has lagged. One of the reasons for this is the limited availability of mutant resources that can enable the identification of phenotype to genotype associations useful in deciphering gene functions at the molecular level.

For example, rice functional genomics research currently benefits from a number of publicly available mutant resources that include insertion (Tos17, T-DNA, DNA transposon), ethyl methanesulfonate (EMS), and fast neutron deletion mutants (Krishnan et al., 2009), By comparison only EMS mutants have been generated in sorghum (Xin et al., 2008; Blomstedt et al., 2012; Jiao et al., 2016). EMS is an alkylating agent and alkylates guanines (G) in DNA causing base pairing with thymine (T) instead of cytosine (C), which can lead to G/C to A/T transitions (McCallum et al., 2000; Greene et al., 2003; Henikoff and Comai, 2003). Mutations by transversion, deletion or other chromosomal changes are rare and account ~1% in EMS population (Greene et al., 2003). Although the rate of mutation is dependent on dosage and exposure time, EMS generally results in very high mutation frequency with reported rates as high as ~1000/genome in Arabidopsis (Till et al., 2003; Henikoff et al., 2004) and ~7000/genome in sorghum (Jiao et al., 2016). Fast neutron mediated mutagenesis, on the other hand, causes mostly deletions of DNA fragments ranging in size from 1 bp to several megabases (Bruggmann et al., 1996; Li et al., 2001; Li and Zhang, 2002; Men et al., 2002; Belfield et al., 2012). EMS primarily causes point mutations, while fast neutron mainly causes deletion of variable DNA fragments and can complement the EMS mutagenesis efforts. The density of mutation in fast neutron treatment is also dosage dependent but believed to be much lower than that of EMS. Fast neutron bombardment at 60 Gy causes ~18 mutations/genome in Arabidopsis (Belfield et al., 2012).

There are several agronomic challenges associated with sorghum growth including biotic (pests and diseases), and abiotic (drought, salinity, heat, etc) stresses that globally affect sorghum production (Morgan and Finlayson, 2000; Smith and Frederiksen, 2000; Song et al., 2007; Zheng, 2010; Sathya et al., 2016). As a relatively abiotic stress resistant and C4 type

photosynthetic plant, sorghum provides several advantages to study the mechanism of stress resilience and photosynthetic efficiency. A suitable mutant resource in sorghum would help facilitate sorghum molecular genetics and genomics studies. In this study, I report a comprehensive investigation of fast neutron mutagenesis in the sorghum reference genome BTx623, mutant screening, and characterization of the sorghum *svd* mutant. The study covers three sections:

1) The first part deals with the generation of a mutant population with fast neutron bombardment, and whole-genome sequencing to identify a specific mutation that is associated with a strong developmental phenotype. This section begins with irradiating sorghum seeds with fast neutrons followed by selfing and screening strategies to obtain mutants that are responsive to different environmental signals. BTx623 sorghum seeds were mutagenized at three doses (25Gy, 30Gy, and 35Gy) and the seeds were grown in the field (Wes Watkins agricultural research station in Lane, Oklahoma) for selfing. Individual plants were collected and screened the next season by planting ~15 seeds from each line. After screening, the target mutant was backcrossed to the parental line and genotyping was performed followed by sampling from M2 plants for whole-genome sequencing with 30X coverage. Next, computational methods were used to identify the cause of the mutant phenotype. Quality control analysis was performed at different stages. Trimming was performed using Trimmomatic v0.39, alignment to the reference genome and identification of mutations was done using FNBtools. From three sequenced mutant lines, FN-556, FN-264, and FN-987 (*svd* mutants). FN-987 had a unique seed phenotype, and the fewest number of deletions. As a result, this mutant was selected for further characterization in this thesis.

2) The second section encompassed the characterization of the *svd* mutants and the understanding of the molecular function of SVD in sorghum development. To confirm which loci affected the *svd* mutant phenotype, sequencing of segregating lines showed that plants that showed the *svd* phenotype contained a deletion at a specific locus. However, other plants from the same line showing no phenotype did not have a deletion at that specific locus. Yeast complementation was performed to determine which of sorghum *SVD lines* can complement the *ccc2* yeast mutant, and which were unable to grow in copper and iron-deficient media. Phylogenetic analysis and multiple sequence alignments were performed to study the similarity of *SbSVD* with other homologous genes. A high throughput ionome study was performed to compare the total ionome composition of *svd* mutants and WT (BTx623) under normal conditions. Subcellular localization of the SVD protein was determined using SVD-GFP fusion protein infiltrated into tobacco leaves and imaged after 48-60 hrs. To further characterize SVD lines, potential partner proteins were identified using STRING, a protein-protein interactors prediction program. This program was designed to predict a functional protein association network for more than 2000 organisms based on the sequence homology of previously annotated model organisms. Two of the predicted interactors were confirmed by yeast two hybrid (Y2H) and bimolecular fluorescence complementation (BiFC) analyses. This section also included analysis of the induction of *SbSVD* expression by heavy metals and treatment of the *svd* mutant with various levels of heavy metals in an attempt to understand SVD's function.

3) The third section is comprised of transcriptome analysis and gene expression study of the *svd* mutant. RNA-seq was performed on three-week-old seedlings. The experiment was executed to study the transcriptome changes between the wild type and the *svd* mutant. Bioinformatic analysis showed that there were 951 differentially expressed genes (DEGs) in

the mutant lines as compared to the wild type (WT). Most of these DEGs were upregulated genes and grouped primarily under the kinase, kinesins and histone gene families. Selection of important genes from the DEG pool was done based on their fold change value, expression and co-expression pattern of the DEGs within the *SVD lines*. A total of 38 genes were selected for further study and validation using RNA-seq and quantitative real time PCR. The selected DEGs from the RNA-seq study were used to study their relationship with *SVD* in cellular trafficking.

The last chapter is a generalized conclusion and reflects on some of the major findings in terms of overall significance and perspectives.

1.2. References

Belfield, E.J., Gan, X., Mithani, A., Brown, C., Jiang, C., Franklin, K., Alvey, E., Wibowo, A., Jung, M., Bailey, K., Kalwani, S., Ragoussis, J., Mott, R., and Harberd, N.P. (2012). Genome-wide analysis of mutations in mutant lineages selected following fast-neutron irradiation mutagenesis of Arabidopsis thaliana. Genome Res **22,** 1306-1315.

Blomstedt, C.K., Gleadow, R.M., O'Donnell, N., Naur, P., Jensen, K., Laursen, T., Olsen, C.E., Stuart, P., Hamill, J.D., Moller, B.L., and Neale, A.D. (2012). A combined biochemical screen and TILLING approach identifies mutations in Sorghum bicolor L. Moench resulting in acyanogenic forage production. Plant Biotechnol J **10,** 54-66.

Bruggmann, E., Handwerger, K., Essex, C., and Storz, G. (1996). Analysis of Fast neutron generated mutants at the arabidopsis HY4 locus. The Plant Journal. 58: 668-681 **10,** 755-760.

Greene, E.A., Codomo, C.A., Taylor, N.E., Henikoff, J.G., Till, B.J., Reynolds, S.H., Enns, L.C., Burtner, C., Johnson, J.E., Odden, A.R., Comai, L., and Henikoff, S. (2003). Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in Arabidopsis. Genetics **164,** 731-740.

Henikoff, S., and Comai, L. (2003). Single-nucleotide mutations for plant functional genomics. Annu Rev Plant Biol **54,** 375-401.

Henikoff, S., Till, B.J., and Comai, L. (2004). TILLING. Traditional mutagenesis meets functional genomics. Plant Physiol **135,** 630-636.

Jiao, Y., Burke, J., Chopra, R., Burow, G., Chen, J., Wang, B., Hayes, C., Emendack, Y., Ware, D., and Xin, Z. (2016). A Sorghum Mutant Resource as an Efficient Platform for Gene Discovery in Grasses. Plant Cell **28,** 1551-1562.

Krishnan, A., Guiderdoni, E., An, G., Hsing, Y.I., Han, C.D., Lee, M.C., Yu, S.M., Upadhyaya, N., Ramachandran, S., Zhang, Q., Sundaresan, V., Hirochika, H., Leung, H., and Pereira, A. (2009). Mutant resources in rice for functional genomics of the grasses. Plant Physiol **149,** 165-170.

Li, X., and Zhang, Y. (2002). Reverse genetics by fast neutron mutagenesis in higher plants. Funct Integr Genomics **2,** 254-258.

Li, X., Song, Y., Century, K., Straight, S., Ronald, P., Dong, X., Lassner, M., and Zhang, Y. (2001). A fast neutron deletion mutagenesis-based reverse genetics system for plants. Plant J **27,** 235-242.

McCallum, C.M., Comai, L., Greene, E.A., and Henikoff, S. (2000). Targeted screening for induced mutations. Nat Biotechnol **18,** 455-457.

Men, A.E., Laniya, T.S., Searle, I.R., Iturbe-Ormaetxe, I., Gresshoff, I., Jiang, Q., Carroll, B.J., and Gresshoff, P.M. (2002). Fast Neutron Mutagenesis of Soybean (Glycine soja L.) Produces a Supernodulating Mutant Containing a Large Deletion in Linkage Group H. Genome Letters **1,** 147-155.

Morgan, P., and Finlayson, S. (2000). Physiology and genetics of maturity and height. In Sorghum: Origin, History, Technology, and Production. Edited by Smith CW, Frederiksen RA. New York: Wiley Series in Crop Science; 2000:240-242.

Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., Schmutz, J., Spannagl,

M., Tang, H., Wang, X., Wicker, T., Bharti, A.K., Chapman, J., Feltus, F.A., Gowik, U., Grigoriev, I.V., Lyons, E., Maher, C.A., Martis, M., Narechania, A., Otillar, R.P., Penning, B.W., Salamov, A.A., Wang, Y., Zhang, L., Carpita, N.C., Freeling, M., Gingle, A.R., Hash, C.T., Keller, B., Klein, P., Kresovich, S., McCann, M.C., Ming, R., Peterson, D.G., Mehboob ur, R., Ware, D., Westhoff, P., Mayer, K.F., Messing, J., and Rokhsar, D.S. (2009). The Sorghum bicolor genome and the diversification of grasses. Nature **457,** 551-556.

Sathya, A., Kanaganahalli, V., Rao, P.S., and Gopalakrishnan, S. (2016). Cultivation of Sweet Sorghum on Heavy Metal-Contaminated Soils by Phytoremediation Approach for Production of bioethanol. In Bioremediation and Bioeconomy, pp. 271-292.

Smith, C.W., and Frederiksen, R.A. (2000). Sorghum: Origin, History, Technology, and Production. Eds Smith CW, Frederikson RA (John Wiley & Sons, New York), pp 191–223.

Song, B., Wang, G., and Yuan, J. (2007). Measurement and characterization of singlet oxygen production in copper ion-catalyzed aerobic oxidation of ascorbic acid. Talanta **72,** 231-236.

Till, B.J., Reynolds, S.H., Greene, E.A., Codomo, C.A., Enns, L.C., Johnson, J.E., Burtner, C., Odden, A.R., Young, K., Taylor, N.E., Henikoff, J.G., Comai, L., and Henikoff, S. (2003). Large-scale discovery of induced point mutations with high-throughput TILLING. Genome Res **13,** 524-530.

Xin, Z., Wang, M.L., Barkley, N.A., Burow, G., Franks, C., Pederson, G., and Burke, J. (2008). Applying genotyping (TILLING) and phenotyping analyses to elucidate gene function in a chemically induced sorghum mutant population. BMC Plant Biol **8,** 103.

Zheng, S.J. (2010). Crop production on acidic soils: overcoming aluminium toxicity and phosphorus deficiency. Ann Bot **106,** 183-184.

CHAPTER II

2. FAST NEUTRON MUTAGENESIS AND IDENTIFICATION OF MUTANTS IN SORGHUM (*Sorghum bicolor* L.)

2.1. Introduction

Mutagenesis is a potent tool for producing genetic variability in different organisms (Kumawat et al., 2019), and a fundamental approach in biology to identify genes and study gene functions (Tadege et al., 2009). There are different types of mutagenesis techniques. Chemical mutagenesis, physical mutagenesis, and insertional mutagenesis are among the forward genetic screening schemes. Several of these methods were optimized for different mutation rates and frequency depending on various conditions including the type of the organism. Of those, T-DNA tagging, transposon tagging, EMS, gamma-rays and fast neutron radiation were among the most commonly used methods (Koornneef et al., 1982; McCallum et al., 2000; Li et al., 2001; Li et al., 2002; Tadege et al., 2005; Wu et al., 2005; Kim et al., 2006; Rogers et al., 2009; Bolon et al., 2011). Consequently, there are different types of mutations that could occur due to these different mutagenesis approaches. For instance, EMS mutagenesis can cause three different types of mutations..

These include: 1) nonsense mutations, which can result from single base changes that convert an amino acid codon into a stop codon, 2) missense mutations, which result in single base changes that alter the amino acid encoded by a particular codon. These can further be categorized into those resulting in conservative and non-conservative substitutions. Conservative substitution causes an amino acid change; where the properties of the new amino acid is the same. On the other hand, non- conservative substitution results in a translation into an amino acid with different properties, and 3) silent mutations, which result when a single base change to a codon does not alter the encoded amino acid. These changes are usually, but not exclusively, the result of mutations that alter the third base of a codon. Generally, nonsense mutations most likely alter gene function underscoring the importance of the expected frequency of each class of mutation (McCallum et al., 2000). However, EMS mutagenesis results in fewer strand breaks that results in either an inversion or deletion mutants (Koornneef et al., 1982; Kim et al., 2006). These types of mutations results in less or no effect on the activity of the mutagenized protein where only about 5% of outcomes cause a knockout mutation in Arabidopsis (McCallum et al., 2000). On the contrary, fast neutron mutagenesis gives rise to deletions where every event contains a knockout allele due to frame shifting effect (Henikoff and Comai, 2003). Although, other types of mutations are also common in fast neutron radiation, 81% of the mutations are from deletions, which range from a single base pair to large deletion fragments of up to 60kb or more (Rogers et al., 2009; Bolon et al., 2011; Islam et al., 2019).

Despite being one of the earliest and most efficient methods, fast neutron mutagenesis was considered labor intensive due to the difficulty of detecting large deletion

fragments (Emmanuel and Levy, 2002). Nevertheless, the advancement of mutation detection methodology such as pooling DNA samples and sequencing of mutants has increased the efficacy of radiation mutagenesis, and consequently there has been a renewed interest in fast neutron radiation combined with next generation sequencing methods. However, radiation mutagenesis still requires an optimized screening strategy especially in cases where conspicuous phenotypes are not readily visible. After irradiation, the first step is growing the plants in the field and identifying and collecting putative mutants independently after selfing which is the next critical step to avoid any contamination with wild type and other mutants of little interest. In addition, downstream screening of the selfed lines at M2 stage depends of the objectives of the research project that could be applied by phenotype screening (Meinke et al., 2003), composition screening (Bolon et al., 2011), or other methods like PCR (Li et al., 2002), ionome screening (Baxter et al., 2008; Salt et al., 2008; Baxter, 2009) and MutMap (Abe et al., 2012; Kumawat et al., 2019).

Mutation detection strategies for fast neutron mutants have improved over the years. Some of the common strategies include, the PCR approach, restriction enzyme, and MutMap. The PCR method (Li et al., 2002), used a PCR by altering the extension time interval which effectively deletes alleles that efficaciously competes for amplification of deleted genes from very large population of mutants. However, the problem with this PCR strategy is that when the deletion only includes a small proportion of the amplified region ranging form one to few basepairs, it is undifferentiable by changing the amplification time as both the wildtype and mutant amplicons have approximately the same size and changing the extention time is negligble in this case (Rogers and Oldroyd, 2008). The second method is using restriction enzyme destruction depending on the digesting of highly complex DNA

pools with a restriction enzyme which cuts once within the target sequence leading to destruction of the template. However, in the case of mutants it is presumed that the deletion misplaces the restriction enzyme places is not intact which leads the assumption that the mutant DNA will not be digested by the restriction enzyme and will get the chance to be amplified (Edgley et al., 2002; Rogers and Oldroyd, 2008). The third and recently common approach in identifying the fast neutron mutants is, the MutMap approach; considered an efficient system in mutation identification in small genomes (Kumawat et al., 2019). The MutMap approach uses whole genome sequencing of a pooled DNA from segregating populations of mutagenized plants (Abe et al., 2012). The mutant line is first back crossed to the wild type to create F1 generation and then selfed to form an F2 population. The F2 population is then further screened repeatedly for visible mutant phenotype. Phenotypic evaluation of F2 progenies gives two contrasting bulks representing either wild type phenotype or mutant phenotype. Then sequencing DNA samples from F2 plants and mapping to the reference genome gives SNPs and InDels (Abe et al., 2012; Kumawat et al., 2019).

Although there are many research efforts to create mutant collections and databases by EMS mutagenesis (Xin et al., 2008; Blomstedt et al., 2012; Jiao et al., 2016), the EMS technique faces one severe drawback due to the EMS method causing too many single nucleotide substitutions (Mohd-Yusoff et al., 2015), which results in a very high mutation density (56-100 fold more than fast neutron) (Martin et al., 2009) creating another challenge to identify mutations by whole genome sequencing (WGS) in EMS populations. Thus, it is very important to create a feasible method utilizing forward genetic strategy in

sorghum which will open the door for functional genomic analysis of sorghum and other large genome crop species genes.

2.2. **Objectives**

The overall objective of this study was to create a sorghum mutant resource using fast neutron mutagenesis to facilitate sorghum functional genomics and complement current EMS mutant lines.

Specific objectives:

1. To develop a sorghum deletion mutant population that will be used as a resource for functional genomics studies.

2. To screen the sorghum mutant population under field conditions and identify developmental mutants.

3. To identify and characterize mutations for selected mutants by sequencing in comparison to the whole genome sorghum sequence .

2.3**. Materials and Methods**

2.3.1. Plant material and growth condition

Plant materials used in this study were sorghum, [*Sorghum bicolor* L.) Moench], BTx623 seeds, planted in the field at the Wes Watkins Agricultural research station in Lane, OK, and in green house and in growth chamber conditions in the institute for Agricultural Biosciences, Ardmore, Ok. In the field, plants were grown from late April to August during the growing season with moisture provided via irrigation. In the greenhouse, plants were

grown in one gallon pots under long-day (LD) conditions with 16/8 hours light/dark cycle at 28-32˚C temperature, and in growth chamber under long-day conditions of 16/8 hours light /dark cycle at 25-28◦C, 70-80% relative humidity, and 150 μmol.m^2 light intensity.

2.3.2. Mutant screening and phenotyping

The fast neutrons used for sorghum mutagenesis were generated in a nuclear reactor at the University of California, McClellan Nuclear Research Center in Davis, California and the Hungarian Academy of Sciences Center for Energy Research in Budapest, Hungary. Mutant seed propagation and screening was conducted in the field in Lane, Oklahoma. For mutant confirmation and characterization, plants were grown in a greenhouse in one-gallon pots under long-day conditions with 16/8 hrs light/dark cycle at 28-32◦C temperature.

2.3.3. DNA extraction and whole genome sequencing

Genomic DNA was extracted using DNeasy Plant Mini Kit (Qiagen). Library preparation with 350 bp insert size and WGS (Illumina PE150) at 30x genome coverage and an output quality of Q30 ≥ 80% was performed using the Illumina Sequencing system (Novogene Corp, CA).

2.3.4. Sequence analysis and deletion identification

To keep only high-quality reads for analysis, we used Trimmomatic v0.39 (Bolger et al., 2014) to trim the NGS reads (threads 20 ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36). FNBtools was used to identify unique deletions in sorghum mutant samples (Sun et al., 2018). All High-quality trimmed reads were first aligned to *Sorghum bicolor* v3.1.1 reference genome using fnbalign function in FNBtools. After screening of mutant and heterozygous phenotypes,

we considered the mutant phenotype as a sample line, and the heterozygous and wild type as control line. We used fnbscan function in FNBtools to filter out both homozygous and heterozygous deletions in control samples to identify unique homozygous deletions in the *svd* mutant.

2.4. Results

2.4.1. Fast neutron mutagenesis and mutant screening

With the goal of establishing a deletion mutant functional genomics resource for sorghum, this work was initiated by mutagenizing the sorghum genome with fast neutron bombardment. Seeds of the reference variety BTx623 were irradiated with different doses of fast neutron (20-60Gy with 5Gy increments) initially at the Hungarian Academy of Sciences Center for Energy Research in Budapest (400 M2 lines), and later at the UC Davis McClellan Nuclear Research Center (6000 M2 lines). Three doses of fast neutron 25Gy, 30Gy, and 35Gy were selected for final treatment optimized for mutation frequency, survival and fertility by testing plants in the field and greenhouse up to M3 seed production. This combination treatment had M1 seed germination rate of 80-90% but of these 10-15% of the did not reach maturity and another 40% of matured M1 plants were sterile. Under irrigated field conditions in Lane, Oklahoma, less than 20% of selfed M1 plants were fertile and produced M2 seeds. The seeds from each M1 plant were collected separately and designated M2 family seeds (Fig. 1). Using this approach, we generated approximately 6,000 M2 families over the past several years with variable seed number in each family. Approximately 15 seeds from each M2 family, were planted in the field to identify mutants and catalog phenotypes (Fig. 1).

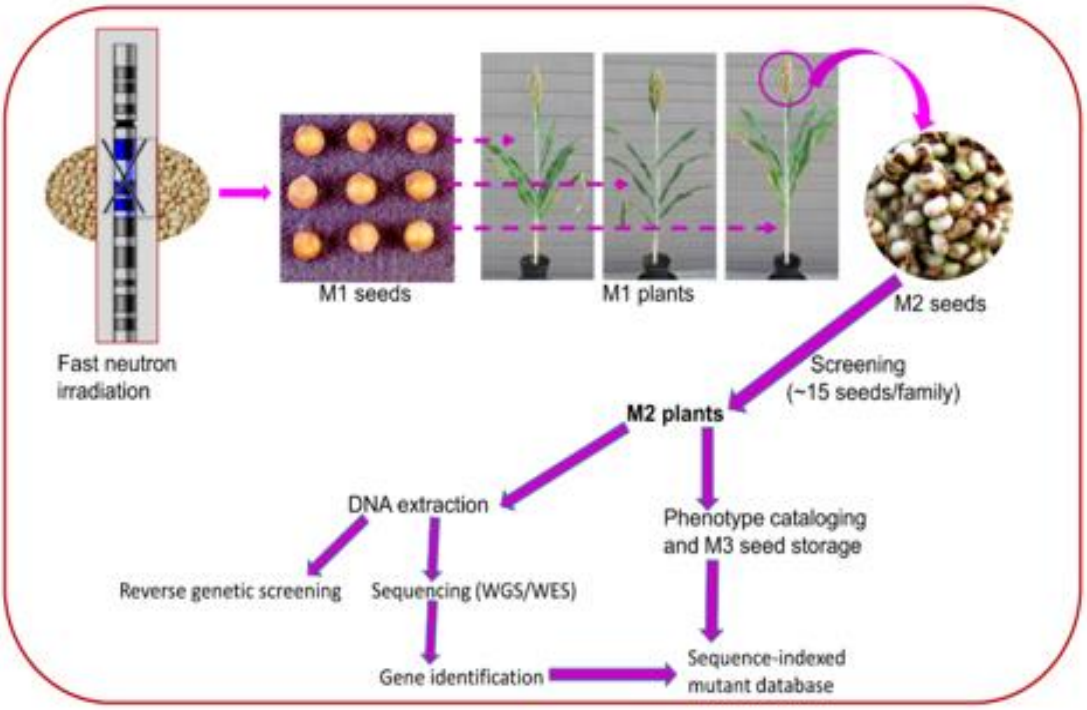

Fig. 1) Schematic representation of sorghum fast neutron mutant generation and phenotype screening pipeline

The mutagenized seeds were labeled M1 seeds and planted individually in field or greenhouse to give M1 plants. M1 plants were self-pollinated in sorghum head bags and seeds from each plant were collected separately and labeled M2. Approximately 15 seeds from each M2 family were planted in the field to screen for visible above ground phenotypes in M2 plants. Genomic DNA was extracted from M2 family plants as pooled leaves for sequencing and reverse screening. M3 seeds were collected from the phenotypic mutants and labeled by their mutant phenotype and the rest were pooled together as M3 seeds.

Out of 1,200 M2 families (~18,000 seeds) planted in the field, 211 above ground developmental mutants including albino, pale green leaf, early flowering, late flowering, narrow leaf, broad leaf, long leaf, coiled leaf, brown midrib, pink stem, dwarf, thick stem, thin stem, multiple tillers, branched panicle, seed color, shriveled seed, and homeotic conversion of floral organs were identified. Several representative phenotypes obtained from the field grown mutant screening are shown in Fig. 2.

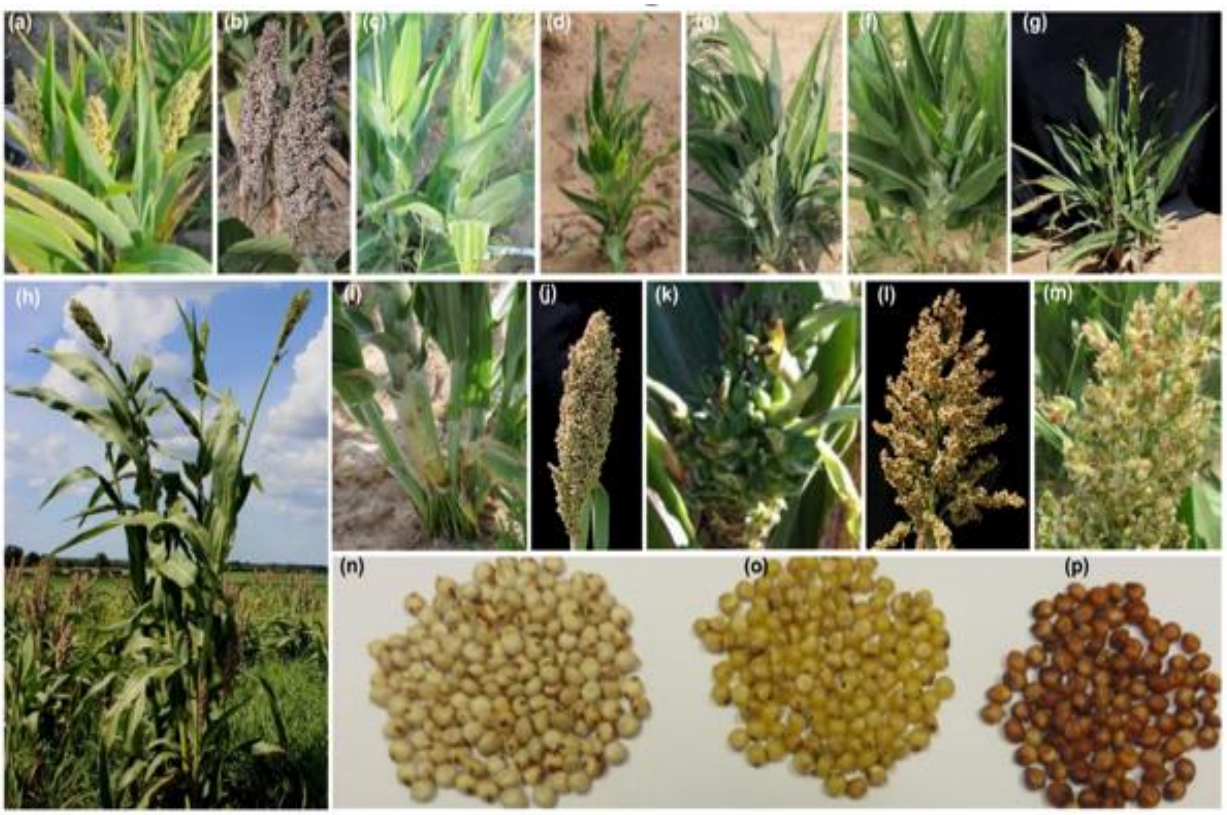

Fig. 2) Representative phenotypes of field grown independent M2 lines

(a-i), the same age vegetative organ mutants; WT (a), early flowering (b), late flowering (c), dwarf, stunted growth, and late flowering (d), dwarf, bushy, late flowering (e), dwarf, late flowering (f), dwarf, bushy, normal flowering (g), extremely tall (h), multiple tillers (i). (j-p), reproductive organ mutants; WT panicle control (j), homeotic mutant showing spikelet converted to leaves (k), branched panicle with normal fertility (l), branched panicle with highly reduced fertility (m), WT seed color control (n), creamy seeds (o), brown seeds (p).

Eleven of these selected dwarf, thin stem and late flowering mutant phenotypes were confirmed under greenhouse growth conditions (Fig. 3, Fig. 4).

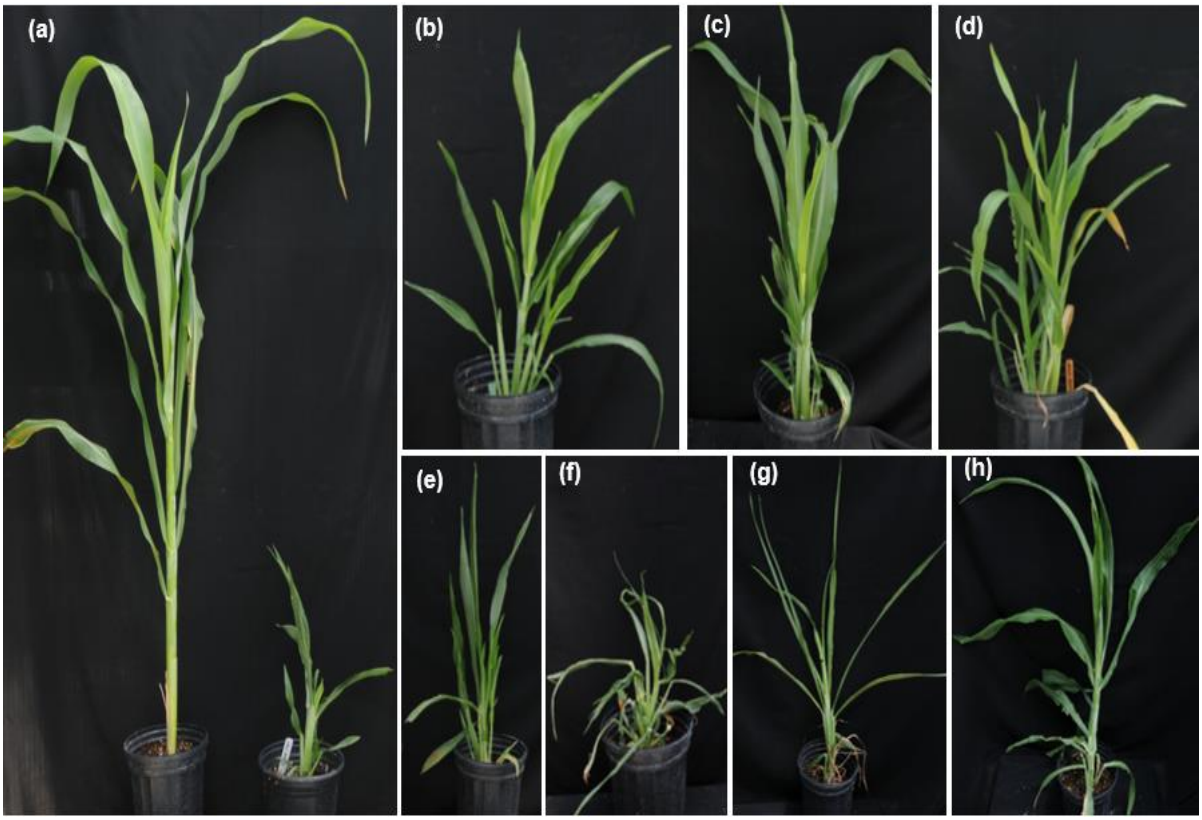

Fig. 3) Phenotypes of representative field isolated plant height and leaf mutants confirmed under greenhouse conditions

(a) WT control (left) and extreme dwarf (right), (b) Dwarf and bushy, modestly narrow leaves, (c) Dwarf with normal leaves, (d)Dwarf and bushy, (e) dwarf, bushy, very narrow, darker green and erect leaves, (f) Very dwarf, bushy, coiled and very narrow leaves, (g) Dwarf and extremely narrow leaves, (h) Moderately dwarf and wilting leaves.

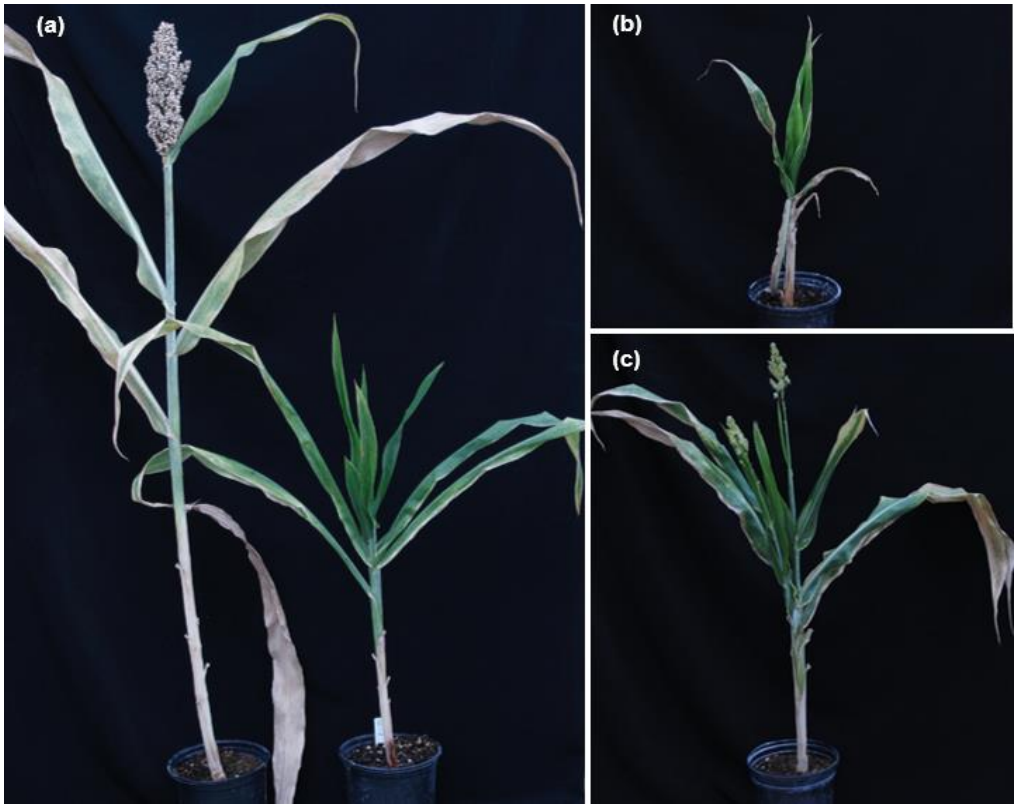

Fig. 4) Examples of developmentally delayed late flowering mutants confirmed in greenhouse shown at maturity

(a) WT control 92 days after germination (left) and dwarf, retarded growth, very late flowering (right), (b) Dwarf and growth arrested, extremely delayed flowering, (c) Dwarf and moderately late flowering (heading ~85 days after germination, while WT is heading ~62 days).

2.4.2. Developmental phenotypes of the *svd* mutant

The *svd* mutant displayed severe defects throughout developmental stages from seed germination to grain filling (Fig. 5). The mutant is significantly delayed in germination and root elongation is restricted compared to the WT BTx623. While the WT started to germinate within 48 hours of imbibition, the *svd* seeds remained dormant for three days (Fig. 5a-b). After 2-3 weeks of growth on standard MS media, the *svd* mutants produced very short roots with few, if any lateral roots while the WT roots proliferated aggressively

with extensive lateral root development (Fig. 5c-e), and after 4-6 weeks of growth in soil, *svd* plants achieved only about a third of the plant height, root length and fresh weight of the WT (Fig. 5f). The leaves were small and appeared necrotic, stressed, and bleached, especially at the tips when grown in soil (Fig. 5f). Seeds of *svd* mutants were small and shriveled (Fig. 5g) with a 1000 seed weight of approximately a third of the WT (Fig. 5h). Compared to WT, root development in *svd* was severely affected when grown in ½ MS-media for three weeks (Fig. 6). These observations indicate that the *svd* mutant is affected in plant growth at all stages of development including grain filling.

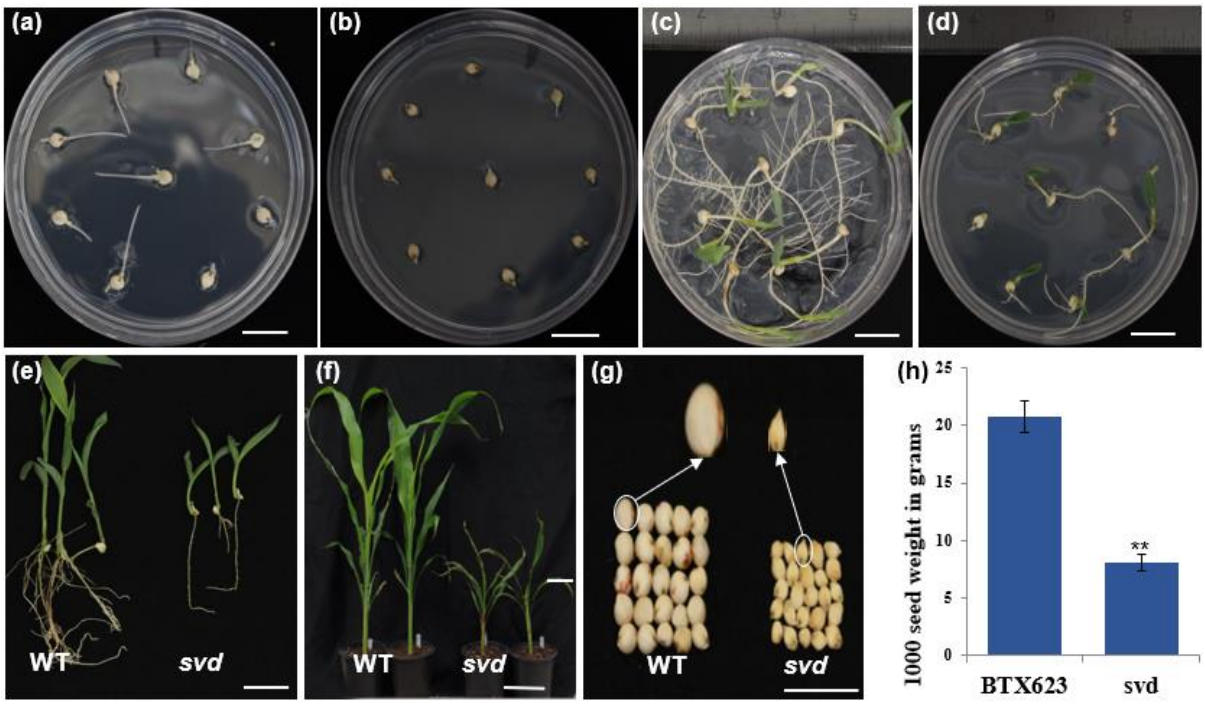

Fig.5) *svd* mutation caused a shriveled seed in grain sorghum BTx623 genotype

(a - b) Germination rate of BTx623 WT lines and svd mutant lines, respectively, which were incubated in the dark for 72 hours with water only, c - d) Phenotypes of BTx623 WT lines compared with svd mutant lines grown in MS media for two weeks. e, f, and g) Phenotypes of the svd mutants and Wild types grown in the greenhouse under long-day (LD) condition for four weeks six weeks and eight weeks, respectively. h) Seed phenotype of svd mutant and BTx623 wild type seeds.

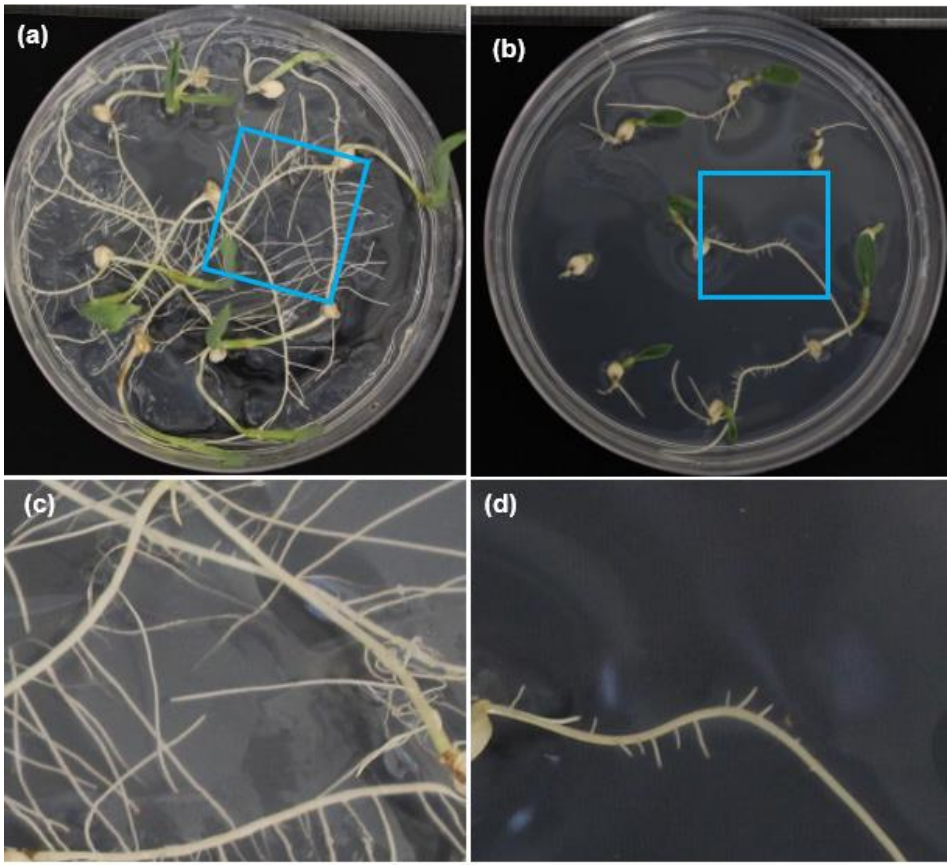

Fig.6) Root elongation and lateral root emergence were severely compromised in svd mutant

(a,b) Root elongation during growth on MS media for 2 weeks of WT (a), and svd mutants (b), (c) higher magnification inset in (a) showing numerous and long lateral roots, (d) higher magnification of inset in (b) showing very few and very short lateral roots.

Root length, plant height, and above ground total fresh weight were similarly significantly affected (Fig. 7). Quantitative traits were scored for plant height (PH), root length (RL), thousand seed weight (TSW), and fresh biomass weight (BW). The average TSW of *svd* was 6.92 grams compared to 20.8 grams in wild type BTx623 (n=3) (Fig.7). Plant height, fresh weight, and root length also showed significantly reduced value ($P < 0.05$, n=10).

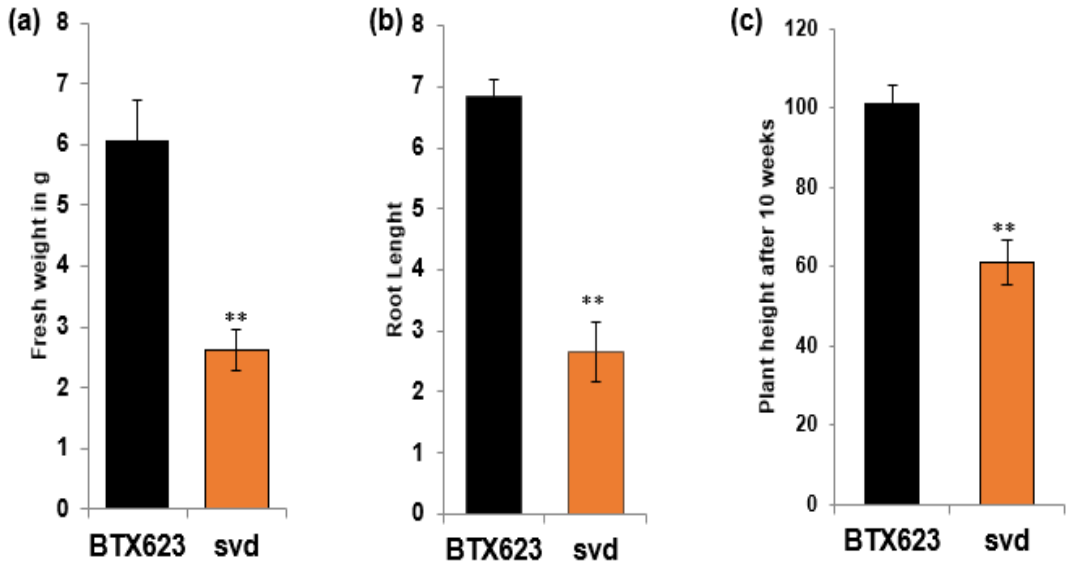

Fig. 7) Yield and yield related traits showed a significant difference between *svd* and BTx623 WT genotype

a) Fresh biomass weigh difference for the wild type and the svd mutant c) Mean root length difference for the wild type and the svd mutant e) Mean root length difference for the wild type and the svd mutant , number of plants taken for plant height , root length and fresh biomass weight were 10 and p=0.05

2.4.3. Whole genome sequencing and identification of the *SVD* gene

To identify the mutated *svd* gene and assess the number of mutations per genome in the population, WGS was performed on *svd* mutant and another two loss-of-function mutant lines with the parental BTx623 using Illumina sequencing. Genomic DNA was extracted from homozygous M3 seedlings and whole genome sequencing (WGS) was carried out at 30X genome coverage to identify mutations. Sequence analysis using FNBtools (Sun *et al*., 2018) was performed based on the assumption that the candidate mutation must be homozygous in the mutant locus of interest. Using this system, a wide range of variation in the number of candidate deletions were identified in the three mutant lines; 7 in line FN-987 (*svd*), 39 in line FN-556 and 5,379 in line FN-264 including exon and non-exon

deletions. Due to this wide variation in the number of unique deletions, determining the average number of deletions per genome in mutant population needs further sequencing and analysis of more lines. The *svd* mutant had seven homozygous and unique deletions (Table 1). Among the seven deletions in the *svd* mutant a two base pair "GC" deletion at chromosome 06, position 52982340-52982342 was in the coding region and located in exon 5 of an annotated gene (Sobic.006G173800). The other six deletions were in intergenic regions at least 7 kb away from protein coding genes (Fig. 8; Table 1).

Table 1: Table showing the deletions across different chromosomes, the break point starting points, and the break points end points.

DEL #	Chr	Breakpoint Start	Breakpoint End	Del-Leng	Gap Starts_ position	GapEnd_ position	Del_ mutant	Del_ control	Homo_ Unique
1	Chr02	12781839	12781840	1	12781839	12781840	Yes	No	Yes
2	Chr05	47512023	47512024	1	47512023	47512024	Yes	No	Yes
3	Chr06	15786510	15786511	1	15786510	15786511	Yes	No	Yes
4	Chr06	37621748	37621749	1	37621748	37621749	Yes	No	Yes
5	Chr06	52982340	52982342	2	52982340	52982342	Yes	No	Yes
6	Chr09	26192583	26192598	15	26192586	26192598	Yes	No	Yes
7	Chr09	27438568	27438569	1	27438568	27438569	Yes	No	Yes

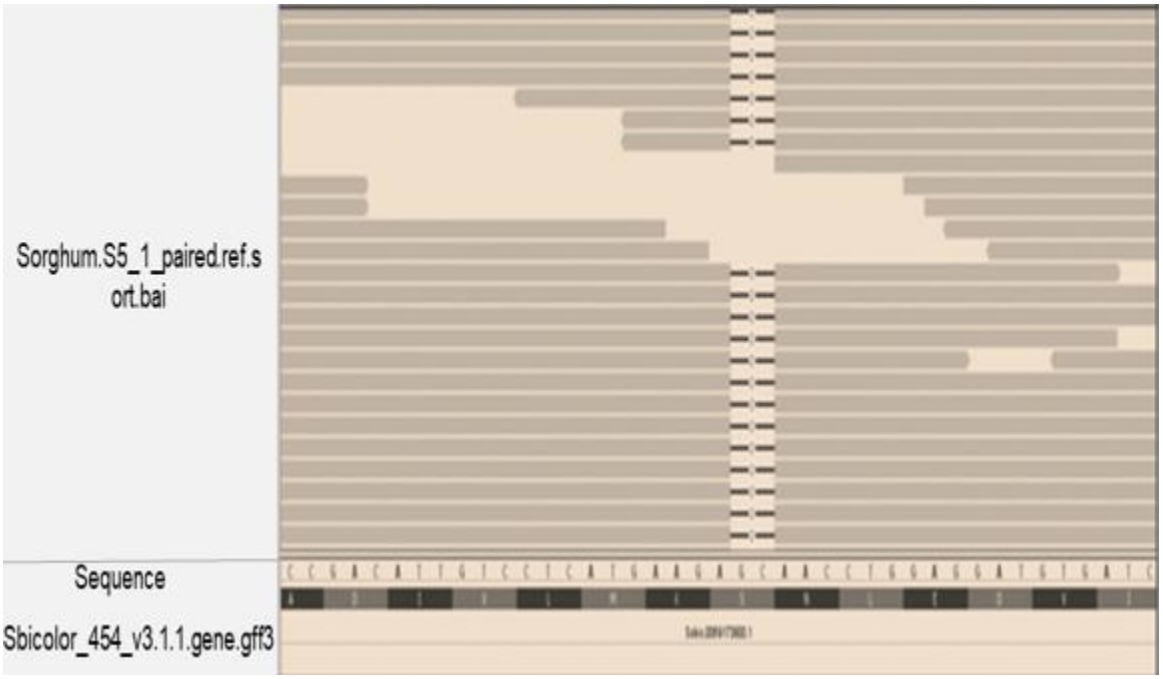

Fig. 8) Deletion Sobic.006G173800 in IGV (Integrated Genome Viewer)

The discovery of putative variants from WGS aligned data was viewed by integrated genome viewer (Fig. 8). The coverage plot and alignment was performed using the integrated genome viewer version 2.3 (Robinson et al., 2011; Thorvaldsdottir et al., 2013; Robinson et al., 2017). The reference sequence Sbicolor_454_V3.1.1.gene.gff3 was used for alignment, and high-resolution view enables to discover the variants. Two base pairs "GC" were deleted 25 times out of 30 sequences. This deletion happened in the locus Sobic.006G173800 from the reference sequence (Fig. 8).

2.5. Discussion

Creating a link between phenotypes and genotypes requires extensive characterization and functional genomic analyses. Mutation breeding is considered to be one of the pioneering technologies in this regard and both natural mutation and induced mutation techniques are

used for studying gene functions. Although induced mutations are considered to produce high frequency of mutation as compared to natural mutation, it still requires considerable effort to identify causative mutations. Fast neutron mutagenesis methods have been used since 1927 after Muller's pioneering work on the artificial transmission of a gene (Muller, 1927) and report of induced mutations in barley by X-rays and radium a year later (Stadler, 1928). Fast neutron radiation has been considered as an effective mutagen for decades and has gained widespread recognition as a gene knockout approach for plants. Deletions from fast neutrons range from 1bp to >678kb (Li et al., 2016), with most deletions at 1bp to 4kb in size (Li et al., 2016; Kumawat et al., 2019).

In this study, 6,000 M2 sorghum lines were developed using fast neutron bombardment, which will be a useful resource for forward and reverse genetics screening to identify gene function in sorghum. Fast neutron bombardment causes primarily deletion of DNA fragments with a high potential of generating knockout mutants with discernible phenotypes (Li et al., 2002). The observation of 17.6% visible above ground phenotypes in independent M2 families under field conditions confirms this assertion. Since several of the mutations in M2 are still heterozygous, they segregate in M3 and M4 to homozygous mutant lines. Taking 10 seeds from each M2 family and 10 seeds from the each resulting M3 family, our population would produce 600,000 M4 pedigreed lines/families that would be ideal for screening, indicating that this is a substantial resource for sorghum functional genomics.

The strategies to identify mutations in fast neutron mutagenized populations have traditionally relied on using map based cloning, transcriptome profiling, PCR-based screening of DNA pools, and genomic hybridizations (Shirley et al., 1992; Bruggmann et

al., 1996; Li and Zhang, 2002; Gong et al., 2004; Mitra et al., 2004; Bruce et al., 2009; Bolon et al., 2011; Bolon et al., 2014), which are all cumbersome and time consuming. However, with the rapidly decreasing cost of next generation sequencing (NGS), mutations in a deletion population can now be conveniently identified with whole genome or whole exome sequencing (WGS/WES) (Belfield et al., 2012; Jia et al., 2018). WGS analysis in Arabidopsis indicated that a typical fast neutron irradiation results on average 18 mutations per genome (Belfield et al., 2012), suggesting a potential for developing a sequence-indexed mutant database for the relatively large (more than six times Arabidopsis) sorghum genome. Sequencing three mutant plants from M3 seedlings of three independent lines at 30X genome coverage identified a total of 7, 39, and 5,379 homozygous deletions per line, indicating wide variation in the number of deletions per genome. Sequencing of a few hundred lines perhaps at 40X coverage, and in pooled M5 seedlings of pedigreed M4 parents will likely provide the average number of mutations per genome to determine whether this approach is suitable for establishing a sequence-indexed mutant database in sorghum and other large genome crop species, where insertional mutagenesis is impractical. Nevertheless, the data demonstrate that fast neutron mutagenesis combined with WGS is a powerful tool for sorghum functional genomics and circumvents most of the difficulties associated with positional cloning and genomic hybridizations as exemplified by the identification of *SVD*.

2.6. References

Abe, A., Kosugi, S., Yoshida, K., Natsume, S., Takagi, H., Kanzaki, H., Matsumura, H., Yoshida, K., Mitsuoka, C., Tamiru, M., Innan, H., Cano, L., Kamoun, S., and Terauchi, R. (2012). Genome sequencing reveals agronomically important loci in rice using MutMap. Nat Biotechnol **30,** 174-178.

Baxter, I. (2009). Ionomics: studying the social network of mineral nutrients. Curr Opin Plant Biol **12,** 381-386.

Baxter, I.R., Vitek, O., Lahner, B., Muthukumar, B., Borghi, M., Morrissey, J., Guerinot, M.L., and Salt, D.E. (2008). The leaf ionome as a multivariable system to detect a plant's physiological status. Proc Natl Acad Sci U S A **105,** 12081-12086.

Belfield, E.J., Gan, X., Mithani, A., Brown, C., Jiang, C., Franklin, K., Alvey, E., Wibowo, A., Jung, M., Bailey, K., Kalwani, S., Ragoussis, J., Mott, R., and Harberd, N.P. (2012). Genome-wide analysis of mutations in mutant lineages selected following fast-neutron irradiation mutagenesis of Arabidopsis thaliana. Genome Res **22,** 1306-1315.

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

Bolon, Y.T., Haun, W.J., Xu, W.W., Grant, D., Stacey, M.G., Nelson, R.T., Gerhardt, D.J., Jeddeloh, J.A., Stacey, G., Muehlbauer, G.J., Orf, J.H., Naeve, S.L., Stupar, R.M., and Vance, C.P. (2011). Phenotypic and genomic

analyses of a fast neutron mutant population resource in soybean. Plant Physiol **156,** 240-253.

Bolon, Y.T., Stec, A.O., Michno, J.M., Roessler, J., Bhaskar, P.B., Ries, L., Dobbels, A.A., Campbell, B.W., Young, N.P., Anderson, J.E., Grant, D.M., Orf, J.H., Naeve, S.L., Muehlbauer, G.J., Vance, C.P., and Stupar, R.M. (2014). Genome resilience and prevalence of segmental duplications following fast neutron irradiation of soybean. Genetics **198,** 967-981.

Bruce, M., Hess, A., Bai, J., Mauleon, R., Diaz, M.G., Sugiyama, N., Bordeos, A., Wang, G.L., Leung, H., and Leach, J.E. (2009). Detection of genomic deletions in rice using oligonucleotide microarrays. BMC Genomics **10,** 129.

Bruggmann, E., Handwerger, K., Essex, C., and Storz, G. (1996). Analysis of Fast neutron generated mutants at the arabidopsis HY4 locus. The Plant Journal. 58: 668-681 **10,** 755-760.

Edgley, M., D'Souza, A., Moulder, G., McKay, S., Shen, B., Gilchrist, E., Moerman, D., and Barstead, R. (2002). Improved detection of small deletions in complex pools of DNA. Nucleic Acids Res **30,** e52.

Emmanuel, E., and Levy, A.A. (2002). Tomato mutants as tools for functional genomics. Curr Opin Plant Biol **5,** 112-117.

Gong, J.M., Waner, D.A., Horie, T., Li, S.L., Horie, R., Abid, K.B., and Schroeder, J.I. (2004). Microarray-based rapid cloning of an ion accumulation deletion mutant in Arabidopsis thaliana. Proc Natl Acad Sci U S A **101,** 15404-15409.

Henikoff, S., and Comai, L. (2003). Single-nucleotide mutations for plant functional genomics. Annu Rev Plant Biol **54,** 375-401.

Islam, N., Stupar, R.M., Qijian, S., Luthria, D.L., Garrett, W., Stec, A.O., Roessler, J., and Natarajan, S.S. (2019). Genomic changes and biochemical alterations of seed protein and oil content in a subset of fast neutron induced soybean mutants. BMC Plant Biol **19,** 420.

Jia, S., Li, A., Zhang, C., and Holding, D. (2018). Deletion Mutagenesis and Identification of Causative Mutations in Maize. Methods Mol Biol **1676,** 97-108.

Kim, Y., Schumaker, K.S., and Zhu, J.K. (2006). EMS mutagenesis of Arabidopsis. Methods Mol Biol **323,** 101-103.

Koornneef, M., Dellaert, L.W., and van der Veen, J.H. (1982). EMS- and radiation-induced mutation frequencies at individual loci in Arabidopsis thaliana (L.) Heynh. Mutat Res **93,** 109-123.

Kumawat, S., Rana, N., Bansal, R., Vishwakarma, G., Mehetre, S.T., Das, B.K., Kumar, M., Kumar Yadav, S., Sonah, H., Sharma, T.R., and Deshmukh, R. (2019). Expanding Avenue of Fast Neutron Mediated Mutagenesis for Crop Improvement. Plants (Basel) **8**.

Li, G., Chern, M., Jain, R., Martin, J.A., Schackwitz, W.S., Jiang, L., Vega-Sanchez, M.E., Lipzen, A.M., Barry, K.W., Schmutz, J., and Ronald, P.C. (2016). Genome-Wide Sequencing of 41 Rice (Oryza sativa L.) Mutated Lines Reveals Diverse Mutations Induced by Fast-Neutron Irradiation. Mol Plant **9,** 1078-1081.

Li, X., and Zhang, Y. (2002). Reverse genetics by fast neutron mutagenesis in higher plants. Funct Integr Genomics **2,** 254-258.

Li, X., Lassner, M., and Zhang, Y. (2002). Deleteagene: a fast neutron deletion mutagenesis-based gene knockout system for plants. Comp Funct Genomics **3,** 158-160.

Li, X., Song, Y., Century, K., Straight, S., Ronald, P., Dong, X., Lassner, M., and Zhang, Y. (2001). A fast neutron deletion mutagenesis-based reverse genetics system for plants. Plant J **27,** 235-242.

Martin, B., Ramiro, M., Martinez-Zapater, J.M., and Alonso-Blanco, C. (2009). A high-density collection of EMS-induced mutations for TILLING in Landsberg erecta genetic background of Arabidopsis. BMC Plant Biol **9,** 147.

McCallum, C.M., Comai, L., Greene, E.A., and Henikoff, S. (2000). Targeted screening for induced mutations. Nat Biotechnol **18,** 455-457.

Meinke, D.W., Meinke, L.K., Showalter, T.C., Schissel, A.M., Mueller, L.A., and Tzafrir, I. (2003). A sequence-based map of Arabidopsis genes with mutant phenotypes. Plant Physiol **131,** 409-418.

Mitra, R.M., Gleason, C.A., Edwards, A., Hadfield, J., Downie, J.A., Oldroyd, G.E., and Long, S.R. (2004). A Ca2+/calmodulin-dependent protein kinase required for symbiotic nodule development: Gene identification by transcript-based cloning. Proc Natl Acad Sci U S A **101,** 4701-4705.

Mohd-Yusoff, N.F., Ruperao, P., Tomoyoshi, N.E., Edwards, D., Gresshoff, P.M., Biswas, B., and Batley, J. (2015). Scanning the effects of ethyl methanesulfonate on the whole genome of Lotus japonicus using second-generation sequencing analysis. G3 (Bethesda) **5,** 559-567.

Muller, H.J. (1927). Artificial Transmission of a gene. Science **66,** 84-87.

Robinson, J.T., Thorvaldsdottir, H., Wenger, A.M., Zehir, A., and Mesirov, J.P. (2017). Variant Review with the Integrative Genomics Viewer. Cancer Res **77,** e31-e34.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat Biotechnol **29,** 24-26.

Rogers, C., and Oldroyd, G. (2008). Fast Neutron Mutagenesis for Functional Genomics. Methods and Protocols **299**.

Rogers, C., Wen, J., Chen, R., and Oldroyd, G. (2009). Deletion-based reverse genetics in Medicago truncatula. Plant Physiol **151,** 1077-1086.

Salt, D.E., Baxter, I., and Lahner, B. (2008). Ionomics and the study of the plant ionome. Annu Rev Plant Biol **59,** 709-733.

Shirley, B.W., Hanley, S., and Goodman, H.M. (1992). Effects of ionizing radiation on a plant genome: analysis of two Arabidopsis transparent testa mutations. Plant Cell **4,** 333-347.

Stadler, L.J. (1928). Mutations in Barley induced by X-rays and Radium. Science **68,** 186-187.

Sun, L., Ge, Y., Bancroft, A.C., Cheng, X., and Wen, J. (2018). FNBtools: A Software to Identify Homozygous Lesions in Deletion Mutant Populations. Front Plant Sci **9,** 976.

Tadege, M., Ratet, P., and Mysore, K.S. (2005). Insertional mutagenesis: a Swiss Army knife for functional genomics of Medicago truncatula. Trends Plant Sci **10,** 229-235.

Tadege, M., Wang, T.L., Wen, J., Ratet, P., and Mysore, K.S. (2009). Mutagenesis and beyond! Tools for understanding legume biology. Plant Physiol **151,** 978-984.

Thorvaldsdottir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform **14,** 178-192.

Wu, J.L., Wu, C., Lei, C., Baraoidan, M., Bordeos, A., Madamba, M.R., Ramos-Pamplona, M., Mauleon, R., Portugal, A., Ulat, V.J., Bruskiewich, R., Wang, G., Leach, J., Khush, G., and Leung, H. (2005). Chemical- and irradiation-induced mutants of indica rice IR64 for forward and reverse genetics. Plant Mol Biol **59,** 85-97.

CHAPTER III

3. SHRIVELED SEED (*SVD*), A P_{1B}-TYPE HMA5 TRANSPORTER ATPASE, IS INVOLVED IN COPPER DETOXIFICATION TO CONTROL SORGHUM GROWTH AND DEVELOPMENT

3.1. Introduction

Some heavy metals and metalloids such as Aluminum (Al), Cadmium (Cd), Lead (Pb), Mercury (Hg), and Arsenic (As) are considered non-essential to plants but can adversely affect crop performance and many of them pose serious risks to human health (Gall et al., 2015; Clemens and Ma, 2016; Rai et al., 2019). Other metals such as Copper (Cu), Iron (Fe), Manganese (Mn), Cobalt (Co), Nickel (Ni) and Zinc (Zn), on the other hand, are required by plants in trace amounts as they make up important components of cytochromes and enzymes that catalyze critical metabolic reactions including photosynthesis, respiration and scavenging reactive oxygen species. The lack of such elements can compromise plant growth and productivity but also cause diseases in humans that range from Menkes syndrome (Vairo et al., 2019) for Cu deficiency to anemia for Fe deficiency (Miller et al., 2013). However, at higher concentrations, these elements become toxic inhibiting plant growth and development, and causing health risks in humans.

As an essential micronutrient, Cu is required in many physiological processes in plants including photosynthetic electron transport, mitochondrial respiration, oxidative stress responses, protein and carbohydrate metabolism, hormone signaling, and acts as a structural element in regulatory proteins (Marschner and Rietbrock, 1995; Raven et al., 1999; Rehman and Rather, 2019). Cu ions act as cofactors in many enzymes such as Cu/Zn superoxide dismutase (SOD), cytochrome c oxidase, amino oxidase, laccase, plastocyanin and polyphenol oxidase, affecting several fundamental physiological processes including signaling of transcription and protein trafficking, oxidative phosphorylation and iron mobilization (Ravet and Pilon, 2013; Rehman and Rather, 2019). But, at high concentrations, Cu is extremely toxic causing symptoms such as chlorosis and necrosis, stunting, leaf discoloration and inhibition of root growth (Van Assche and Clijsters, 1990; Marschner and Rietbrock, 1995; Chandrasekhar and Ray, 2017) due to inhibition of enzyme activity or protein function, impaired cell transport processes, and oxidative damage (Meharg, 1994). Thus, Cu homeostasis is critical for proper plant growth and development.

Despite the critical importance of Cu in both deficiency and toxicity, the molecular mechanism of Cu homeostasis in plants is still poorly understood. In Arabidopsis, two types of Cu transporters have been described, CTR and HMA type. The COPT1 gene and its homologues belong to the copper transporter (CTR) family and encode high affinity Cu transporters that allow the entrance of Cu into plant cells (Kampfenkel et al., 1995; Puig et al., 2002; Sancenon et al., 2004) . The second type of Cu transporters control Cu homeostasis by delivery and removal of Cu into and out of intracellular sites and encode P-type ATPases that use energy from ATP hydrolysis to drive Cu transport across

membranes. P-type ATPases transport a variety of ions including K^+, H^+, Na^+/K^+, H^+/K^+, Ca^{2+}, heavy metals, and phospholipids, and have been divided into 5 major groups based on sequence motifs and substrate specificity, namely P_{1B}, P_{2A}, P_{2B}, P_{3A}, P_4-type ATPases and a 6^{th} group (P_5) with yet unidentified substrate (Axelsen and Palmgren, 2001; Arguello, 2003; Palmgren and Nissen, 2011; Rosenzweig and Arguello, 2012). The P_{1B}-type heavy-metal ATPases (HMAs) are found in all organisms from bacteria to humans, and transport a variety of heavy metal ions across membranes including those that are required for life processes, such as Cu and Zn and those that are not required such Cd and Pb (Rosenzweig, 2001; Rosenzweig and Arguello, 2012).

A few Cu transporter HMAs have been identified by genetic analysis in plants (Abdel-Ghany et al., 2005; Andres-Colas et al., 2006; Klaumann et al., 2011; Zheng et al., 2012; Deng et al., 2013; Garcia-Molina et al., 2013; Huang et al., 2016). P_{1B} -type ATPases (HMAs), are represented by 8 members in Arabidopsis, denoted HMA1 to HMA8 (Baxter et al., 2003). While HMA1 to HMA4 may be involved in Zn, Cd, Pb homeostasis (Mills et al., 2003; Hussain et al., 2004), HMA5 to HMA8 are believed to play a role in the homeostasis of Cu and Ag (Axelsen and Palmgren, 2001). Out of these, *RAN1* corresponds to *HMA7* and delivers Cu to ethylene receptor (Hirayama et al., 1999; Woeste and Kieber, 2000). *PAA1* (*HMA6*) and *PAA2 (HMA8)* are required for Cu delivery into the chloroplast (Abdel-Ghany et al., 2005). HMA5 is reported to detoxify Cu in Arabidopsis roots presumably by extruding Cu out of the cells (Andres-Colas et al., 2006). In rice, while OsHMA5 is reported to be involved in Cu loading into the xylem for root-to-shoot translocation (Deng et al., 2013), OsHMA4 is shown to prevent Cu accumulation in the rice grain by sequestering it to the vacuole (Huang et al., 2016). These P_{1B} type ATPase

transporters complete their work by interacting with metallochaperones at the protein level to control metal concentrations in the cell and transport to different cellular regions safely (Rosenzweig, 2002b).

Metallochaperones are very important proteins that are required for efficient and safe transport of metal ions into different cellular sections (Abreu-neto et al., 2013). These metallochaperones have specific cysteine rich sites MT/HCxxC, where the second amino acid is either threonine or histidine and the 'xx' can be any of the 20 amino acids (Arnesano et al., 2002). The CxxC domain is a binding motif within the first loop of a ferredoxin-like structural fold (βαββαβ), which is the heavy metal binding domain (Rosenzweig, 2002b; Shoshan et al., 2011; de Abreu-Neto et al., 2013).

A number of different strategies are used to study the intracellular location of protein, and interaction between proteins. These include subcellular localization using fluorescence tags and protein-protein interaction studies. One of the most widely used methods for the latter include yeast two-hybrid, bimolecular fluorescence complementation, and co-immunoprecipitation (Fields and Song, 1989; Fields and Sternglanz, 1994; Luo et al., 1997; Young, 1998; Kerppola, 2006b, a; Fan et al., 2008; Kerppola, 2008; Bruckner et al., 2009; Hamdi and Colas, 2012).

Subcellular distribution of proteins aids to understand the mechanism of protein functions. Advancements in fluorescence microscopy imaging has made this approach convenient and reliable (Glory and Murphy, 2007). The yeast two-hybrid method, on the other hand, has many applications including testing of known proteins for possible interaction, identifying specific domains of proteins critical for interaction, and yeast library screening for identifying unknown protein partners (Fields and Sternglanz, 1994).

The yeast two-hybrid Gal4 transcription activator protein system developed in 1989 by Fields and Song is used system in many laboratories for the identification of possible protein-protein interactors in bacteria, plant, and animal cells (Fields and Song, 1989). Yeast two-hybrid system remains a preferred systems for the study of protein-protein interactions for its low cost and simple implementation and the Gal4 system has been modified for ease of in different organisms (Fields and Sternglanz, 1994; Young, 1998; Bruckner et al., 2009; Snider et al., 2010; Hamdi and Colas, 2012; Stynen et al., 2012). The overall principle of the yeast two-hybrid system is the detection of interaction between two proteins using a prey and bait system to reconstitute a Gal4 protein and enable growth on a specific media or color reaction notifying the two proteins in question to have a likely chance of interaction in the cell to accomplish a specific physiological process in the organism. However, one of the problems with the Yeast two-hybrid method is non-specific interaction that can produce false positive results and failure of legitimate interactions, resulting in false negatives (Bruckner et al., 2009). Hence, to avoid bias, yeast two-hybrid results must be validated using other methods.

Bimolecular fluorescence complementation (BiFC) is a reliable alternative method to identify protein binding *in vivo* and provides important confirmation of protein-protein interactions. Like the yeast two-hybrid system the BiFC system employs a similar principle of bait and prey system where two rationally designed non-fluorescent fragments of a fluorescent protein (CYFP and NYFP) are fused to protein ligands and come into a close proximity through interaction between the ligands to reconstitute fluorescence. This system has the capability to detect weak protein interactions and it is often used as a validation of other protein-protein interaction techniques (Shyu et al., 2006; Morell et al.,

2008b; Morell et al., 2008a). Here we describe the functional characterization of a developmental mutant in sorghum named *shriveled seed* (*svd*), which displays severe defects throughout plant development from seed germination to grain filling.

3.2. Objectives

The main objective of this work was to characterize the *svd* mutant and to understand the molecular mechanism of SVD protein.

3.3. Materials and Methods

3.3.1. RNA extraction, cDNA synthesis and Gene cloning for transformation

Total RNA was isolated using TRIzol Reagent (Invitrogen) and after purification with ethanol, complementary DNA (cDNA) synthesis was completed using *SSIV* reverse transcriptase enzyme using an oligo (dT) prime following manufacturer's instruction. RT-PCR and RT-qPCR were performed as described by Biassoni and Walker (Biassoni and Walker, 2014) by using actin as a control. *SVD* gene sequence was retrieved from the sorghum genome using a BLAST search in GenBank using gene id 'Sobic.006G173800'. Full-length of *SVD* gene was amplified by PCR from cDNA synthesized from sorghum (BTx623) leaves using *SSIV* reverse transcriptase. RT primers were designed by an NCBI online tool Primer3plus (Barrett et al., 2007), using a forward primer $^{5'}$ATGGTGGCGACTACTCGAGCTCT$^{3'}$ from the transcription start site and a reverse primer $^{5'}$TCAACTGCCAACAATCTTTGGAGCCT$^{3'}$ to amplify the gene. For sequencing, the *SVD* gene was amplified by attaching *attB* sites and the purified product was cloned to *pDONOR207* and the sequence identity was confirmed by alignment with the reference gene sequence.

Local sequence alignment was performed to confirm whether the cause the mutation was due to the frameshift mutation caused by the "GC" deletion. A short specific region targeting the deletion (GC) region was used to design short primers; forward: $^{5'}$ATAATGGTGACTGGCGACAACTGG$^{3'}$ 300bp just before the deletion and reverse: $^{5'}$GGCAAATGATATGCCCGTCTTCCA$^{3'}$ 243 after the deletion were used to amplify a 543 base pair representing the target region in the selected lines. After amplifying with RT-PCR, the purified product was cloned into pGEM-T Easy vector and prescreening was done using X-gal and IPTG to identify false positives and finally, the positive lines were sequenced using the Sanger sequencing system.

Gene cloning and making constructs were carried out using the Gateway cloning technique and transformed to pDONOR207. All constructs were transformed into pMDC32 Gateway destination vector under control of the *UBI* promoter.

3.3.2. High throughput Ionomics

Heavy metal content of *svd* was determined using an inductively coupled plasma Mass Spectroscopy (ICP-MS) elemental analysis system (Baxter, 2009) at the Danforth Plant Science Center facility in St. Louis MO, USA. Leaf and seed samples for analysis were collected in five replications from three week old plants grown in standard half strength MS media. Samples were oven dried in 50^{o}C for 48 hours, weighed without contamination of any metals and sent to Danforth Plant Science center for digestion.

3.3.3. Yeast complementation

The yeast (*Saccharomyces cerevisiae*) wild type strain BY4741 (*MATa his3*Δ*1 leu2*Δ*0 met15*Δ*0 ura3*Δ*0*), and BY4741-derived mutant ccc2 (*MATa his3*Δ*1 leu2*Δ*0*

met15Δ0 ura3Δ0 YΔR270w::kanMX4) were obtained from Open Biosystems. Yeast complementation was performed as reported (Huang et al., 2016). Briefly, the full length *SVD* gene was cloned in the pYES2/CT yeast expression vector by adding HindIII and BamHI restriction sites to *SVD* and ligating this into the pYES2/CT vector, creating pYES2/CT-SVD plasmid. The plasmid propagated in *E. coli* and purified with standard miniprep for yeast transformation. The new pYES2/CT- SVD plasmid construct and empty vector pYES2/CT were transformed into BY4741 WT and Δccc2 mutant strains using the LiAc method (Gietz and Schiestl, 2007). Then, transformants were cultured at 30°C overnight in 3ml of *SD-Ura* media (6.7g l1 yeast nitrogen base, 1.92g l1dropout mix without uracil) containing 2% (w/v) glucose. Cells were washed three times with 10ml sterile deionized water and the optical density at 600nm adjusted to 0.2 with sterile distilled water. After sequential 10-fold dilutions, 10ml of cell suspension of each genotype was spotted on Fe-limited, Fe-sufficient or Cu-sufficient media, and the plates were incubated at 30°C for 3 days. The Fe-limited media contained 0.17% (w/v) yeast nitrogen base without $CuSO_4$ and $FeCl_3$ (BIO 101 Systems), 0.2% (w/v) dropout mix without uracil, 2% (w/v) galactose, 1% (w/v) raffinose, 50 mM MES (pH 6.1), 1mM 3-(2-pyridyl)-5, 6-bis (4-sulfophenyl)-1, 2, 4-triazine disodium salt (Ferrozine disodium salt; Sigma), 50mM $Fe(NH_4)_2(SO_4)_2$,1mM $CuSO_4$ and 2% (w/v) agar. The Fe-sufficient and Cu-sufficient media were modified from Fe-limited media by increasing the concentration of added Fe $(NH_4)_2(SO_4)_2$ and $CuSO_4$ to 350 and 500mM, respectively, and omitting Ferrozine disodium salt.

3.3.4. Sequence alignment and Phylogenetic analysis

Sequences of the SVD protein from sorghum and other related species were obtained from NCBI GenBank. Multiple sequence alignment was performed using muscle and Phylogenetic analysis was done using MEGA7 software with 1000 bootstrap replicates to construct the neighbor joining Phylogenetic tree.

3.3.5. Yeast two-hybrid analysis

Yeast two-hybrid analysis was performed using the Clontech Two-Hybrid system (Invitrogen). The full-length coding sequence of SVD was cloned in pGBKT7-GW as bait, while the full-length sequence partner proteins were cloned in pGADT7-GW as prey and sets of constructs were co-transformed into Y2H Gold yeast strain (Clontech). Yeast transformants were grown on minimal double dropout culture media deficient in SD/-Leu/-Trp, and protein interactions were assessed on triple dropout media (TDO) deficient in SD/-His/-Trp/-Leu. MtWOX9 and MtSTF interaction was used as a positive control.

3.3.6. Bimolecular fluorescence complementation assay

BiFC assays were conducted according to (Lu et al., 2010). Briefly, the *SVD* coding sequence was cloned into N-YFP, while coding sequences different partner chaperon proteins (*SbATX1, SbFRN3, SbHMT1, SbATOX1, SbDETO, SbIPP23* and others) were cloned into C-YFP, by LR reaction. Each construct was transformed into *Agrobacterium* strain *GV2260* by freeze-heat shock method. Combined pairs were co-infiltrated into four week old *N. benthamiana* leaves. P19 was used to inhibit transgene silencing. Yellow fluorescent protein (YFP) signal was assessed after 48 to 60hrs of infiltration using a TCS SP2 AOBS confocal laser scanning microscope (Leica Microsystems).

3.3.7. Sub-cellular localization

Full-length *SVD* coding sequence was cloned from BTx623 genotype using a Gateway cloning system by removing the stop codon and sequenced in a pDONOR207, and then ligated to the destination vector pGBW505 with GFP in the C- terminal region. The final construct was transformed into Agrobacterium strain *GV2260*. A new culture from a single colony was grown and infiltrated into N. *benthamiana* leaves. P19 was used to inhibit transgene silencing. Screening and imaging using a TCS SP2 AOBS confocal laser scanning microscope were done after 48 hours.

3.4. Results

3.4.1. Backcrossing of *svd* mutant to the Wild type BTx623 genotype

Homozygous *svd* mutant plants were backcrossed to WT and segregation of the *svd* phenotype was analyzed in the F2 progeny. A total of 240 F2 plants were grown in the greenhouse in one gallon pots under long-day (LD) conditions with 16/8 hours light/dark cycle at 28-32°C temperature. Of these plants,164 showed the wild type phenotype (Fig.9a), while 44 plants displayed the *svd* phenotype (Fig. 9b), indicating a segregation ratio of 3.7:1 (164:44) for the wild-type to mutant phenotype that is not statistically different from the predicted 3:1 segregation ratio of 3:1 (X^2 value of 7.66 and p-value=0.005). Growth of leaves of plants with the mutant phenotype was strongly reduced while the wild type segregants showed healthy leaves (Fig.9c). These data indicate that *svd* segregates as a recessive allele.

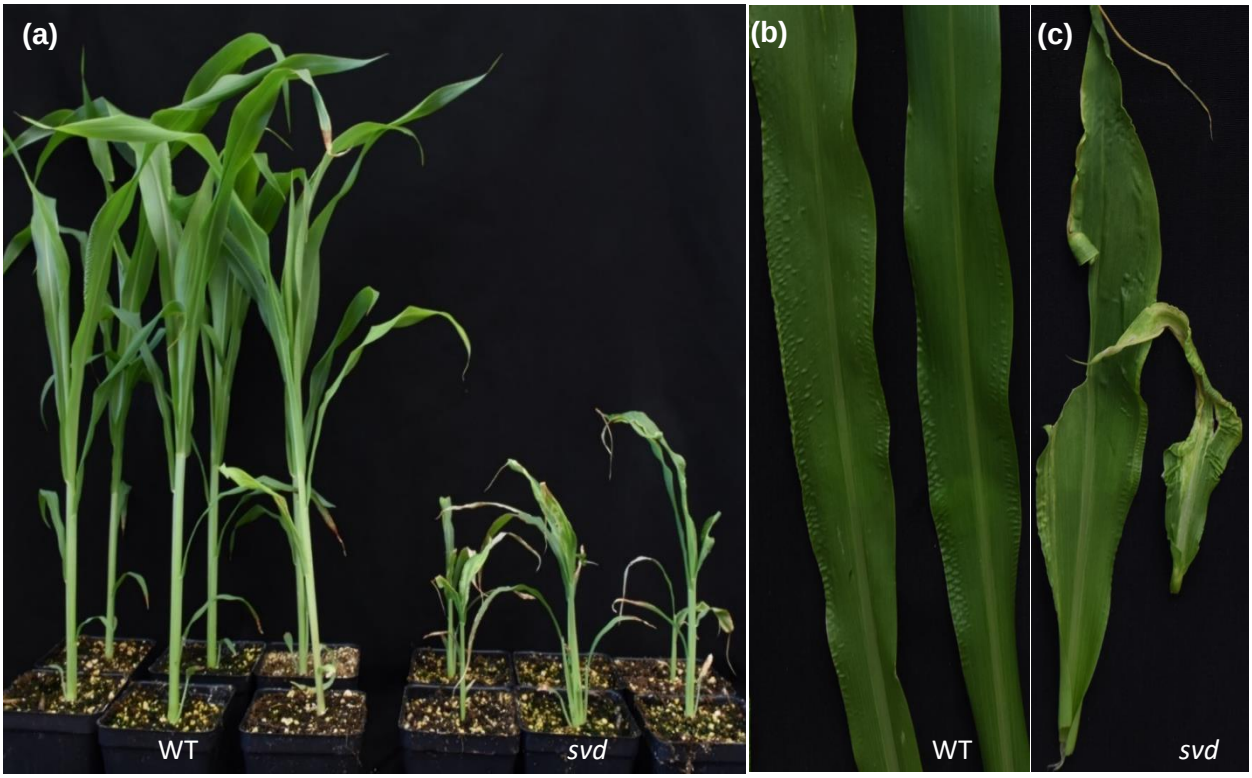

Fig. 9) Back crossing of *svd* mutant and re-screening of the F2 generation.

Whole plant and detached leaf Phenotypes of *svd* mutant in soil four weeks after germination. (a) WT control (left) and typical *svd* mutant plants grown for four weeks in greenhouse under long day conditions, (b) Close up view of detached WT leaves same age as in (a), (c) Close up view of detached *svd* mutant leaves same age as in (a).

3.4.2. Sequence segregation analysis

Based on EMBOSS Transeq online protein translation tool, the deletion in the exon resulted in a frameshift after the 914th amino acid, resulting in a longer open reading frame (1016 amino acids) than the annotated reading frame (998 amino acids). However, the *svd* transcript was not expressed to any detectable levels, indicating that *svd* is likely to be a null (loss of function) allele (data not shown). To evaluate if the 2 bp deletion identified in the Sobic.006G173800 locus segregates with the *svd* mutant phenotype, the region

surrounding the deletion was amplified and sequenced from 20 F2 siblings showing the mutant phenotype and 3 that showed the wild type phenotype, along with one plant with the parental BTx623 genotype. The results showed that all the 20 *svd* siblings harbored the deletion while the corresponding sequences from the 3 wild type plants were intact and matched that of the parental genotype (Fig. 10). Thus, the 20 *svd* segregants were homozygous for the deletion, while the three wild type segregants carried at least one copy of the intact coding sequence a. These observations indicate that the 2 bp deletion co-segregates with the *svd* phenotype and this frameshift mutation is likely to result in the loss of function allele that is causative for the *svd* phenotype. .

Fig.10*)* Local alignment shows a homozygous deletion in lines showing phenotype

Sequence alignment of a specific region (300bp upstream of the deletion site and 243 downstream of the deletion site) to determine whether the cause of the mutant

phenotype to the deletion in the SVD gene. The Parental BTx623 line was represented by WT while H1, H2 and H3 represent F2 segregants with the wild-type phenotype and *svd* segregants are represented by T3- T26

3.4.3. The *SVD* gene encodes a putative P_{1B}-type copper transporting ATPase

The Sobic.006G173800 locus *(SVD* gene*)* has five exons and the deletion happened in the fifth exon. SVD encodes a peptide of 998 amino acids with sequence similarity to known Cu transporter ATPases (Fig. 11a). Topology predictions of SVD with HMMTOP (Tusnady and Simon, 2001) showed eight putative transmembrane domains, along with conserved features of metal transporters including two heavy metal binding motifs (GMTCAAC) and (GMACKYC) near the N-terminus, and an ion transduction (CPC) motif within the sixth transmembrane domain (Fig. 11b). Moreover, it has an extended hydrophilic loop with phosphorylation (DKTG), ATP binding (SEHPL) and phosphatase (GDGIND) domains (Petris et al., 2002; Vonk et al., 2012) between transmembrane domains 6 and 7 (Fig. 11b, Fig.12), typical of a P_{1B}-type heavy metal ATPase (HMA).

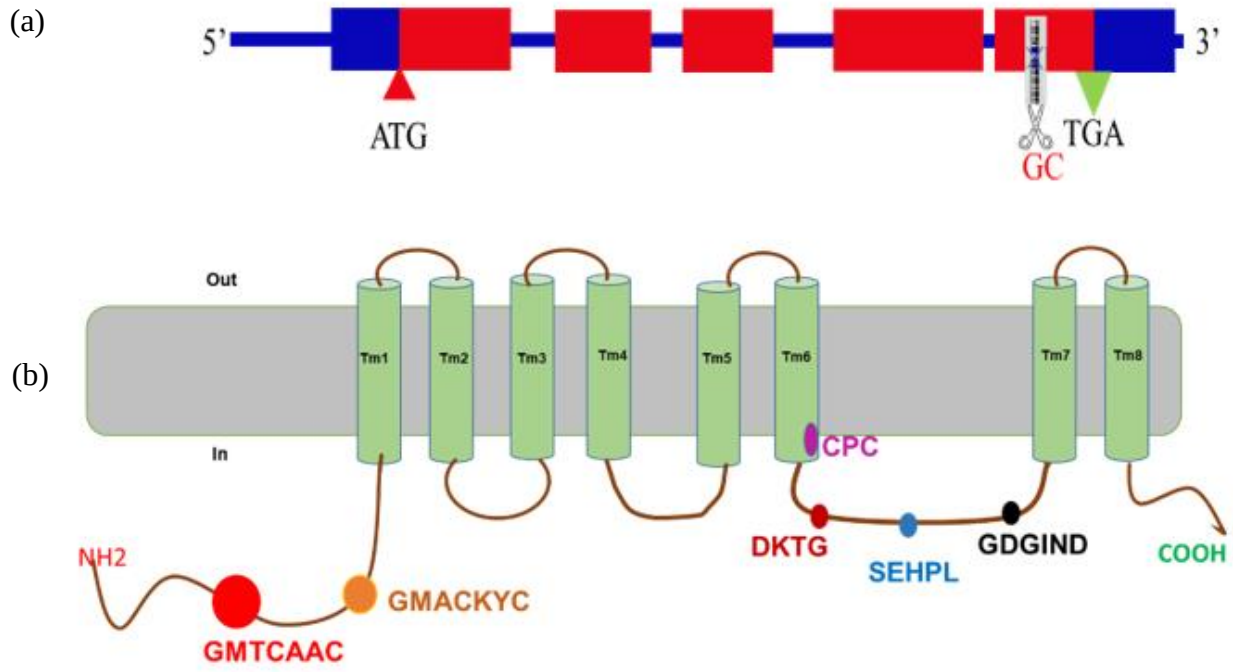

Fig.11) Topology of SVD contains one GMTCxxC and one CPx binding motifs

a) Prediction of gene structure using Gene Scan, b) Predicted topology of SVD and its predicted protein structures including its eight transmembrane regions and other important domains like metal binding and ion binding motifs.

3.4.4. Evolutionary relationship based on phylogenetic analysis shows *SVD* is a close homologue of *HMA5*

Multiple sequence alignment showed SVD is most closely related to the previously characterized copper transporter ATPases OsHMA5 (85.7% identity) and AtHMA5 (69.72% identity) from rice and Arabidopsis, respectively (Fig. 12).

Fig.12) Sequence alignment with MEGA 7.0 method showed GMTCxxC and CPC motifs conserved

Alignment performed using HHpred in all SVD, AtHMA5, OsHMA5, OsHMA4 and AtHMA7 showed GMTCxxC is conserved across all of the proteins, while in the second GMTCxxC motif (marked in black arrows), the third amino acid "T" is substituted with "A". The CPC domain (marked in black arrows) is also conserved across all the species).

Additionally, to identify the most closely related orthologs from different species, phylogenetic analysis was performed, and the closest homologue was from Sugarcane

(*Saccharum* species) with 96.5% identity (Fig.13). The next most closely related sequences were HMA5 orthologs from maize, and *Setaria*, which shared 93.3% and 91% identity, respectively (Fig.13).

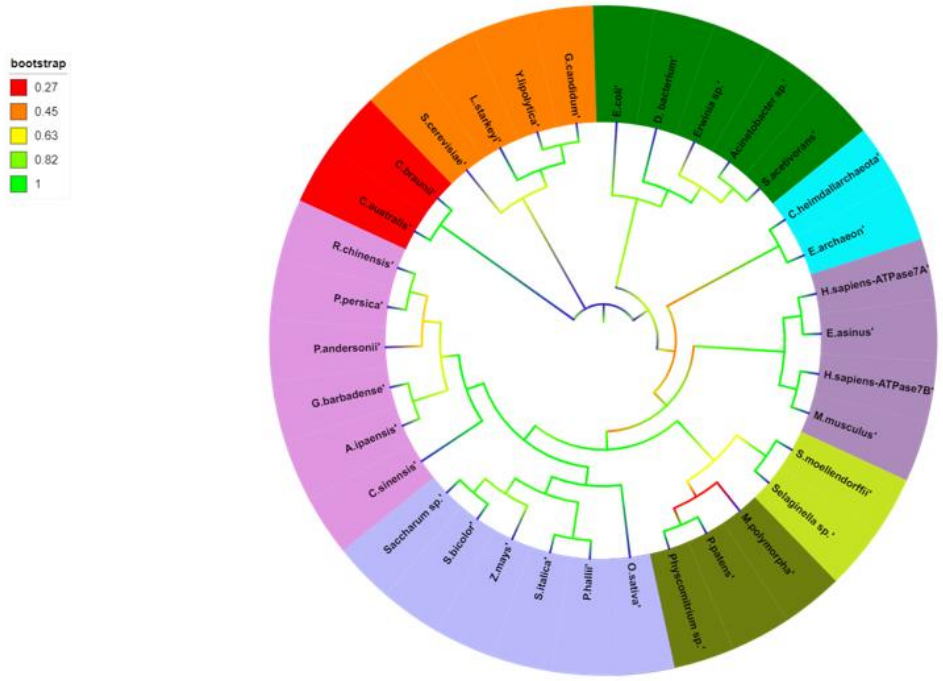

Fig. 13) Phylogenetic study shows SVD is Conserved across different species

Multiple amino acid sequence alignment was performed using MUSCLE and the neighbor joining phylogenetic tree was constructed using MEGA 7.0 with 1000 bootstrap replicates and poisson model. The exported tree in newick format is edited using iTOL: interactive tree of life program (Letunic and Bork, 2019).

There are eleven P_{1B}-type HMAs annotated in the reference sorghum genome (Fig. 14), of which HMA3 and HMA5 appear to be represented by duplicate genes each, SbHMA3 and SbHMA3-like (SbHMA3L), and SbHMA5 and SbHMA5-like (SbHMA5L) (Fig. 14). Out of the eleven sorghum HMAs, seven (SbHMA4-SbHMA9) including SVD/SbHMA5 and SbHMA5L form a distinct phylogenetic cluster with Arabidopsis copper transporters

HMA5-HMA8, while the other four (SbHMA1- SbHMA3L) cluster with HMA1-HMA3 (Fig. 14). reported to be Cd and Zn transporters (Mills et al., 2003; Eren and Arguello, 2004; Mills et al., 2005; Kim et al., 2009; Cai et al., 2019).

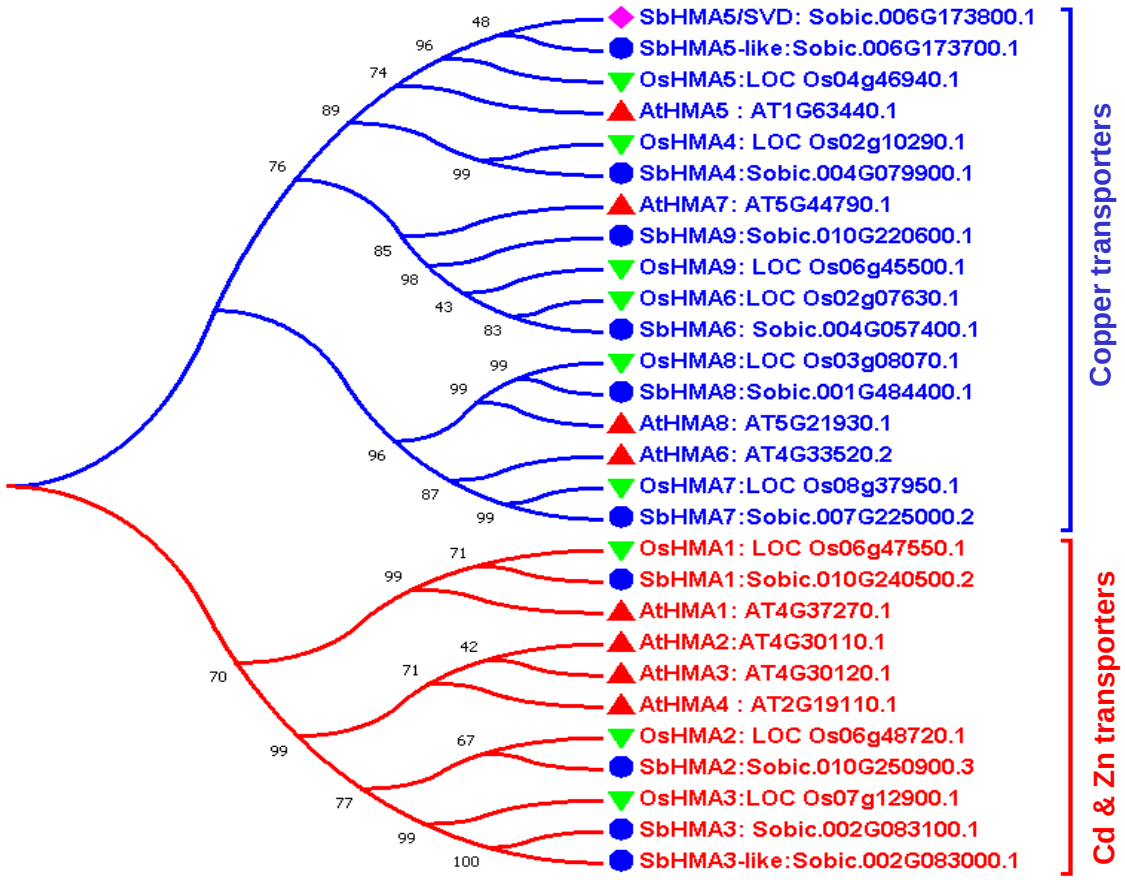

Fig.14) Phylogenetic analysis of SVD/SbHMA5 with other species

Maximum likelihood neighbor joining phylogenetic tree of P_{1B}-type heavy metal ATPases in sorghum (blue circles), rice (green triangles), and Arabidopsis (red triangles). The SVD protein (pink diamond shape) clusters with the Cu transporters.

A 20kb genomic region close to the Sobic.006G173800 locus and its homolog from other species (*Setaria italic, Panicum halli, Ozysa sativa, Brachypodium distachyon, and Arabidopsis thaliana)* was retrieved from Phytozome. Comparative analyses of these regions showed that sorghum, Setaria and Panicum have two tandemly linked HMA5 genes

while rice, *Brachypodium* and Arabidopsis have only a single HMA5 gene at this locus (Fig. 15). This suggests the divergence of the gene between different families.

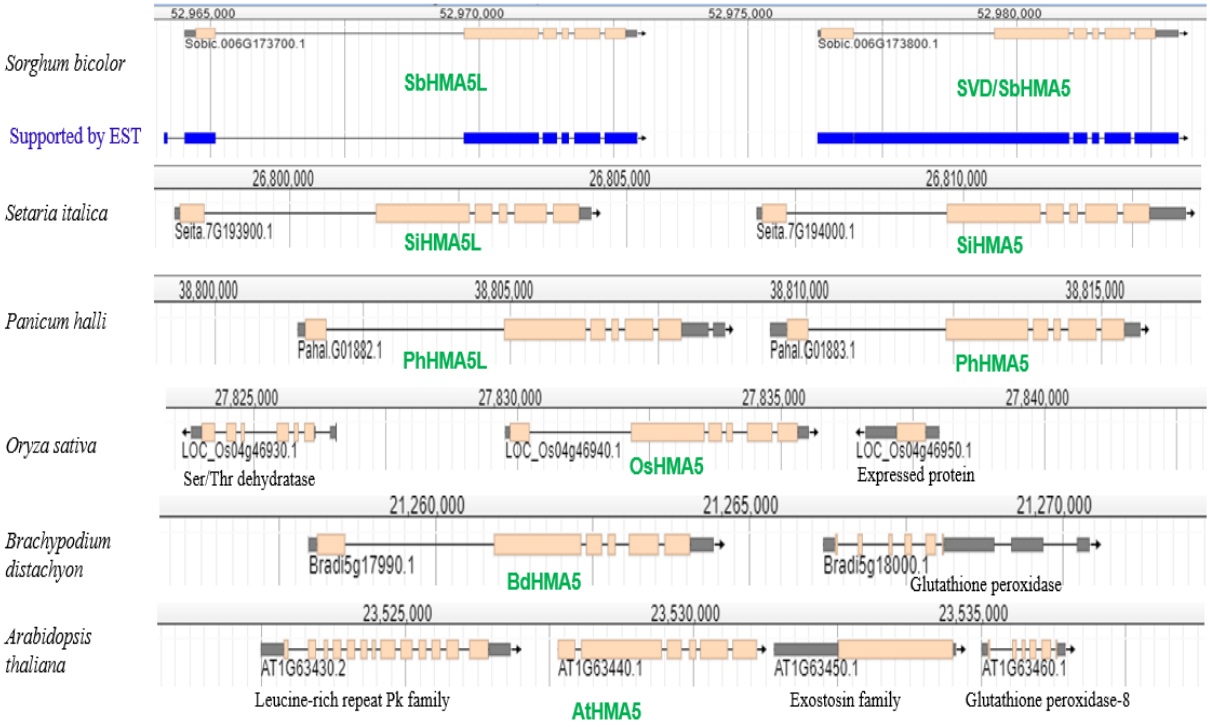

Fig.15) The HMA5 gene was duplicated in sorghum, *Setaria*, and *Panicum* .

Gene Browser showing a 20 kb genomic region around the HMA5 locus in six species. *Sorghum bicolor*, *Setaria italica*, *Panicum halli*, *Ozysa sativa*, *Brachypodium distachyon*, and *Arabidopsis thaliana*. Note that HMA5 is represented by two genes in sorghum, *Setaria* and *Panicum* but by a single gene in rice, *Brachypodium* and Arabidopsis.

3.4.5. *SVD* complements the yeast *ccc2* knockout mutant defective in copper transport

To test if SVD indeed can transport copper across membrane, we complemented the yeast *ccc2* Cu transport defective mutant with the wild type *SVD* gene. Yeast complementation was performed according to (Huang et al., 2016), using pYES2/CT yeast expression vector. The *ccc2* gene is a Cu-transporting P-type ATPase that delivers Cu to the multicopper oxidase Fet-3P, required for high affinity Fe uptake at the plasma membrane, and

disruption of the *ccc2* gene results in a defect in respiration and iron uptake (Yuan et al., 1995), making the *ccc2* yeast mutant incapable of growing on iron-limited media. The full-length coding sequence of *SVD* was cloned into the *BamHI* and *HindIII* sites of pYES2/CT vector to generate pYES2/CT-SVD. The pYES2/CT-SVD plasmid and empty pYES2/CT vector were transformed into the *ccc2* mutant and BY4741 wild type control, and plated on Fe-limited, Fe-sufficient, and Cu-sufficient media (Fig. 16). Growth was evaluated after three days of incubation at 30ºC. In Cu- or Fe-sufficient media, both BY4741 and *ccc2* mutant strains grew well with or without SVD (Fig. 16a,b). But in Fe-limited media, while the *ccc2* mutant transformed with pYES2/CT alone failed to grow, the *ccc2* mutant transformed with pYES2/CT-SVD grew well to the level of BY4741 (Fig. 16c), indicating that *SVD* fully complemented the yeast *ccc2* mutant. This suggests that sorghum SVD is capable of transporting copper across the yeast plasma membrane, like the rice and Arabidopsis HMA5 proteins.

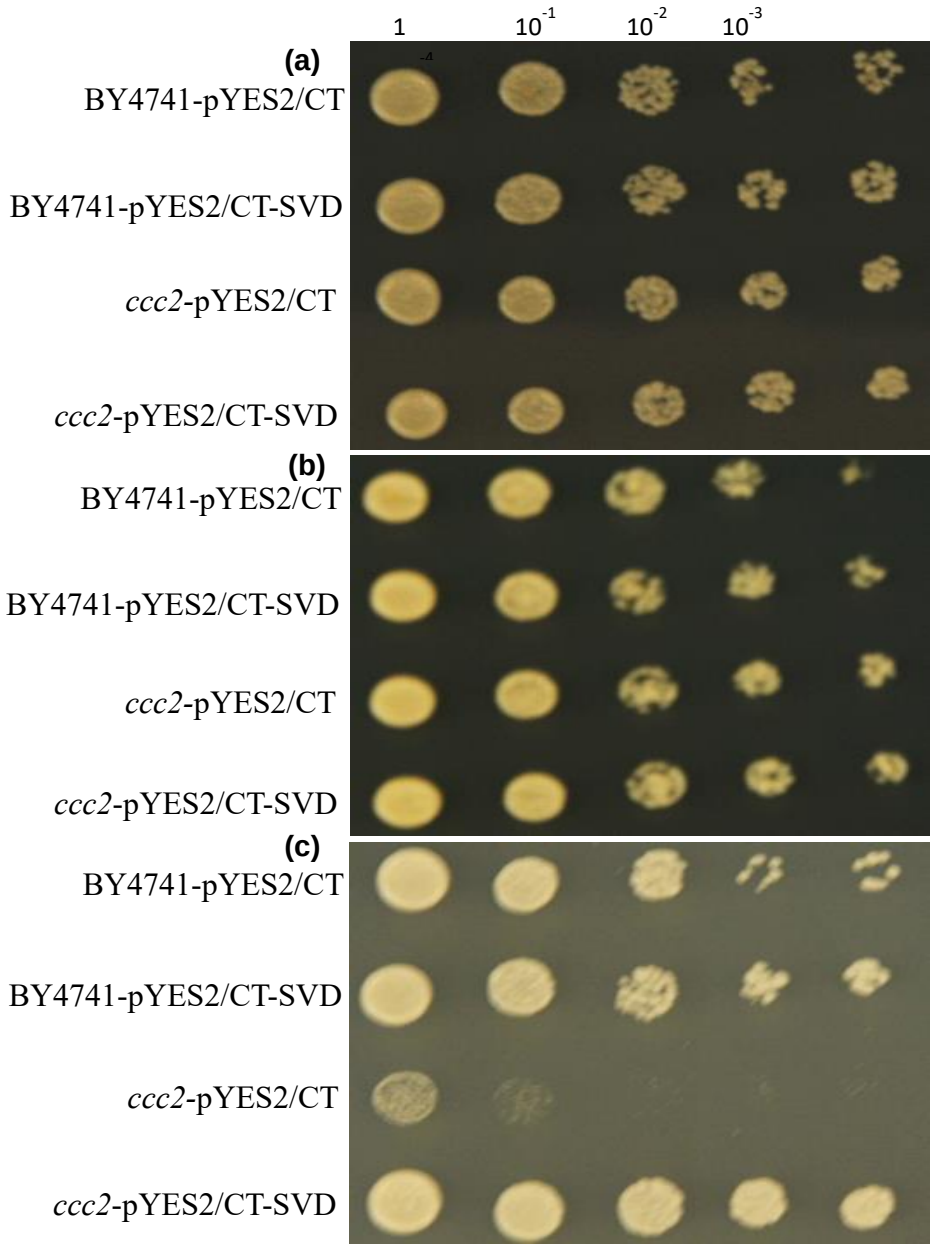

Fig.16) Yeast complementation and confirmation of mutation by local alignment

Yeast strains BY4741 and *Δccc2* transformed with empty vector pYES2/CT(Vector) or an *SVD*-expression vector were cultivated on Fe-limited, Fe-sufficient, and Cu-sufficient plates as indicated a) The WT strain (BY4741) and its knock out (Δ*ccc*2) mutant grown in copper sufficient media with and without the SVD gene, b) The WT strain (BY4741) and its knockout *ccc*2 grown in iron sufficient media with and without the SVD gene, c) The WT strain (BY4741) and its knockout ccc_2 grown in iron limited media with and without the *SVD* gene.

3.4.6. The *svd* mutant accumulates more Cu and is hypersensitive to $CuSO_4$ treatment

To examine the heavy metal ion concentrations in plant tissues, we determined the concentration of Cu, Mn, Zn, and Fe in leaf tissues of *svd* mutant and WT plants grown in MS-media, and seed samples from *svd* and WT grown in 1 gallon pot in the greenhouse were also included to evaluate these metal concentrations by inductively coupled plasma mass spectrometry (ICP-MS) (Baxter, 2009; Baxter and Dilkes, 2012). We found that, while the concentration of Mn, Zn and Fe remained similar between *svd* and WT, significantly higher ($P<0.01$) amounts of Cu were detected in both the leaves and seeds of *svd* mutants compared to WT (Fig. 17a-d). This indicates that the *svd* mutant is unable to remove excess Cu from the tissue, suggesting the possible involvement of SVD in Cu efflux and homeostasis.

To determine if the *svd* mutant plants are sensitive to specific heavy metals, we germinated seeds on tissue culture media and grew them for three weeks at different concentrations of heavy metals. Significant differential in growth between WT and *svd* was observed only for CuSO4 treatment (Fig.17e-i). With the addition of 30 µM $CuSO_4$ to culture media, growth of the *svd* mutant was completely arrested while the WT grew similar to the untreated control (Fig. 17e,f). Addition of 3 mM MnSO4 or 0.9 mM ZnSO4, on the other hand, affected growth of both WT and *svd* although the effect of Mn appeared to be slightly more pronounced in the *svd* mutant (Fig. 17g,h), but addition of 3 mM FeSO4 had no effect similar to the untreated control (Fig. 17i). These observations indicate that the *svd* mutant is specifically sensitive to additional CuSO4 treatment and support the

notion that SVD may be involved in Cu toxicity detoxification, possibly by extruding excess Cu out of the cell.

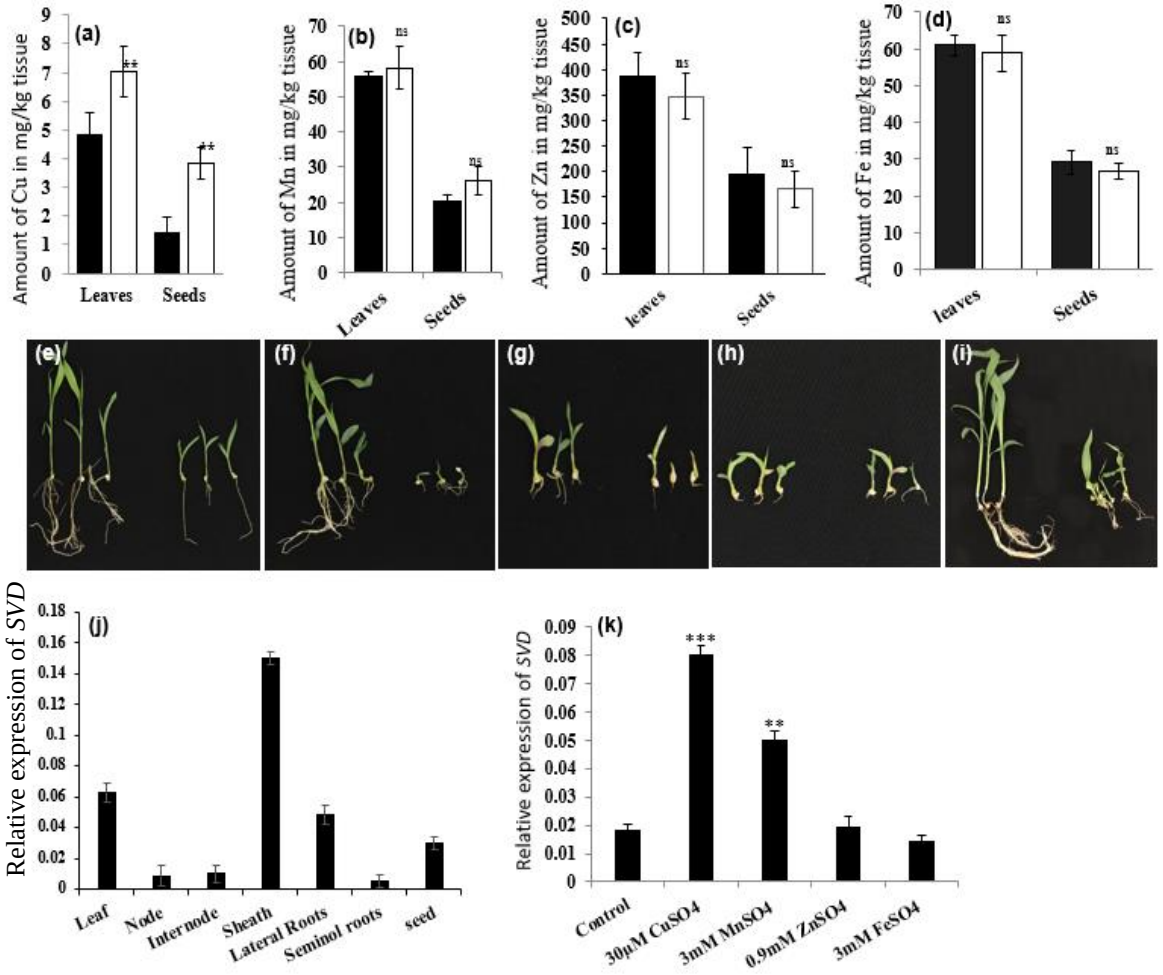

Fig. 17) ICP-MS Ionomics analysis in the leaves and seeds of svd mutant and SVD expression in the leaf in response to heavy metals treatment.

(a-d) Heavy metal content in mg per kg of seeds and leaves from 6 week old plants grown in soil. Shown are (a) Cu, (b) Mn, (c) Zn and (d) Fe in WT (dark bars) and svd mutant (open bars). (e-i) Growth of WT (left) and *svd* mutant (right) in MS media for three weeks in control with standard MS (e), addition of 30 µM CuSO4 (f), 3mM MnSO4 (g), 0.9 mM ZnSO4 (h), and 3 mM FeSO4 (i), (j) Relative expression of SVD by RT-qPCR in different tissues of sorghum, (k) Induction of SVD expression in sorghum leaves by treatment with different concentrations of

heavy metals. Asteriks indication significant differences from the control (**p < 0.01, ***p < 0.001, student t-test).

3.4.7. *SVD* gene expression is induced by Cu and Mn treatment

Reverse transcription quantitative PCR (RT-qPCR) indicated that the *SVD* gene is expressed in most tissues including leaves, roots and seeds with highest expression in the leaf sheath (Fig. 17j). To evaluate the effect of heavy metal treatment on SVD expression, plants were grown on culture media supplemented with 30 μM CuSO4, 3mM MnSO4, 0.9mM ZnSO4, and 3mM FeSO4, and *SVD* expression was analyzed in leaves by RT-qPCR. We found that *SVD* expression was highly induced by $CuSO_4$ and to a lower extent by $MnSO_4$ but not by $ZnSO_4$ and $FeSO_4$ (Fig. 17k). This suggests that cells can sense excessive levels of these ions and induce the expression of *SVD*, perhaps to remove the excess Cu, consistent with the possible role of *SVD* in maintaining Cu homeostasis as an efflux transporter.

3.4.8. SVD is localized primarily in the plasma membrane

To determine the subcellular localization of SVD, the full length *SVD* CDS was fused to the C-terminus of GFP in the Gateway vector pGWB505 and the resulting plasmid transformed into *Agrobacterium tumefaciens* strain GV2260. pGWB505 was used as a negative control. GV2260 was then infiltrated into 4 week old *Nicotiana benthamiana* leaves and examined by confocal microscopy 48-60 hrs after infiltration. Images were taken together with red channel for chloroplast auto florescence and DIC for cell outline as controls. While no GFP florescence was observed in the pGWB505 control leaves (Fig. 18a-d), strong green florescence was observed in leaves infiltrated with the pGWB505-

SVD construct localized along the plasma membrane (Fig. 18e-h). These results indicate that SVD is primarily localized to the plasma membrane. This is consistent with the putative role of SVD in pumping excess Cu out of the cell to maintain Cu homeostasis in the cytosol. Some diffused GFP signal was also detected in the endoplasmic reticulum (Fig. 18e).

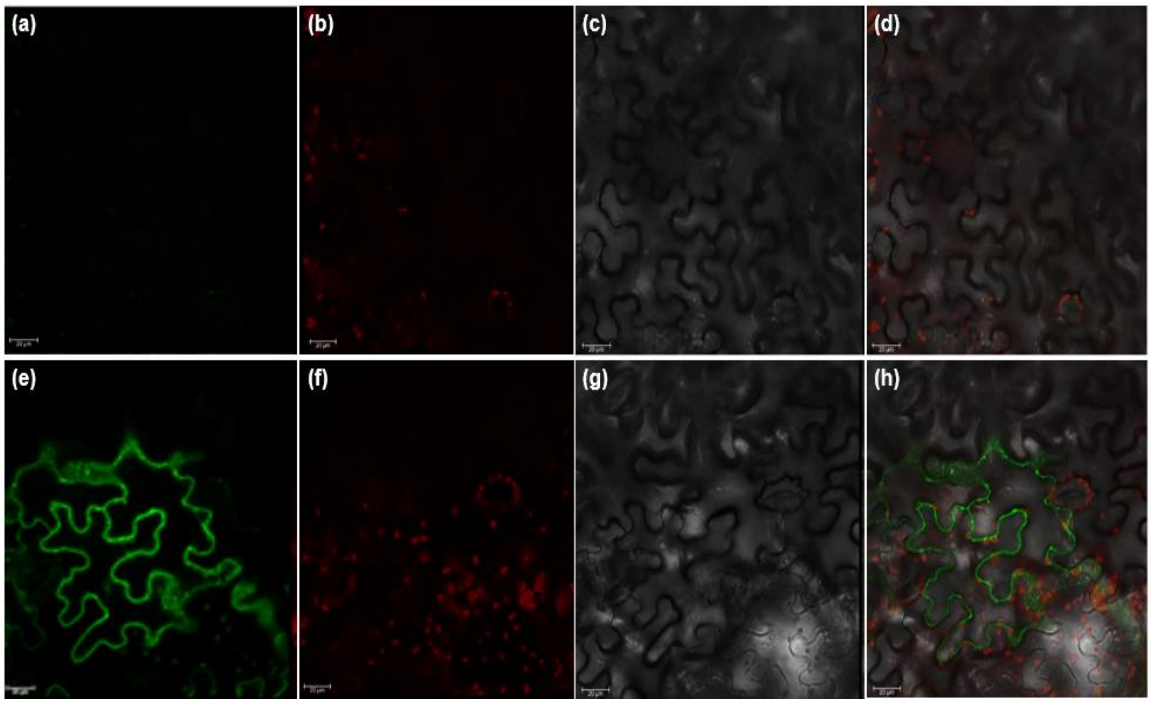

Fig. 18) SVD-GFP fusion protein is predominantly localized in the plasma membrane

a) YFP channel image (excitation 514 nm, emission 524-599 nm) for the control, b) red channel for the auto fluorescence of the chloroplast (excitation 631 nm, emission 680-755 nm) for the control, c) DIC image (Bright field) for the control, d) Merged all above for the control, e) YFP channel image (excitation 514 nm, emission 524-599 nm) for the sample, f) red channel for the auto fluorescence of the chloroplast (excitation 631 nm, emission 680-755 nm) for the sample, g) DIC image (Bright field) for the sample, h) Merged all above for the sample

3.4.9. SVD physically interacts with SbATX1 and SbFRN3 but not with SbHMT1 and SbCuB-like

P_{1B}-type heavy metal ATPases often function in a complex with other proteins, called metallochaperones, that are reported to serve as metal receptors and facilitate safe

transport of the metal ions (Arnesano et al., 2002; de Abreu-Neto et al., 2013). Thus, based on bioinformatic prediction using the STRING v11 database (Szklarczyk et al., 2019), ten proteins were identified in the sorghum genome that could potentially interact with SVD. This includes homologues of the yeast antioxidant protein ATX1 (SbATX1; Sb08g017782.1), farnesylated related protein (SbFRN3; Sb07g020020.1), heavy metal related protein (SbHMT1; Sb08g022260.1), SbCuB-like (SbCuBL; Sb10g021320.1) and six others (Fig. 19). Nine of the ten predicted interactors share a single conserved CxxC (red arrows) motif and a few other residues in the 70aa heavy metal binding regions (Fig. 20).

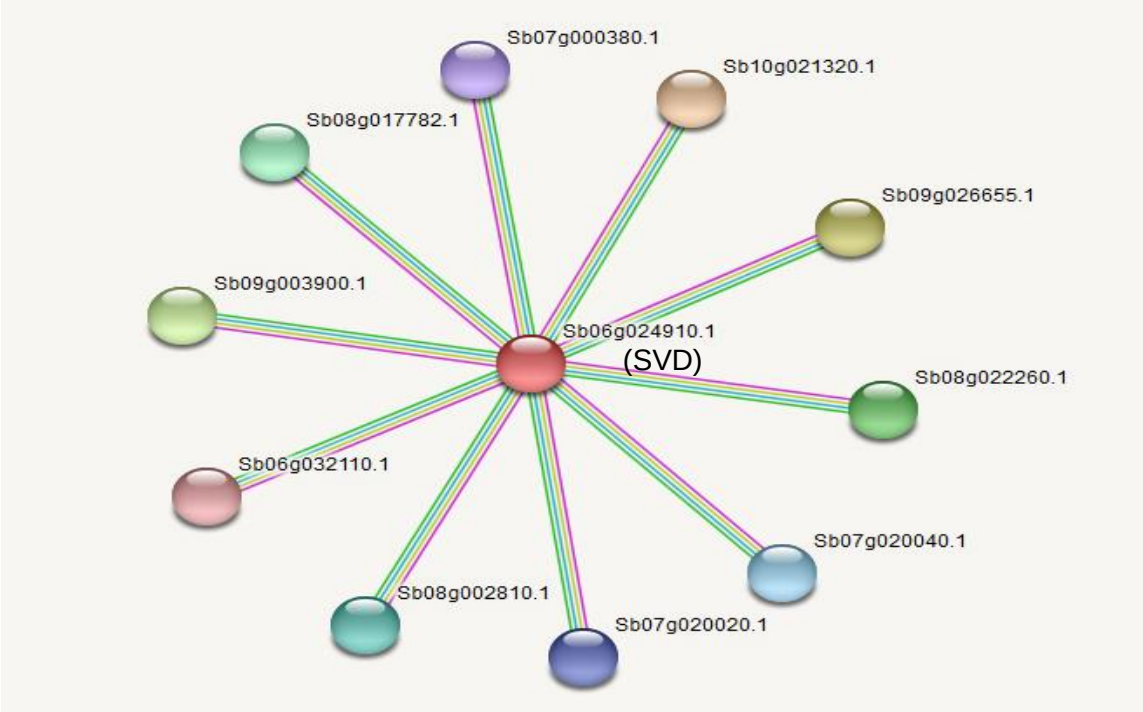

Fig. 19) In silico predicted potential SVD protein interaction partners in the sorghum genome using STRING.

Blast search of the genes from the string program gives characterized and predicted protein names such as ATX1 (SbATX1; Sb08g017782.1), farnesylated related protein (SbFRN3; Sb07g020020.1), heavy metal related protein (SbHMT1; Sb08g022260.1), SbCuB-like (SbCuBL; Sb10g021320.1), a detoxification like

protein (SbDETO; Sb07g000380), SbB0811B10; Sb09g026655, (IPP23; Sb07g020040), SbIPP33; Sb08g002810, SbATOX1;Sb06g032110, SbJNBa; Sb09g003900 were used.

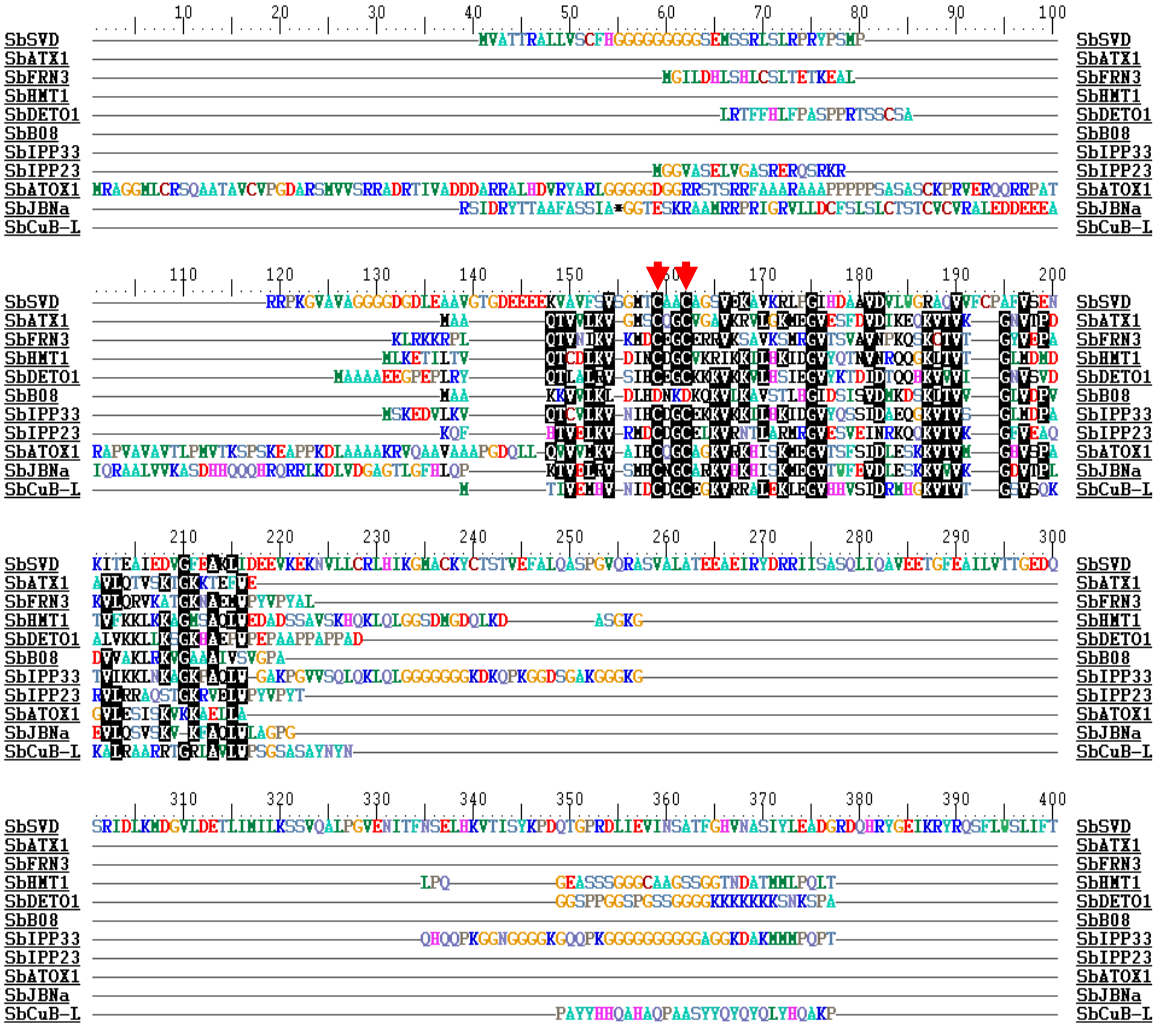

Fig. 20) Multiple amino acid alignment of SVD with bioinformatically predicted potential interactor metallochaperone like proteins

Multiple amino acid sequence alignment was performed using MUSCLE to see the conserved domains of the potential SVD interactors. Metallochaperone proteins identified using the STRING program including ATX1 (SbATX1; Sb08g017782.1), farnesylated related protein (SbFRN3; Sb07g020020.1), heavy metal related protein (SbHMT1; Sb08g022260.1), SbCuB-like (SbCuBL; Sb10g021320.1), a detoxification like protein (SbDETO; Sb07g000380),

SbB0811B10; Sb09g026655, (IPP23; Sb07g020040), SbIPP33; Sb08g002810, SbATOX1;Sb06g032110, SbJNBa; Sb09g003900 were used.

We tested four of these by yeast two hybrid (Y2H) analysis to see which of these proteins indeed physically interact with SVD. The SVD protein was cloned in the pGBKT7-GW plasmid as bait and each of the candidate proteins was cloned in pGADT7-GW plasmid as prey, and pairwise interaction was evaluated in yeast Gold strain (Takara Bio). We found that the SVD protein interacted with SbATX1 and SbFRN3 but not with SbHMT1and SbCuBL (Fig. 21a-d). These Y2H interactions were validated in living plant cells with bimolecular fluorescence complementation (BiFC) using split YFP in which SVD was fused to the N-terminal half of YFP and the partner proteins were fused to the c-terminal half of YFP. Consistent with the Y2H result, yellow fluorescence due to the close proximity of N-YFP and C-YFP brought about by protein-protein interaction was observed in SVD-SbATX1 and SVD-SbFRN3 pairs but not with SVD-SbHMT1 and SVD-SbCuB-like pairs (Fig. 21e-l, Fig. 22), confirming that SVD physically interacts with SbATX1 and SbFRN3 in yeast and plant cells. Taken together, our data indicate that SVD/SbHMA5 is critically required at all stages of sorghum growth and development, and functions in Cu homeostasis by extruding excess Cu out of the cell via plasma membrane, perhaps with the help of metallochaperones SbATX1 and SbFRN3.

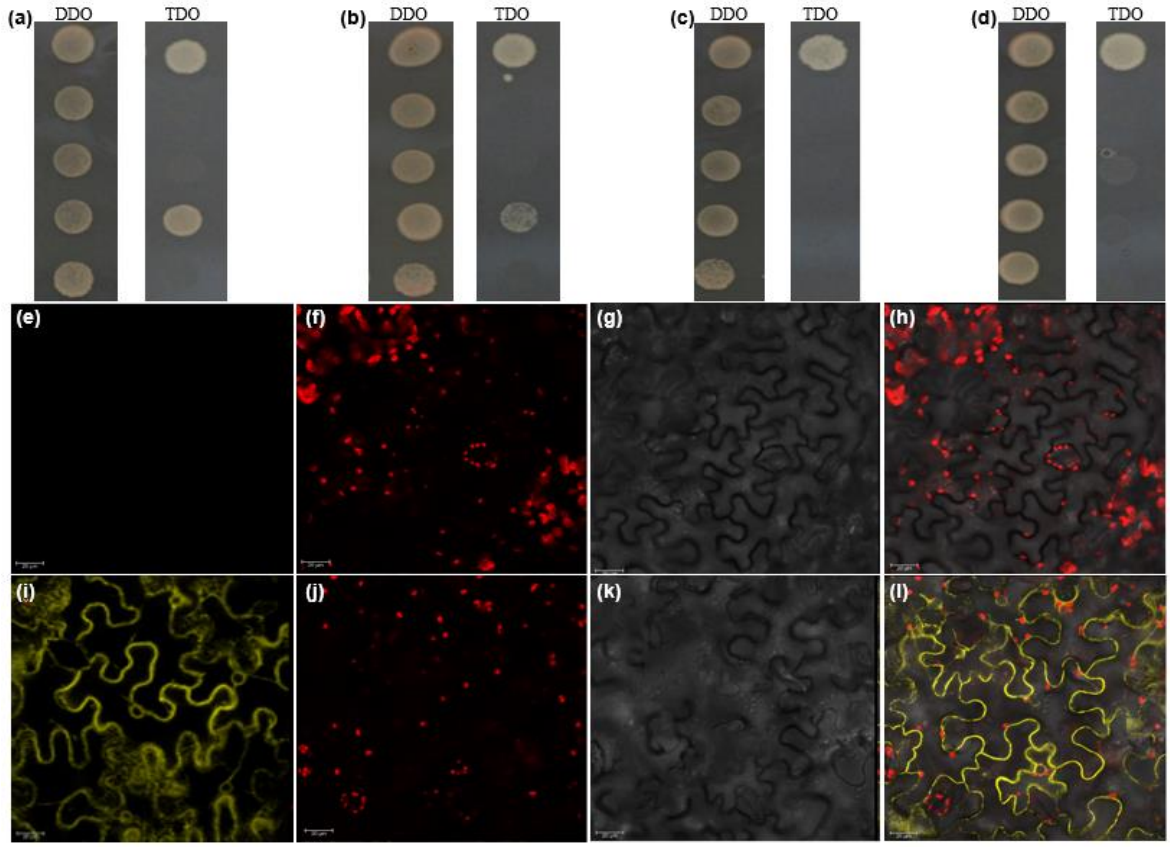

Fig. 21) SVD physically interacts with metallochaperones SbATX1 and SbFRN3.

(a-d) Yeast two hybrid (Y2H) interaction of SVD with SbATX1, SbFRN3, SbHMT1, and SbCuB-like. Shown are yeast growth analysis on synthetic minimal double dropout media (DDO) deficient in SD/-Leu/-Trp for control and synthetic minimal triple dropout media(TDO) deficient in SD/-His/-Trp/-Leu for testing the interactions.(a) SVD with SbATX, (b) SVD with SbFRN3, (c) SVD with SbHMT1, (d) SVD with SbCuB-like. First row - positive control (MtSTF-AD and MtWOX9-BD), second row – negative controls (empty AD and BD) and fifth row – SVD-AD and empty BD for all (a-d). Third row- negative controls; empty AD and SbATX1-BD (a), empty AD and SbFRN3-BD (b), empty AD and SbHMT1-BD (c) and empty AD and SbCuBL-BD (d). Fourth row – actual testing SVD-AD and ATX1-BD (a), SVD-AD and SbFRN3-BD (b), SVD-AD and SbHMT1-BD (c), and SVD-AD and SbCuBL-BD (d),(e-l) BiFC analysis of SVD and SbATX1 in tobacco leaf cells, (e-h) Negative control (SVD-NYFP and CYFP), (i-l) Test (SVD-NYFP and SbATX1-CYFP), (e,i) YFP channel (excitation 514 nm and emission 524-599 nm), (f,j) Red channel autofluorescence (excitation 631 nm and emission 680-755 nm), (g,k) DIC image (bright field), (h,l) Merged image. Bars = 20 µM.

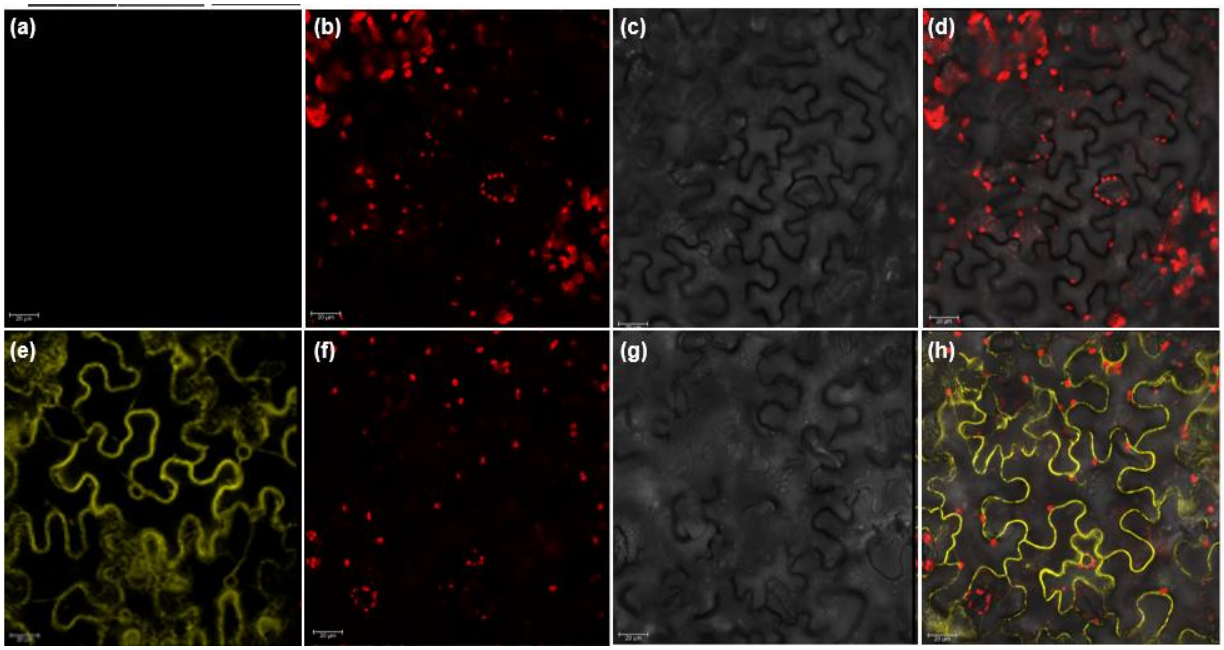

Fig.22) BiFC analysis of the interaction between SVD and SbFRN3 in tobacco leaves

a) YFP channel image (excitation 514nm, emission 524-599nm) for the control b) red channel for the autofluorescence of the chloroplast (excitation 631nm, emission 680-755nm) for the control c) DIC image (Bright field) for the control d) Merged all above for the control e) YFP channel image (excitation 514nm, emission 524-599nm) for the sample f) red channel for the autofluorescence of the chloroplast (excitation 631nm, emission 680-755nm) for the sample g) DIC image (Bright field) for the sample h) Merged all above for the sample

3.5. Discussion

3.5.1. The SVD gene encodes a P_{1B}-type Cu ATPase involved in copper detoxification

We showed that the *shriveled seed* (*svd*) mutant of sorghum has a severe defect in plant growth throughout development from seed germination to grain filling, accumulates more copper in the leaves and seeds, and is more sensitive to $CuSO_4$ treatment compared to WT plants (Figs. 6, 17). We also demonstrated that these phenotypes co-segregated with a 2-bp deletion within the fifth exon. The *SVD* encodes a plasma membrane localized *HMA5* homologue (Fig. 11), a P_{1B}-type Cu ATPase that is likely to be involved in Cu homeostasis.

Although Cu is required for cellular functions in organisms from bacteria to humans, its concentration in the cell must be tightly regulated as excess Cu is toxic, causing oxidative damage to DNA, membrane lipids, and proteins leading to disruption of cellular functions (Pena et al., 1999; Mitra et al., 2004; Burkhead et al., 2009; Rosenzweig and Arguello, 2012). In humans, while Cu deficiency due to failure of the ATP7A gene causes Menkes disease, which affects hematologic, neurologic, cardiovascular and immune systems (Kaler, 2011; Guthrie et al., 2020), Cu accumulation due to mutation of the ATP7B gene causes Wilson disease, a life-threatening condition that affects the brain, liver, kidney and other organs (Patil et al., 2013; Aggarwal and Bhatt, 2020). In Arabidopsis, PAA1/AtHMA6 and PAA2/AtHMA8 deliver Cu to the chloroplast and the *paa1 paa2* double mutant is seedling lethal (Abdel-Ghany et al., 2005), indicating the critical importance of Cu to photosynthesis. AtHMA5, the closest Arabidopsis homologue of SVD/SbHMA5, on the other hand, is reported to be involved in Cu detoxification in the roots by extruding Cu out of the cytosol (Andres-Colas et al., 2006). Similar to the *svd* mutant, the *hma5* mutants are sensitive to excess Cu and display arrested root growth with Cu treatment (Andres-Colas et al., 2006). However, the severe developmental defects observed in *svd* mutants in soil and standard MS media were not displayed by *hma5* mutants of Arabidopsis, suggesting that other members may have some level of redundancy with HMA5 function in Arabidopsis. Nevertheless, the topology of the predicted SVD protein is very similar to AtHMA5 and OsHMA5 with multiple highly conserved motifs (Fig. 11) typical of P_{1B}-type Cu ATPases.

Among the nine P_{1B}-type ATPases (OsHMAs) found in the rice genome, OsHMA2 and OsHMA3 are reported to transport Zn and Cd (Satoh-Nagasawa et al., 2012; Takahashi

et al., 2012; Yamaji et al., 2012; Sasaki et al., 2014). OsHMA5 is the closest rice homolog to sorghum SVD (SbHMA5) sharing 85.7% amino acid identity over the full length protein (Fig. 14). Similar to SVD, it is located at the plasma membrane, and it is involved in Cu loading to the xylem of rice roots for root-to-shoot translocation (Deng et al., 2013). OsHMA4, the next closest rice homologue of SVD (Fig.14), is involved in Cu sequestration in the root vacuoles, preventing Cu accumulation in the rice grain (Huang et al., 2016). OsHMA6 also localizes to the plasma membrane and likely extrudes Cu (Wenli et al., 2020). OsHMA9, on the other hand, is reported to be an efflux carrier for Cu, Zn and Pb, removing these heavy metals from cells (Lee et al., 2007). Though OsHMA9 is involved in multiple heavy metal detoxification, it is related to SVD/SbHMA5 in being plasma membrane localized, induced by Cu treatment, and pumps excess Cu from the cell (Lee et al., 2007). Functional validation is not yet available for the remaining OsHMAs but conserved domains and phylogenetic analysis predict that six of the nine rice HMAs (OsHMA4 – OsHMA9) may be involved in Cu transport (Wenli et al., 2020) analogous to the four (AtHMA5 – AtHMA8) P_{1B}-type Cu ATPases described in Arabidopsis (Burkhead et al., 2009). In *Silene vulgaris* adapted to Cu-polluted soils, two HMA5 paralogs, SvHMA5I and SvHMA5II, function in Cu detoxification via sequestration in the vacuole and extruding through the plasma membrane, respectively (Li et al., 2017). Our conserved domain and phylogenetic analysis (Figs. 11, 14) indicate that SVD/SbHMA5 belongs to this group of Cu transporters and localizes primarily to the plasma membrane although some signal in the endoplasmic reticulum could be detected (Fig.18). The accumulation of more Cu in both leaf and seed tissues of *svd* mutants and their hypersensitivity to exogenous Cu treatment is consistent with SVD/SbHMA5 being involved in Cu

detoxification and homeostasis by preventing accumulation of Cu in sorghum tissues through extruding Cu out of the cells.

Compared with nine OsHMAs in rice and eight AtHMAs in Arabidopsis, there are eleven SbHMAs in the sorghum genome. Out of these, HMA5 and HMA3 appeared to be duplicated resulting in SbHMA5 and SbHMA5L, and SbHMA3 and SbHMA3L, respectively (Fig. 14). These were tandem duplications very close to each other; HMA5 on chromosome six and HMA3 on chromosome two (Fig. 15) suggesting that the duplication occurred after the separation of the sorghum and rice lineages , but before the sorghum and *Panicum* divergence. Whether the duplicate genes perform identical functions is unknown. It would be interesting to investigate the significance of this duplication and the higher number of HMAs in sorghum with respect to Cu toxicity tolerance and plant development in general.

3.5.2. SVD physically interacts with metallochaperones SbATX1 and SbFRN3

Metallochaperones are a class of soluble metal ion receptor proteins that bind and deliver metal ions to intracellular destinations (Rosenzweig, 2002a; Robinson and Winge, 2010; Sekhon, 2010; Capdevila et al., 2017). They are required to prevent the deleterious reactions of heavy metals and facilitate safe transport to different cellular compartments (Rosenzweig, 2002a; de Abreu-Neto et al., 2013). Cu(I) and Cu(II), for example, are highly reactive ions that form tight complexes with organic molecules causing damage through generation of reactive oxygen species (Robinson and Winge, 2010). These ions are not available free in the cytosol, instead, they are bound by copper chaperones and delivered to P_{1B}-type Cu ATPases for transport to copper enzymes and endomembrane sites (Rosenzweig, 2002a; Gonzalez-Guerrero and Arguello, 2008; Robinson and Winge, 2010).

The delivery of Cu from the Cu chaperone to the Cu ATPase is proposed to be mediated by protein-protein interaction through a conserved metal binding CxxC motif found in both partners (Fig. 19). Using a web-based tool, we looked for potential interactors of SVD in the sorghum genome and found ten potential metallochaperone-like candidates with conserved metal binding domains (Figs. 19, 20). Of these, two were confirmed to physically interact with SVD in yeast and plant cells through Y2H and BiFC analysis (Figs. 21, 22). The positive interactors were homologues of ATX1 and FRN3. FRN3/HIPP22 is not well-characterized in Arabidopsis, except that it is reported to complement the Cd-sensitive yeast *ycf1* mutant (Tehseen et al., 2010). The yeast antioxidant protein (ATX1) was reported to deliver Cu to and physically interact with the yeast P-type ATPase Ccc2 (Lin and Culotta, 1995; Pufahl et al., 1997; Huffman and O'Halloran, 2000). Similarly, the human ATX1 homologue, ATOX1 was shown to interact with ATP7A and ATP7B to safely and efficiently transport Cu to intracellular destinations (Hung et al., 1998; Hamza et al., 1999; Walker et al., 2002). In plants, while Arabidopsis AtATX1 was demonstrated to interact with HMA5 and RAN1/HMA6 to detoxify Cu and deliver Cu to the ethylene receptor, respectively (Andres-Colas et al., 2006; Puig et al., 2007; Li et al., 2017), rice OsATX1 is shown to interact with OsHMA4, OsHMA5, OsHMA6 and OsHMA9 (Zhang et al., 2018), suggesting that OsATX1 may deliver Cu to these P_{1B}-type Cu ATPases for both delivery to intracellular locations and removal from the cytosol. Thus, while the interaction of SVD with SbATX1 could be functionally significant to Cu homeostasis, the contribution of SbATX1 has not yet been evaluated. Taken together, our data demonstrate that Cu homeostasis is critically required for plant growth and development in sorghum from seed germination to grain filling and this is primarily accomplished by a P_{1B}-type Cu

ATPase SVD (SbHMA5) that efflux Cu from the cytosol. However, while my data is not conclusive in this regard, there is still a possibility that there may also be other *SVD* gene functions which needs further investigation.

3.6. References

Abdel-Ghany, S.E., Day, I.S., Simmons, M.P., Kugrens, P., and Reddy, A.S. (2005). Origin and evolution of Kinesin-like calmodulin-binding protein. Plant Physiol **138,** 1711-1722.

Aggarwal, A., and Bhatt, M. (2020). Wilson disease. Curr Opin Neurol.

Andres-Colas, N., Sancenon, V., Rodriguez-Navarro, S., Mayo, S., Thiele, D.J., Ecker, J.R., Puig, S., and Penarrubia, L. (2006). The Arabidopsis heavy metal P-type ATPase HMA5 interacts with metallochaperones and functions in copper detoxification of roots. Plant J **45,** 225-236.

Arguello, J.M. (2003). Identification of ion-selectivity determinants in heavy-metal transport P1B-type ATPases. J Membr Biol **195,** 93-108.

Arnesano, F., Banci, L., Bertini, I., Ciofi-Baffoni, S., Molteni, E., Huffman, D.L., and O'Halloran, T.V. (2002). Metallochaperones and metal-transporting ATPases: a comparative analysis of sequences and structures. Genome Res **12,** 255-271.

Axelsen, K.B., and Palmgren, M.G. (2001). Inventory of the Superfamily of P-Type Ion Pumps in Arabidopsis. Plant Physiol **126**.

Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Rudnev, D., Evangelista, C., Kim, I.F., Soboleva, A., Tomashevsky, M., and Edgar, R. (2007). NCBI GEO: mining tens of millions of expression profiles--database and tools update. Nucleic Acids Res **35,** D760-765.

Baxter, I. (2009). Ionomics: studying the social network of mineral nutrients. Curr Opin Plant Biol **12,** 381-386.

Baxter, I., and Dilkes, B.P. (2012). Elemental profiles reflect plant adaptations to the environment. Science **336,** 1661-1663.

Baxter, I., Tchieu, J., Sussman, M.R., Boutry, M., Palmgren, M.G., Gribskov, M., Harper, J.F., and Axelsen, K.B. (2003). Genomic comparison of P-type ATPase ion pumps in Arabidopsis and rice. Plant Physiol **132,** 618-628.

Bruckner, A., Polge, C., Lentze, N., Auerbach, D., and Schlattner, U. (2009). Yeast two-hybrid, a powerful tool for systems biology. Int J Mol Sci **10,** 2763-2788.

Burkhead, J.L., Reynolds, K.A., Abdel-Ghany, S.E., Cohu, C.M., and Pilon, M. (2009). Copper homeostasis. New Phytol **182,** 799-816.

Cai, H., Huang, S., Che, J., Yamaji, N., and Ma, J.F. (2019). The tonoplast-localized transporter OsHMA3 plays an important role in maintaining Zn homeostasis in rice. J Exp Bot **70,** 2717-2725.

Capdevila, D.A., Edmonds, K.A., and Giedroc, D.P. (2017). Metallochaperones and metalloregulation in bacteria. Essays Biochem **61,** 177-200.

Chandrasekhar, C., and Ray, J.G. (2017). Copper accumulation, localization and antioxidant response in Eclipta alba L. in relation to quantitative variation of the metal in soil. Acta Physiologiae Plantarum **39**.

Clemens, S., and Ma, J.F. (2016). Toxic Heavy Metal and Metalloid Accumulation in Crop Plants and Foods. Annu Rev Plant Biol **67,** 489-512.

de Abreu-Neto, J.B., Turchetto-Zolet, A.C., de Oliveira, L.F., Zanettini, M.H., and Margis-Pinheiro, M. (2013). Heavy metal-associated isoprenylated plant protein (HIPP): characterization of a family of proteins exclusive to plants. FEBS J **280,** 1604-1616.

Deng, F., Yamaji, N., Xia, J., and Ma, J.F. (2013). A member of the heavy metal P-type ATPase OsHMA5 is involved in xylem loading of copper in rice. Plant Physiol **163,** 1353-1362.

Eren, E., and Arguello, J.M. (2004). Arabidopsis HMA2, a divalent heavy metal-transporting P(IB)-type ATPase, is involved in cytoplasmic Zn2+ homeostasis. Plant Physiol **136,** 3712-3723.

Fan, J.Y., Cui, Z.Q., Wei, H.P., Zhang, Z.P., Zhou, Y.F., Wang, Y.P., and Zhang, X.E. (2008). Split mCherry as a new red bimolecular fluorescence complementation system for visualizing protein-protein interactions in living cells. Biochem Biophys Res Commun **367,** 47-53.

Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. Nature **340**.

Fields, S., and Sternglanz, R. (1994). The two-hybrid system: an assay for protein-protein interactions. TIG Reviews.

Gall, J.E., Boyd, R.S., and Rajakaruna, N. (2015). Transfer of heavy metals through terrestrial food webs: a review. Environ Monit Assess **187,** 201.

Garcia-Molina, A., Andres-Colas, N., Perea-Garcia, A., Neumann, U., Dodani, S.C., Huijser, P., Penarrubia, L., and Puig, S. (2013). The Arabidopsis COPT6 transport protein functions in copper distribution under copper-deficient conditions. Plant Cell Physiol **54,** 1378-1390.

Gietz, R.D., and Schiestl, R.H. (2007). Quick and easy yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat Protoc **2,** 35-37.

Glory, E., and Murphy, R.F. (2007). Automated subcellular location determination and high-throughput microscopy. Dev Cell **12,** 7-16.

Gonzalez-Guerrero, M., and Arguello, J.M. (2008). Mechanism of Cu+-transporting ATPases: soluble Cu+ chaperones directly transfer Cu+ to transmembrane transport sites. Proc Natl Acad Sci U S A **105,** 5992-5997.

Guthrie, L.M., Soma, S., Yuan, S., Silva, A., Zulkifli, M., Snavely, T.C., Greene, H.F., Nunez, E., Lynch, B., De Ville, C., Shanbhag, V., Lopez, F.R., Acharya, A., Petris, M.J., Kim, B.E., Gohil, V.M., and Sacchettini, J.C. (2020). Elesclomol alleviates Menkes pathology and mortality by escorting Cu to cuproenzymes in mice. Science **368,** 620-625.

Hamdi, A., and Colas, P. (2012). Yeast two-hybrid methods and their applications in drug discovery. Trends Pharmacol Sci **33,** 109-118.

Hamza, I., Schaefer, M., Klomp, L.W., and Gitlin, J.D. (1999). Interaction of the copper chaperone HAH1 with the Wilson disease protein is essential for copper homeostasis. Proc Natl Acad Sci U S A **96,** 13363-13368.

Hirayama, T., Kieber, J.J., Hirayama, N., Kogan, M., Guzman, P., Nourizadeh, S., Alonso, J.M., Dailey, W.P., Dancis, A., and Ecker, J.R. (1999). RESPONSIVE-TO-ANTAGONIST1, a Menkes/Wilson disease-related copper transporter, is required for ethylene signaling in Arabidopsis. Cell **97,** 383-393.

Huang, X.Y., Deng, F., Yamaji, N., Pinson, S.R., Fujii-Kashino, M., Danku, J., Douglas, A., Guerinot, M.L., Salt, D.E., and Ma, J.F. (2016). A heavy metal P-type ATPase OsHMA4 prevents copper accumulation in rice grain. Nat Commun **7,** 12138.

Huffman, D.L., and O'Halloran, T.V. (2000). Energetics of copper trafficking between the Atx1 metallochaperone and the intracellular copper transporter, Ccc2. J Biol Chem **275,** 18611-18614.

Hung, I.H., Casareno, R.L., Labesse, G., Mathews, F.S., and Gitlin, J.D. (1998). HAH1 is a copper-binding protein with distinct amino acid residues mediating copper homeostasis and antioxidant defense. J Biol Chem **273,** 1749-1754.

Hussain, D., Haydon, M.J., Wang, Y., Wong, E., Sherson, S.M., Young, J., Camakaris, J., Harper, J.F., and Cobbett, C.S. (2004). P-type ATPase heavy metal transporters with roles in essential zinc homeostasis in Arabidopsis. Plant Cell **16,** 1327-1339.

Kaler, S.G. (2011). ATP7A-related copper transport diseases-emerging concepts and future trends. Nat Rev Neurol **7,** 15-29.

Kampfenkel, K., Kushnir, S., Babiychuk, E., Inze´, D., and Van Montagu, M. (1995). Molecular Characterization of a Putative Arabidopsis thaliana Copper Transporter and Its Yeast Homologue. THE JOURNAL OF BIOLOGICAL CHEMISTRY **270**.

Kerppola, T.K. (2006a). Complementary methods for studies of protein interactions in living cells. In Nature Methods.

Kerppola, T.K. (2006b). Design and Implementation of Bimolecular Fluorescence Complementation (BiFC) Assays for the Visualization of Protein Interactions in Living Cells. Nat Protoc **1**.

Kerppola, T.K. (2008). Bimolecular fluorescence complementation (BiFC) analysis as a probe of protein interactions in living cells. Annu Rev Biophys **37,** 465-487.

Kim, Y.Y., Choi, H., Segami, S., Cho, H.T., Martinoia, E., Maeshima, M., and Lee, Y. (2009). AtHMA1 contributes to the detoxification of excess Zn(II) in Arabidopsis. Plant J **58,** 737-753.

Klaumann, S., Nickolaus, S.D., Furst, S.H., Starck, S., Schneider, S., Ekkehard Neuhaus, H., and Trentmann, O. (2011). The tonoplast copper transporter COPT5 acts as an exporter and is required for interorgan allocation of copper in Arabidopsis thaliana. New Phytol **192,** 393-404.

Lee, S., Kim, Y.Y., Lee, Y., and An, G. (2007). Rice P1B-type heavy-metal ATPase, OsHMA9, is a metal efflux protein. Plant Physiol **145,** 831-842.

Li, Y., Iqbal, M., Zhang, Q., Spelt, C., Bliek, M., Hakvoort, H.W.J., Quattrocchio, F.M., Koes, R., and Schat, H. (2017). Two Silene vulgaris copper transporters residing in different cellular compartments confer copper hypertolerance by distinct mechanisms when expressed in Arabidopsis thaliana. New Phytol **215,** 1102-1114.

Lin, S.J., and Culotta, V.C. (1995). The ATX1 gene of Saccharomyces cerevisiae encodes a small metal homeostasis factor that protects cells against reactive oxygen toxicity. Proc Natl Acad Sci U S A **92,** 3784-3788.

Luo, Y., Batalao, A., Zhou, H., and Zhu, L. (1997). Mammalian Two-Hybrid System: A Complementary Approach to the Yeast Two-Hybrid System. Bio Techniques **22**.

Marschner, J.P., and Rietbrock, N. (1995). Oxygen release kinetics in healthy subjects and diabetic patients. II: Effects of HbCO. Int J Clin Pharmacol Ther **33,** 263-265.

Meharg, A., A, . (1994). Integrated tolerance mechanisms: constitutive and adaptive plant responses to elevated metal concentrations in the environment. Plant Cell Environ. **17,** 989-993.

Miller, T.A., Wittenberg, J.S., Wen, H., Connor, S., Cui, Y., and Lindenberg, A.M. (2013). The mechanism of ultrafast structural switching in superionic copper (I) sulphide nanocrystals. Nat Commun **4,** 1369.

Mills, R.F., Krijger, G.C., Baccarini, P.J., Hall, J.L., and Williams, L.E. (2003). Functional expression of AtHMA4, a P1B-type ATPase of the Zn/Co/Cd/Pb subclass. Plant J **35,** 164-176.

Mills, R.F., Francini, A., Ferreira da Rocha, P.S., Baccarini, P.J., Aylett, M., Krijger, G.C., and Williams, L.E. (2005). The plant P1B-type ATPase AtHMA4 transports Zn and Cd and plays a role in detoxification of transition metals supplied at elevated levels. FEBS Lett **579,** 783-791.

Mitra, R.M., Gleason, C.A., Edwards, A., Hadfield, J., Downie, J.A., Oldroyd, G.E., and Long, S.R. (2004). A Ca2+/calmodulin-dependent protein kinase required for symbiotic nodule development: Gene identification by transcript-based cloning. Proc Natl Acad Sci U S A **101,** 4701-4705.

Morell, M., Espargaro, A., Aviles, F.X., and Ventura, S. (2008a). Study and selection of in vivo protein interactions by coupling bimolecular fluorescence complementation and flow cytometry. Nat Protoc **3,** 22-33.

Morell, M., Czihal, P., Hoffmann, R., Otvos, L., Aviles, F.X., and Ventura, S. (2008b). Monitoring the interference of protein-protein interactions in vivo by

bimolecular fluorescence complementation: the DnaK case. Proteomics **8,** 3433-3442.

Palmgren, M.G., and Nissen, P. (2011). P-type ATPases. Annu Rev Biophys **40,** 243-266.

Patil, M., Sheth, K.A., Krishnamurthy, A.C., and Devarbhavi, H. (2013). A review and current perspective on Wilson disease. J Clin Exp Hepatol **3,** 321-336.

Pena, M.M., Lee, J., and Thiele, D.J. (1999). A delicate balance: homeostatic control of copper uptake and distribution. J Nutr **129,** 1251-1260.

Petris, M.J., Voskoboinik, I., Cater, M., Smith, K., Kim, B.E., Llanos, R.M., Strausak, D., Camakaris, J., and Mercer, J.F. (2002). Copper-regulated trafficking of the Menkes disease copper ATPase is associated with formation of a phosphorylated catalytic intermediate. J Biol Chem **277,** 46736-46742.

Pufahl, R.A., Singer, C.P., Peariso, K.L., Lin, S.J., Schmidt, P.J., Fahrni, C.J., Culotta, V.C., Penner-Hahn, J.E., and O'Halloran, T.V. (1997). Metal ion chaperone function of the soluble Cu(I) receptor Atx1. Science **278,** 853-856.

Puig, S., Lee, J., Lau, M., and Thiele, D.J. (2002). Biochemical and genetic analyses of yeast and human high affinity copper transporters suggest a conserved mechanism for copper uptake. J Biol Chem **277,** 26021-26030.

Puig, S., Andres-Colas, N., Garcia-Molina, A., and Penarrubia, L. (2007). Copper and iron homeostasis in Arabidopsis: responses to metal deficiencies, interactions and biotechnological applications. Plant Cell Environ **30,** 271-290.

Rai, P.K., Lee, S.S., Zhang, M., Tsang, Y.F., and Kim, K.H. (2019). Heavy metals in food crops: Health risks, fate, mechanisms, and management. Environ Int **125,** 365-385.

Raven, J.A., Evans, M.C., and Korb, R.E. (1999). The role of trace metals in photosynthetic electron transport in O2-evolving organisms. Photosynth. Res. **60,** 111-149.

Ravet, K., and Pilon, M. (2013). Copper and iron homeostasis in plants: the challenges of oxidative stress. Antioxid Redox Signal **19,** 919-932.

Rehman, M.U., and Rather, I.A. (2019). Myricetin Abrogates Cisplatin-Induced Oxidative Stress, Inflammatory Response, and Goblet Cell Disintegration in Colon of Wistar Rats. Plants (Basel) **9**.

Robinson, N.J., and Winge, D.R. (2010). Copper metallochaperones. Annu Rev Biochem **79,** 537-562.

Rosenzweig, A.C. (2001). Copper delivery by metallochaperone proteins. Acc Chem Res **34,** 119-128.

Rosenzweig, A.C. (2002a). Metallochaperones: Bind and Deliver. Chemistry & Biology **9,** 673-677.

Rosenzweig, A.C. (2002b). Metallochaperones: bind and deliver. Chem Biol **9,** 673-677.

Rosenzweig, A.C., and Arguello, J.M. (2012). Toward a molecular understanding of metal transport by P(1B)-type ATPases. Curr Top Membr **69,** 113-136.

Sancenon, V., Puig, S., Mateu-Andres, I., Dorcey, E., Thiele, D.J., and Penarrubia, L. (2004). The Arabidopsis copper transporter COPT1 functions in root elongation and pollen development. J Biol Chem **279,** 15348-15355.

Sasaki, T., Iizuka, A., Watanabe, M., Hongo, T., and Yamasaki, A. (2014). Preparation and performance of arsenate (V) adsorbents derived from concrete wastes. Waste Manag **34,** 1829-1835.

Satoh-Nagasawa, N., Mori, M., Nakazawa, N., Kawamoto, T., Nagato, Y., Sakurai, K., Takahashi, H., Watanabe, A., and Akagi, H. (2012). Mutations in rice (Oryza sativa) heavy metal ATPase 2 (OsHMA2) restrict the translocation of zinc and cadmium. Plant Cell Physiol **53,** 213-224.

Sekhon, S.B. (2010). Metallochaperones - an Overview. Current Chemical Biology **4,** 173-186.

Shoshan, M.S., Shalev, D.E., Adriaens, W., Merkx, M., Hackeng, T.M., and Tshuva, E.Y. (2011). NMR characterization of a Cu(I)-bound peptide model of copper metallochaperones: insights on the role of methionine. Chem Commun (Camb) **47,** 6407-6409.

Shyu, Y.J., Liu, H., Deng, X., and Hu, C.D. (2006). Identification of new fluorescent protein fragments for bimolecular fluorescence complementation analysis under physiological conditions. Biotechniques **40,** 61-66.

Snider, J., Kittanakom, S., Curak, J., and Stagljar, I. (2010). Split-ubiquitin based membrane yeast two-hybrid (MYTH) system: a powerful tool for identifying protein-protein interactions. J Vis Exp.

Stynen, B., Tournu, H., Tavernier, J., and Van Dijck, P. (2012). Diversity in genetic in vivo methods for protein-protein interaction studies: from the yeast two-hybrid system to the mammalian split-luciferase system. Microbiol Mol Biol Rev **76,** 331-382.

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., Jensen, L.J., and Mering, C.V. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res **47,** D607-D613.

Takahashi, R., Ishimaru, Y., Shimo, H., Ogo, Y., Senoura, T., Nishizawa, N.K., and Nakanishi, H. (2012). The OsHMA2 transporter is involved in root-to-shoot translocation of Zn and Cd in rice. Plant Cell Environ **35,** 1948-1957.

Tehseen, M., Cairns, N., Sherson, S., and Cobbett, C.S. (2010). Metallochaperone-like genes in Arabidopsis thaliana. Metallomics **2,** 556-564.

Tusnady, G.E., and Simon, I. (2001). The HMMTOP transmembrane topology prediction server. Bioinformatics **17,** 849-850.

Vairo, F.P.E., Chwal, B.C., Perini, S., Ferreira, M.A.P., de Freitas Lopes, A.C., and Saute, J.A.M. (2019). A systematic review and evidence-based guideline for diagnosis and treatment of Menkes disease. Mol Genet Metab **126,** 6-13.

Van Assche, F., and Clijsters, H. (1990). A biological test system for the evaluation of the phytotoxicity of metal-contaminated soils. Environ Pollut **66,** 157-172.

Vonk, W.I., de Bie, P., Wichers, C.G., van den Berghe, P.V., van der Plaats, R., Berger, R., Wijmenga, C., Klomp, L.W., and van de Sluis, B. (2012). The copper-transporting capacity of ATP7A mutants associated with Menkes disease is ameliorated by COMMD1 as a result of improved protein expression. Cell Mol Life Sci **69,** 149-163.

Walker, J.M., Tsivkovskii, R., and Lutsenko, S. (2002). Metallochaperone Atox1 transfers copper to the NH2-terminal domain of the Wilson's disease protein and regulates its catalytic activity. J Biol Chem **277,** 27953-27959.

Wenli, Z., Chang, L., Yajun, Z., Jingguang, C., Haohua, H., and Guoyou, Y. (2020). Rice Heavy Metal P-type ATPase OsHMA6 Is Likely a Copper Efflux Protein. Rice Science **27,** 143-151.

Woeste, K.E., and Kieber, J.J. (2000). A strong loss-of-function mutation in RAN1 results in constitutive activation of the ethylene response pathway as well as a rosette-lethal phenotype. Plant Cell **12,** 443-455.

Yamaji, N., Chiba, Y., Mitani-Ueno, N., and Feng Ma, J. (2012). Functional characterization of a silicon transporter gene implicated in silicon distribution in barley. Plant Physiol **160,** 1491-1497.

Young, K.H. (1998). Yeast Two-Hybrid: So Many Interactions, (in) So Little Time ... Biology of reproduction **58**.

Yuan, D.S., Stearman, R., Dancis, A., Dunn, T., Beeler, T., and Klausner, R.D. (1995). The Menkes/Wilson disease gene homologue in yeast provides copper to a ceruloplasmin-like oxidase required for iron uptake. Proc Natl Acad Sci U S A **92,** 2632-2636.

Zhang, Y., Chen, K., Zhao, F.J., Sun, C., Jin, C., Shi, Y., Sun, Y., Li, Y., Yang, M., Jing, X., Luo, J., and Lian, X. (2018). OsATX1 Interacts with Heavy Metal P1B-Type ATPases and Affects Copper Transport and Distribution. Plant Physiol **178,** 329-344.

Zheng, Z.J., Ye, F., Zheng, L.S., Yang, K.F., Lai, G.Q., and Xu, L.W. (2012). Copper-catalyzed Huisgen and oxidative Huisgen coupling reactions controlled by polysiloxane-supported amines (AFPs) for the divergent synthesis of triazoles and bistriazoles. Chemistry **18,** 14094-14099.

CHAPTER IV

4. RNA-SEQ DEPENDENT TRANSCRIPTOME PROFILING OF *svd* MUTANT

4.1. Introduction

Transcriptional patterns of genes are tightly controlled by a combination of chromatin structure and transcription factors. Chromatin remodeling alters the availability of DNA regulatory sequences to transcription factors and affects their ability to activate or repress the expression of target genes (Komili and Silver, 2008).

Recently, RNA-seq and microarray have been become the preferred methods for the study of genome wide patterns of gene expression (Kogenaru et al., 2012). Rapid and inexpensive high throughput sequencing technologies are now used for the study of transcriptomes in wide range of organisms (Hanriot et al., 2008). Transcriptome profiling in plants has been reported in Arabidopsis (Schmid et al., 2005), rice (Jiao et al., 2009; Wang et al., 2010), sorghum (Buchanan et al., 2005; Salzman et al., 2005), switchgrass (Zhang et al., 2013), *Panicum halli* (Meyer et al., 2012), maize (Sekhon et al., 2011), Poaceae species (Davidson et al., 2012), *Glycine max* (Libault et al., 2010), and many others. This led to the availability of transcriptome databases to provide researchers with access to data from numerous experiments. These databases include; Phytozome (Goodstein et al., 2012), NCBI (Barrett et al., 2007), Gene Atlas (Druka et al., 2006), Gene Atlas for sorghum (Shakoor et al., 2014), GO (gene ontology) (Du et al., 2010), refGenes (Hruz et al., 2011) and Expression Data (Zimmermann et al., 2014). A range of tools are used in the analysis of RNA-seq data, including Integrated Genome Viewer (IGV). IGV can display the positions of single or multiple reads in the reference genome and read distributions between annotated exons, introns or intergenic regions, on adjustable scales. It can also show the read abundance of different regions to demonstrate their expression levels, on an adjustable scale. Moreover, IGV provides annotation information for both genes and splicing isoforms, provides other related annotation information and can be downloaded from remote servers and/or imported from local machines.

4.2. Objectives

The major objective of this transcriptome study was to carry out transcriptome profiling of *svd* mutant plants in comparison to wild type (BTx623) sorghum plants. This

will allow for the identification and evaluation of differentially expressed genes (DEG) in *svd* mutant plants.

4.3. Material and Methods

4.3.1. Plant material and growth condition

Plant materials used in this study were sorghum, [*Sorghum bicolor* L.) Moench], BTx623 seeds, and *svd* mutants. Plants were cultivated in a growth chamber under long-day conditions of 16/8 hours light /dark cycle at 25-28◦C temperature, 70-80% relative humidity, and 150μmol.m^2 light intensity. The MS-media contains 2.22gm/ 500ml of MS-Vitamin and 15 gm of sucrose added before autoclaving.

4.3.2. Sample Collection

Seedlings of BTx623, a standard inbred sorghum line, were grown in MS-media for 16 days before sample collection. While the wild type reached stage two making the collar of the fifth leaf visible in sixteen days *svd* was showing stress symptoms and sample collected at this stage (Vanderlip and Revees, 1972). Three biological replicates and matched controls were collected from each of the two different experiments for each compound analyzed. At each time point, 10 to 12 seedlings were harvested from each biological replicate for RNA extraction.

4.3.3. RNA extraction

Total RNA was isolated using TRIzol Reagent (Invitrogen) and after purification with ethanol. Briefly, samples measured 50 –100 mg frozen in liquid nitrogen were ground into powder using an automated tissue lyser, and the tissue were then homogenized using

1ml of TRIzol by brief vortex and by adding 5µl of DNase enzyme. The sample volume was approximately equal to 10% of the total volume. After five minutes of incubation, 200 µl of Chloroform was added and homogenized by inverting the tube five times. Cells were centrifuged at 12 g for 15 minutes at 4^{o}C. After carefully removing the upper aqueous solution, to this 500µl of isopropanol was added and incubated at room temperature for 10 minutes and another 10 minutes centrifuge was performed with the same process as previous. The supernatant was removed, and 1 ml of 70% ethanol treated with DEPC water was added and the pellet was washed by a brief vortex. Another centrifugation was performed at 7.5g for 5 minutes. The supernatant was removed, and the pellet washed, centrifuged and supernatant discarded to produce a clean pellet. After 10-15 minutes of drying the pellet at room temperature, 50µl of DEPC water was added to dissolve the RNA and the sample was incubated on ice at 4^{o}C for 24hrs to dissolve completely.

4.3.4. Total RNA sample Quality control (QC)

The total RNA sample quality control was performed using three methods to validate the purity, degradation, and integrity of the samples. First, Nanodrop measurement system was used to quantify and test the purity of RNA using OD260/ OD280 ratio. Second, potential degradation and contamination of RNA samples was evaluated using polyacrylamide gel electrophoresis. Finally, the integrity of RNA samples was checked with the Agilent 2100 Bioanalyzer before further processing.

4.3.5. Library Construction

For library construction, the RNA-seq adapter sequences (oligonucleotide sequences of adapters from TruSeqTM RNA and DNA Sample Prep Kits): RNA 5' adapter

(RA5), (Cat number # 15013205) were used. Briefly, after the QC procedures, mRNA from sorghum samples was enriched using oligo (dT) beads (TruSeq, Illumina Inc, San Diego CA). The mRNA was fragmented randomly in a fragmentation buffer, followed by cDNA synthesis using random hexamers and reverse transcriptase. After first-strand synthesis, a custom second-strand synthesis buffer (Illumina) was added with dNTPs, RNase H, and *Escherichia coli* DNA polymerase I to generate the second strand by nick-translation. The final cDNA library was complete after a round of purification, terminal repair, A-tailing, ligation of sequencing adapters, size selection, and PCR enrichment according to the protocol (Gilbert et al., 2014).

4.3.6. Library quality control (QC)

Library concentration was quantified using the Qubit 2.0 fluorimeter (Life Technologies), and then diluted to 1ng/µl before checking the insert size on an Agilent 2100 Bioanalyzer and quantifying by quantitative PCR (qPCR) (library activity >2nM).

4.3.7. Sequencing and Data analysis

The data filtering process was performed to remove reads containing adapters or reads of low quality in order to perform the downstream analyses using clean reads only. The filtering process included removing reads with adaptor contamination, discarding reads when undefined nucleotides constitute more than 10 percent of either read (N>10%), shedding reads of low base quality (base quality less than Q20) when these constituted more than 50 percent of the read. The RNA-seq adapter sequences (Oligonucleotide sequences of adapters from TruSeqTM RNA and DNA Sample Prep Kits), which includes RNA 5' side Adapter sequence (RA5), part number # 15013205:

5'AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3' and on the other side the RNA 3' Adapter (RA3), part number 15013207: 5'GATCGGAAGAGCACACGTCTGAACTCCAGTCAC (6-nucleotide index) ATCTCGTATGCCGTCTTCTGCTTG-3' were used in the filtering process. Assessment of raw data was performed and mapped to the reference genome was completed using HSAT2 version 2.1.0-beta and expression quantification was performed using HTSeq version 0.6.1. Differential gene expression analysis was performed using DEGSeq version 1.12.0 to calculate the log2 fold change. DESeq version 1.10.1, and edgeR version 3.0.8 were used to calculate the p-adj value, and intrinsic gene pathway analyses were also performed using GOSeq and hmmscan release 2.12. Protein-protein analysis was analyzed using the BLAST from string database, and KEGG enrichment analysis was done using KOBAS version 3.0.

4.3.8. Validation of RNA-seq data Using qPCR

Briefly, total RNA was isolated using TRIzol Reagent (Invitrogen) for cDNA synthesis. Reverse transcription (RT) was performed using RNA treated with DNase I (Invitrogen), an oligo (dT) primer and Superscript IV reverse transcriptase (Invitrogen) according to the manufacturer's instruction. After 8X dilution of the cDNA, the quality of the cDNA was checked using agarose gel-electrophoresis. Then expression pattern of selected genes was performed using a StepOne plus quantitative real time PCR (qRT-PCR) with specific forward and reverse primers. *Sorghum bicolor* L. actin primer was used as a standard.

4.4. Results

4.4.1. Quality control from RNA to the RNA seq data showed the reproducibility of the experiment

RNA-seq requires quality control at different stages starting from RNA. For DGE (Differential Gene Expression) libraries, a large variation of sequencing error in the first 6-7 bases was allowed due to the use of random primers in library construction (Finotello and Di Camillo, 2015). GC content is one of the aspects of studying the technical variance of the RNA-seq study. Additionally, GC content, shows the replicability of the sequence data and the percent of lane effect designated as "N" showed only few genes contained poly "N", which indicated a normal distribution and less bias. Hypothetically, G should equal to C, and A should equal to T throughout the whole sequencing process. Based on this evaluation, the GC content was ~60% in the cases of wild type (left) and *svd* (right) (Fig.23).

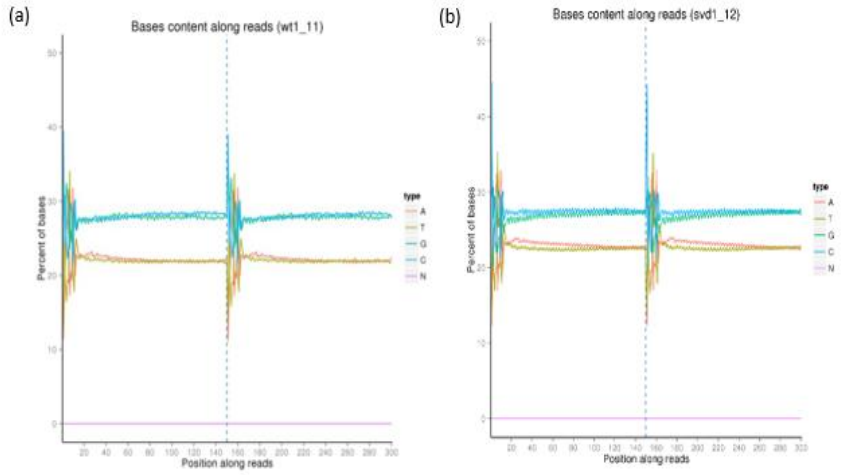

Fig. 23) GC distribution for wild type and svd showed the sequence data is replicable

GC content of RNA-seq shows normal distribution with the x-axis shows each base position within a read and the y-axis shows the percentage of each base, with each base represented by a different color.

4.4.2. Data filtering and classification of Raw reads

In the wild type, a total of 18,812,243 were clean reads and in the *svd* mutant a total of 29462865 (98.32%) were clean reads. Additionally, the number of reads containing N (reads that uncertain nucleotides constituted more than 10 percent of the read) in wild type were 93,794 (0.47%) and in the *svd* mutant 9,467 (0.03%). Low quality reads in wild type were more 414458 (2.06% of total) compared to 116233 (0.39% of total) in the mutant. The low quality reads in the wild type was associated to more ribosomal RNAs in the wild type (Fig. 24). However, the difference was minimal and no impact on the analysis. Adapter related reads were 753109 (3.75%) in wild type and 381894(1.27%) in the mutant. Despite minor discrepancies between the wild type and the *svd* mutant, the sequence clean reads were greater than 93% (Fig. 24).

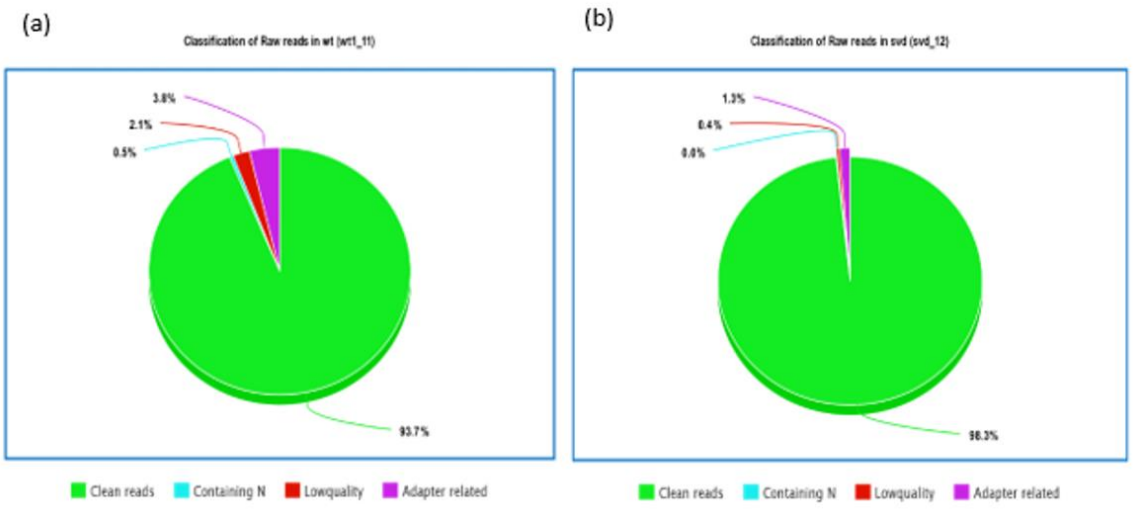

Fig. 24) Raw Reads Components and results in Venn diagram

a) classification of row reads in WT, and b) classification of raw reads in *svd*

Table 2: Overall data production summary in raw reads, clean reads, and error rates, Q20, Q30, and GC content

Sample name	Raw reads	Clean reads	Raw Bases	Clean Bases	Error rate	Q20 (%)	Q30 (%)	GC content (%)
wt1_11	40147028	37624486	6G	5.6G	0.03	97.82	94.47	55.86
wt2_11	62090008	59680992	9.3G	9G	0.02	98.28	95.26	55.66
wt3_11	40556334	38685556	6.1G	5.8G	0.02	98.20	95.14	55.81
*svd*1_12	71845842	69514712	10.8G	10.4G	0.02	98.02	94.55	54.38
*svd*2_12	54620986	52911682	8.2G	7.9G	0.02	98.33	95.31	56.50
*svd*3_12	59940918	58925730	9G	8.8G	0.02	98.25	95.02	56.10

Detail statistics of sequencing data:

(1) Raw Reads: the original sequencing reads counts

(2) Clean Reads: number of reads after filtering

(3) Raw Bases: raw reads number multiply read length, saved in G unit

(4) Clean Bases: clean reads number multiply read length, saved in G unit

(5) Error Rate: average sequencing error rate, which is calculated by Qphred=-10log10 (e)

(6) Q20: percentages of bases whose correct base recognition rates are greater than 99% in total bases

(7) Q30: percentages of bases whose correct base recognition rates are greater than 99.9% in total bases

(8) GC content: percentages of G and C in total bases

The Venn diagrams showing raw reads are shown as percentage of total raw reads, Adapter related (reads that had adapter contamination), containing N (reads that uncertain nucleotides constituted more than 10 percent of the read), Low quality (reads that low quality nucleotides constituted more than 50 percent of the read), Clean reads (reads that passed quality control) (Fig. 24; Table 2). The overall raw data produced ranged from 40,147,028 reads to 62,090,008 reads and from 54,620,986 to 71,845,842 reads in the case of the wild type and *svd,* respectively. The overall error rate calculated by Qphred=-10log10 (e) was ~0.02 and the Q30 value (percentages of bases whose correct base recognition rates are greater than 99.9% in total bases) ranges from 94.47% to 95.26% in wild type and from 94.55 to 95.31 in svd (Table 2).

4.4.3. Mapping to a reference genome

An important mapping quality parameter is the ratio of mapped reads, which is a comprehensive indicator of the total sequencing precision and of the existence of contaminating materials in the sample. Mapping to a reference genome was performed using HISAT2. The quantity of total mapped reads and its percentage of clean reads was calculated, including the quantity of multiple mapped reads and its percentage of clean reads, and the quantity of uniquely mapped reads and its percentage of clean reads (Table

2). The TMR (Total Mapped Reads or Fragments) were larger than 95% and MMR (Multiple Mapped Reads or Fragments) was less than 5% in all cases (Table 2). This showed that the sequencing quality was acceptable in all samples where the minimum requirement for TMR is usually 70% and the MMR should be no more than 10%. The mapped regions are classified based on the structural annotation of the reference genome. Exon-mapped reads should be abundant if the reference genome is well-annotated. Reads mapped to intergenic regions may be due to weak annotation in the sorghum reference genome. In the wild type, 94.8% of the mapped reads were exons, 2% were introns and 3.1 % of the mapped reads were in the intergenic regions. While in *svd*, 95.5% of the mapped reads were exons, 1.6% were introns and 2.9 % of the mapped reads were in the intergenic regions (Fig. 25).

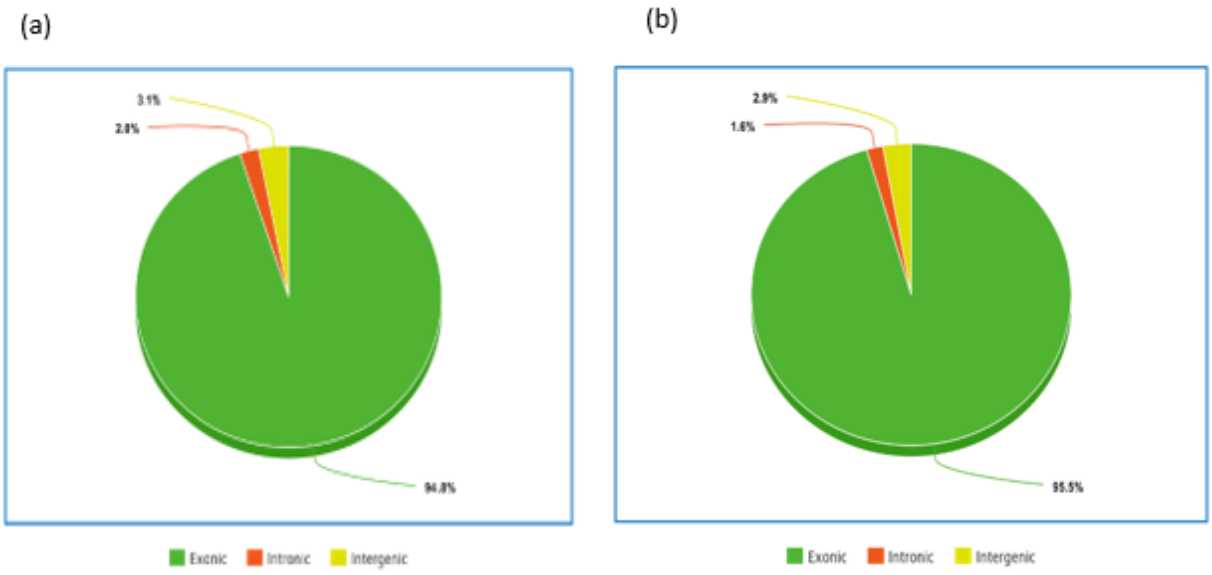

Fig. 25) Classification of Reads according to mapped region

Classification of Reads according to mapped region shows A) Percent of mapped reads to genome regions in wild type (wt1_11), B) Percent of mapped reads to genome regions in in the mutant (*svd*1_12).

4.4.4. Distribution of Mapped Reads in Chromosomes

To obtain an overview of the distribution of mapped reads in chromosomes, the "window size" was set to 1K, the median number of reads mapped to the genome inside the window was calculated and transformed to the $\log_2$value (Fig. 26). In general, the longer the whole chromosome, the more the total number of mapped reads are located within the chromosome (Marquez et al., 2012).

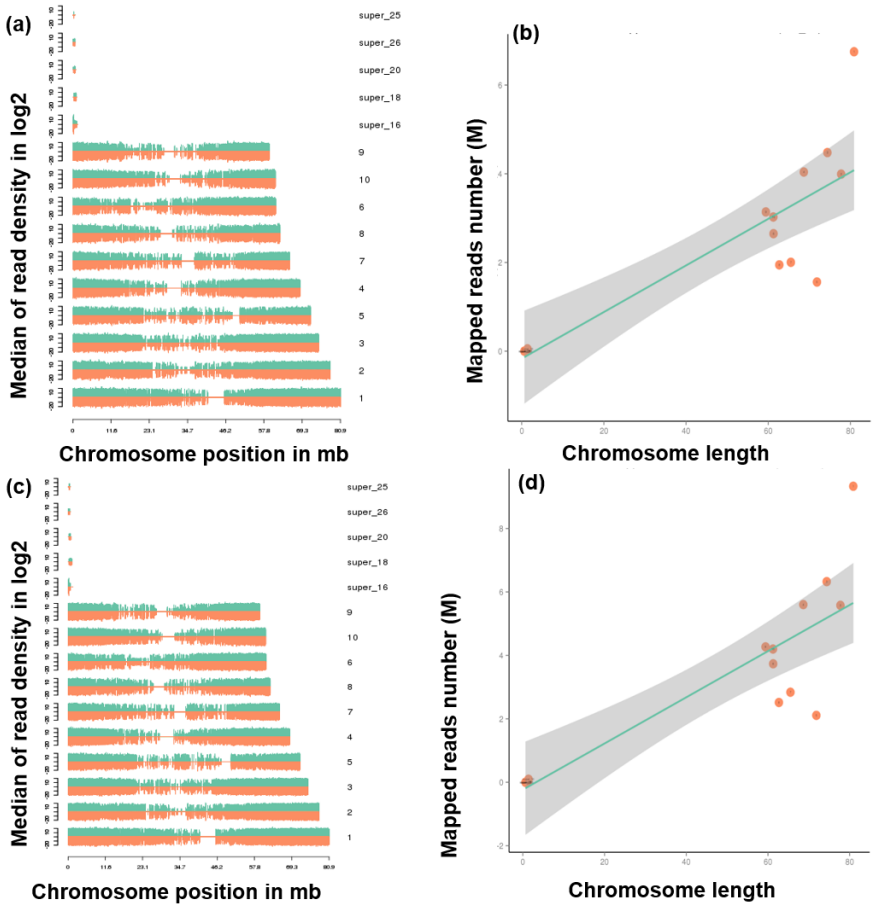

Fig. 26) Distribution Plot of Mapped Reads in Chromosomes

Distribution Plot of Mapped Reads in Chromosomes, a) read density in chromosomes for wild type, b) number of mapped reads in chromosomes for wild type c) read density in chromosomes for *svd*, d) number of mapped reads in chromosomes for *svd*. The X-axis shows the length of the chromosomes (in Mb), and the Y-axis indicates the log2 of the median of read density. Green and red indicate, respectively, the positive and negative strands. In the right panel, the X-axis shows the length of the chromosomes, and the Y-axis indicates the number of mapped reads in each chromosome. The grey region indicates the 95% confidence interval.

4.4.5. Visualization of Mapping Status of Reads

After files were exported in Binary Alignment Map (BAM) format, a standard file format that contains mapping results, from the QC processing steps and the corresponding reference genome and gene annotation file for sorghum was used. The Integrative Genomics Viewer (IGV) was used for visualizing (Fig. 27).

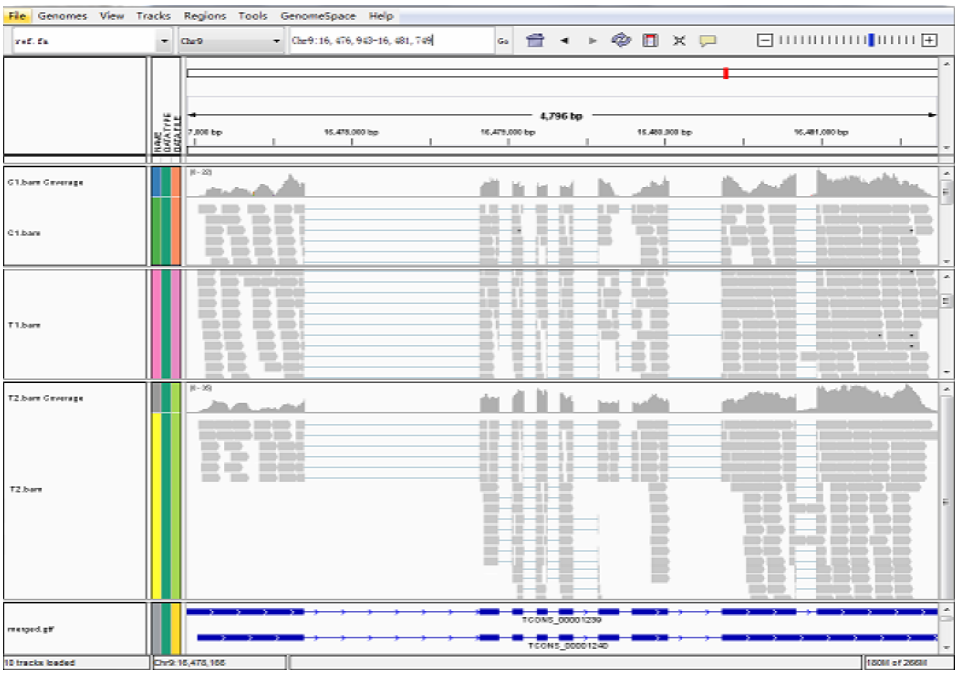

Fig.27) Integrated genome viewer visualizing the mapped reads

The clean filtered reads mapped to the reference genome ranges from 37,624,486 to 59,680,992 reads in the case of wild type and it ranges from 52,911,682 to 69,514,712

reads in *svd* (Table 3). From these clean reads, the totally mapped reads range from 35,833,551 (95.24%) to 57,934,349 (97.07%) reads in wild type and from 50,401,120 (95.26%) to 67,582,944 (97.22%) reads in the wild type (Table 3). Multiple mapped reads (MMR) in the wild type ranged from 1,493,218 (2.5%) to 1,423,899 (3.78%) reads in the case of *svd*, MMR ranges from 1,988,764 (2.86%) to 2,639,039 (4.99%) reads, which is higher compared to the wild type but still below the required standard that is 10%. Uniquely mapped reads range from 91.46 to 94.36 percent of the total mapped reads, and a uniform score in other characteristics of the data like reads mapped to the forward strand, reads mapped to the reverse strand, spliced reads and non-spliced reads mapped were proportional for both the control and samples.

Table 3: Summary of mapping to reference genome; Sorghum bicolor V3.1.1 GFF3 sequence is used for mapping

Sample name	WT1-11	WT2-11	WT3-11	*svd*1_11	*svd*2_11	*svd*3_11
Total reads	37624486	59680992	38685556	69514712	52911682	58925730
Total mapped	35833551 (95.24%)	57934349 (97.07%)	37461494 (96.84%)	67582944 (97.22%)	50401120 (95.26%)	56539407 (95.95%)
Multiple mapped	1423899 (3.78%)	1493218 (2.5%)	1131243 (2.92%)	1988764 (2.86%)	2639039 (4.99%)	2728400 (4.63%)
Uniquely mapped	34409652 (91.46%)	56441131 (94.57%)	36330251 (93.91%)	65594180 (94.36%)	47762081 (90.27%)	53811007 (91.32%)
Reads map to '+'	17199777 (45.71%)	28208496 (47.27%)	18160188 (46.94%)	32787288 (47.17%)	23871538 (45.12%)	26901977 (45.65%)
Reads map to '-'	17209875 (45.74%)	28232635 (47.31%)	18170063 (46.97%)	32806892 (47.19%)	23890543 (45.15%)	26909030 (45.67%)

Non-splice reads	22848547 (60.73%)	37848509 (63.42%)	24440947 (63.18%)	43195242 (62.14%)	31755494 (60.02%)	35411557 (60.1%)
Splice reads	11561105 (30.73%)	18592622 (31.15%)	11889304 (30.73%)	22398938 (32.22%)	16006587 (30.25%)	18399450 (31.22%)

Descriptions of Rows

(1) Total reads: Total number of filtered reads (Clean data).

(2) Total mapped: Total number of reads that can be mapped to the reference genome. In general, this number should be larger than 70% when there is no contamination and the correct reference genome is chosen.

(3) Multiple mapped: Number of reads that can be mapped to multiple sites in the reference genome. This number is usually less than 10% of the total.

(4) Uniquely mapped: Number of reads that can be uniquely mapped to the reference genome.

(5) Reads map to '+', Reads map to '-': Number of reads that map to the positive strand (+) or the minus strand (-).

(6) Splice reads: Splice reads can be segmented and mapped to two exons (also named junction reads), whereas non-splice reads can be mapped entirely to a single exon. The ratio of splice reads depends on the insert size used in the RNA-seq experiments

4.4.6. Gene Expression Quantification

HTSeq software was used to analyze the gene expression levels, using the union mode. The result files present the number of genes with different expression levels and the

expression level of single gene. The FPKM (Fragments per Kilo Bases in million reads) was used to determine the gene expression quantification. Genes with FPKM value 0~1 range from 15,885 (44.66%) to 17,511 (49.23%) in the wild type and from 15,253 (42.89%) to 16,029 (45.07%) in *svd*. Consequently, genes with FPKM value 1~3 ranges from 3,475 (9.77%) to 3,575 (10.05%) in the wild type and from 3,631 (10.21%) to 3,713 (10.44%) in *svd*. Moreover, genes with FPKM value 3~15 ranges from 7,584 (21.32%) to 8,625 (24.25%) in the wild type and from 8,720 (24.52%) to 8,876 (24.96%) in *svd* (Table 4). There are also genes in the FPKM category of 15~60 and >60. In general, an FPKM value of 0.1 or 1 is set as the threshold for determining whether the gene is expressed or not.

Table 4: The number of genes with different expression levels

FPKM interval	WT1-11	WT2-11	WT3-11	*svd*1_11	*svd*2_11	*svd*3_11
0~1	17511(49.23%)	16461(46.28%)	15885(44.66%)	15253(42.89%)	16029(45.07%)	15745(44.27%)
1~3	3475(9.77%)	3544(9.96%)	3575(10.05%)	3705(10.42%)	3713(10.44%)	3631(10.21%)
3~15	7584(21.32%)	8414(23.66%)	8625(24.25%)	8859(24.91%)	8720(24.52%)	8876(24.96%)
15~60	5334(15.00%)	5471(15.38%)	5794(16.29%)	5966(16.77%)	5444(15.31%)	5586(15.71%)
>60	1663(4.68%)	1677(4.72%)	1688(4.75%)	1784(5.02%)	1661(4.67%)	1729(4.86%)

4.4.7. Comparison of Gene Expression levels

To compare gene expression levels under different conditions, an FPKM distribution diagram and violin Plot were used. For biological replicates, the mean values were used as the final FPKM. Each violin had five statistical magnitudes (max value, upper quartile, median, lower quartile, and min value). The violin width showed the gene density (Fig.28). Both the replicates in each sample showed a similar FPKM density distribution

pattern. The gene expression levels for sample genes in WT and *svd* mutants showed genes with different metabolic and molecular functions are differentially regulated (Table 5).

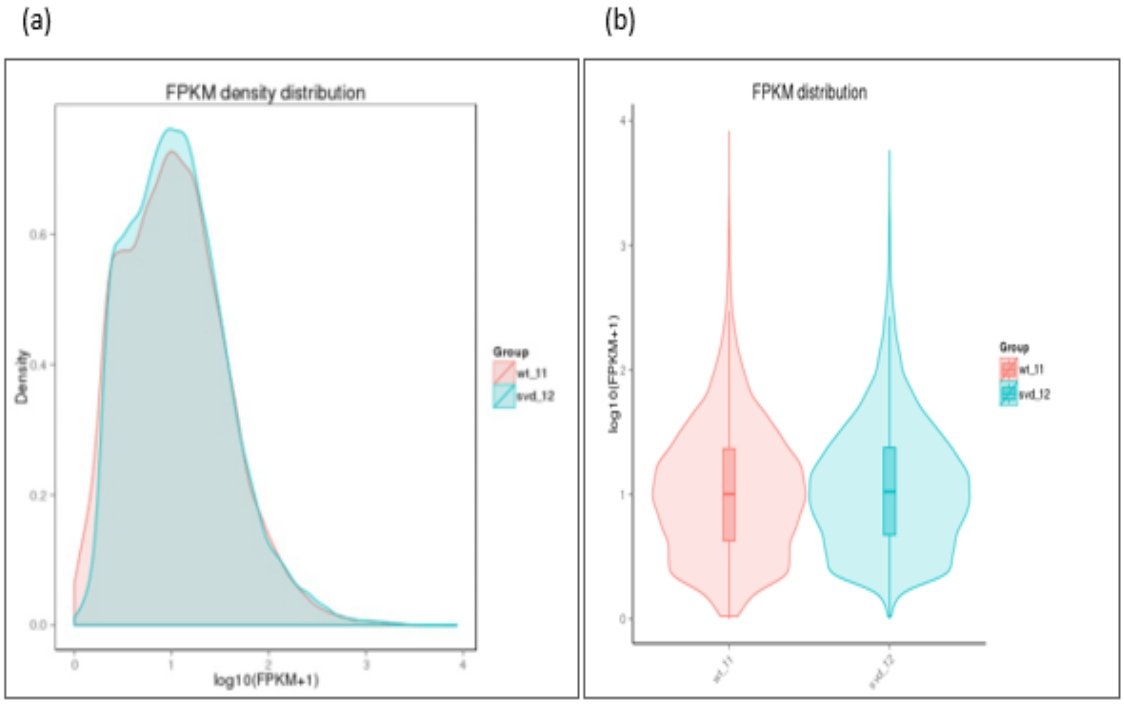

Fig.28) FPKM distribution and FPKM violin Plot

FPKM distribution (a), the x-axis shows the log10 (FPKM+1) and the y-axis shows gene density. (b) FPKM violin Plot, the x-axis shows the sample names and the y-axis shows the log10 (FPKM+1).

Table 5: Gene expression levels for sample genes in WT and *svd* mutants in all the three replicates

Gene ID	WT1-11	WT2-11	WT3-11	*svd*1_11	*svd*2_11	*svd*3_11
SORBI_3001G053900	46.64331	22.01627	31.166636	24.02481	22.33601	23.24590
SORBI_3009G163700	5.36633	12.46120	14.73572	14.73559	12.87902	13.92371
SORBI_3004G244300	4.93841	6.11360	6.82915	7.92656	3.60121	4.21626
SORBI_3002G221000	0.124885	0	0.0234168	0.156026	0	0

4.4.8. The co-expression venn diagram of gene expression differences

The co-expression_venn diagram presents the number of genes that are uniquely expressed within each sample, with the overlapping regions showing the number of genes that are expressed in two or more samples. While a total of 484 genes were expressed specifically to the wild type and 1233 genes were expressed specific to the *svd* mutant, most of the DEG genes (18865 genes) were coExpressed together (Fig.29).

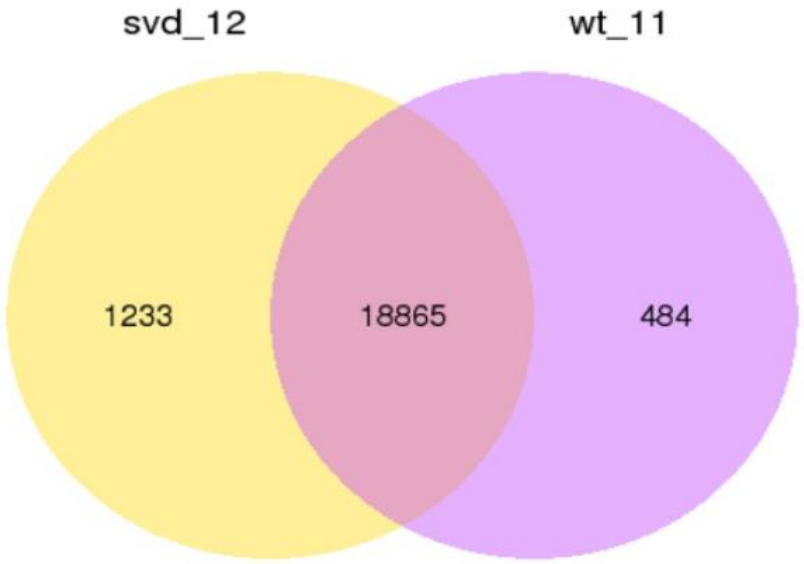

Fig. 29) Total number of expressed genes and co-expressed genes in WT and svd mutant

The sum of the numbers in each circle is the total number of genes expressed within a sample, and the overlap represents the genes expressed in common between samples.

4.4.9. Differential gene expression analysis

Volcano plots were used to infer the overall distribution of differentially expressed genes. A total 147 genes were downregulated as shown in green (Fig. 30; left) where a total of 804 genes were upregulated as shown in red (Fig. 30; right). For experiments without biological replicates, the threshold was normally set as: log2 (Fold Change) > 1 and q-value < 0.005. However, in this case as the DESeq already eliminates the biological variation and the threshold was normally set as: padj < 0.05.

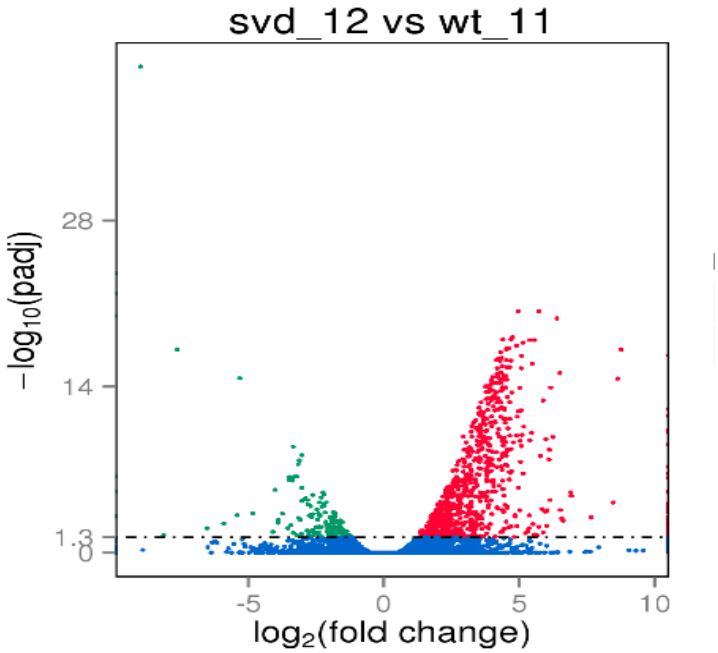

Fig. 30) Volcano plot for log fold change in gene expression

The x-axis shows the fold change in gene expression between different samples, and the y-axis shows the statistical significance of the differences. Significantly up and down regulated genes are highlighted in red and green, respectively. Genes did not express differently between treatment group and control group are in blue.

Important genes including SORBI_3001G261480 (Cytidine deaminase enzyme and function pyrimidine ribonucleosides salvage II), SORBI_3003G078000 (LRR like proteins), and SORBI_3008G079100 were downregulated genes with log2 Fold change of

~8.9, ~9.6, ~11.7, respectively. And SORBI_3004G302700 (Axial regulator YABBY-1 related protein) was one of the most upregulated genes with a value of 5.728 (Fig.30, Table 6).

Table 6: Differentially expressed genes for sample genes

Gene ID	*svd*-12	WT-11	log2 Fold change	p-Val	p-adjusted
SORBI_3001G261480	2.51124	1261.7595	-8.9728	3.5774e-46	1.0477e-41
SORBI_3003G078000	0	201.8133	9.62314	1.9496e-28	2.8548e-24
SORBI_3008G079100	0	176.51194	11.7456	1.3767e-26	1.344e-22
SORBI_3004G302700	502.9049	9.48800	5.728	6.1277e-25	4.4864e-21

Description of rows

(1) Sample1: The readcount values of sample1 after normalization

(2) Sample2: The readcount values of sample2 after normalization

(3) Log2FoldChange: log2 (Sample1/Sample2), (4) p-value (pval): The p-value.

(5) q-value (p-adjusted): the p-value after normalization. The smaller the q-value is, the more significant is the difference

4.4.10. Cluster analysis of gene expression differences

Cluster analysis is used to find genes with comparable expression under various experimental situations. By clustering genes with similar expression patterns, it is possible to infer unknown functions of previously characterized genes or functions of unknown genes. In hierarchical clustering, different groups of genes were represented by different

colors where the strength of these colors matches their expression level, and genes within each cluster may have similar functions or take part in similar biological process. Other clustering methods include the H-cluster, K-means, and SOM cluster methods using log2 (ratios) (Duo et al., 2018; Moussa and Mandoiu, 2018). Genes within the same cluster exhibit the same trends in expression levels under different conditions. The color ranges from red to blue represents the $\log_{10}$(FPKM+1) value from large to small. Lower panel：

log2(ratios) line chart. Each grey line in a subline chart represents the relative expression value of a gene cluster under different experimental conditions, and the blue line represents the mean value (Fig.31a). Some representative DEG genes from the heatmap shows different groups are affected suggesting these *SVD* gene play key roles in different metabolic processes.

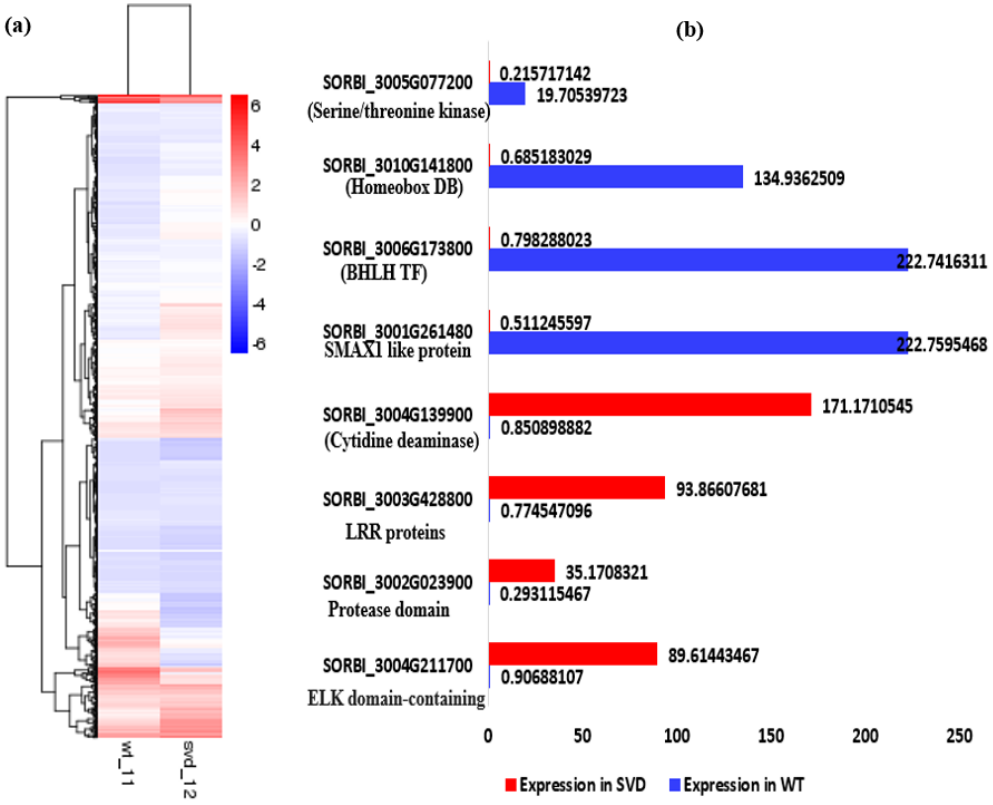

Fig. 31) Heat map for the overall results of FPKM cluster analysis

The overall results of FPKM cluster analysis, clustered using the log10 (FPKM+1) value. (a)Red denotes genes with high expression levels, and blue denotes genes with low expression levels. The color ranges from red to blue represents the log10 (FPKM+1) value from large to small. The x-axis shows the experimental condition and the y-axis shows the relative expression value, (b) Some of the most up and down regulated genes in the heatmap.

4.4.11. GO enrichment analysis of DEGs

Gene ontology study using GOSeq showed different biological, cellular and molecular functions were enriched. Cell cycle and its process, chromosome organization, DNA replication, DNA metabolic processes, microtubule-based movement, movement of cell and subcellular organelles, DNA dependent DNA replication, microtubule-based process, mitotic cell cycle, DNA conformation change from the upregulated enrichments involved in biological processes. Cellular components like chromosomes, chromosomal segments, protein-DNA complexes, DNA packaging complexes, nucleosome, nuclear chromosome part, nucleus, non-membrane bound organelles, intercellular nuclear bound organelles and chromatins were involved. Hence, molecular functions like macromolecular complex bindings, microtubule binding, tubulin binding, microtubule activity, motor activity and DNA binding activities were all affected (Fig. 32a). On the other hand, in downregulated genes also there are different gene enrichments including biological processes, cellular processes and molecular functions. These includes, biological processes like response to wounding , phosphorylation, regulation jasmonic acid methylation, electron transport coupled processes, reductive pentose phosphate cycle, photosynthesis dark reaction, transmembrane transport, jasmonic acid signaling, regulation of signal transduction, regulation of signaling, and regulation of cell communication were

all affected. As a result, molecular functions including phosphatase activity, phosphoric ester hydrolase activity, transcription corepressor activity, inositol 3-alpha-galactosyltransferase, uniporter activity, protein serine/threonine phosphatase activity, ribulose bisphosphate carboxylase activity, rRNA binding, transition metal ion transport, riboflavin kinase activity, ion transmembrane transporter activities were all affected (Fig. 32b).

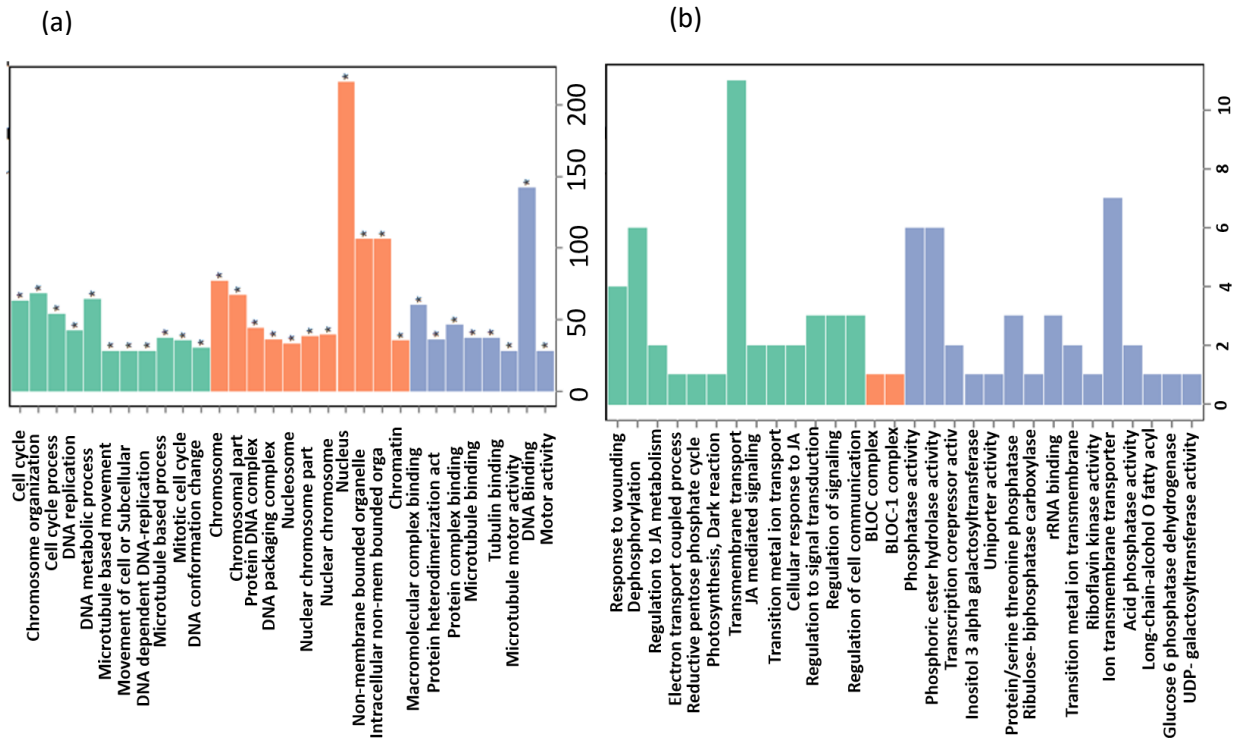

Fig. 32) GO enrichment analysis

a) Upregulated differentially expressed genes in svd, b) Downregulated differentially expressed genes in *svd.*

There are two graphs in each group, the x- axis is GO terms enriched and the y-axis is the number of differential expression genes. Different colors are used to distinct biological process, cellular component, and molecular function, that the enriched GO terms are marked by "*". The GO terms in the figure which are drawn in subsets of graph based on biological process, cellular component, molecular function, and differential expression genes.

4.4.12. Confirmation of gene expression

A total 38 DEG genes from RNA-seq experiment were selected based on the fold change enrichment of genes and based on their expression pattern and nine genes were further selected based on their co-expression patterns with SVD and to validate the RNA-seq experiment a quantitative real time expression analysis.

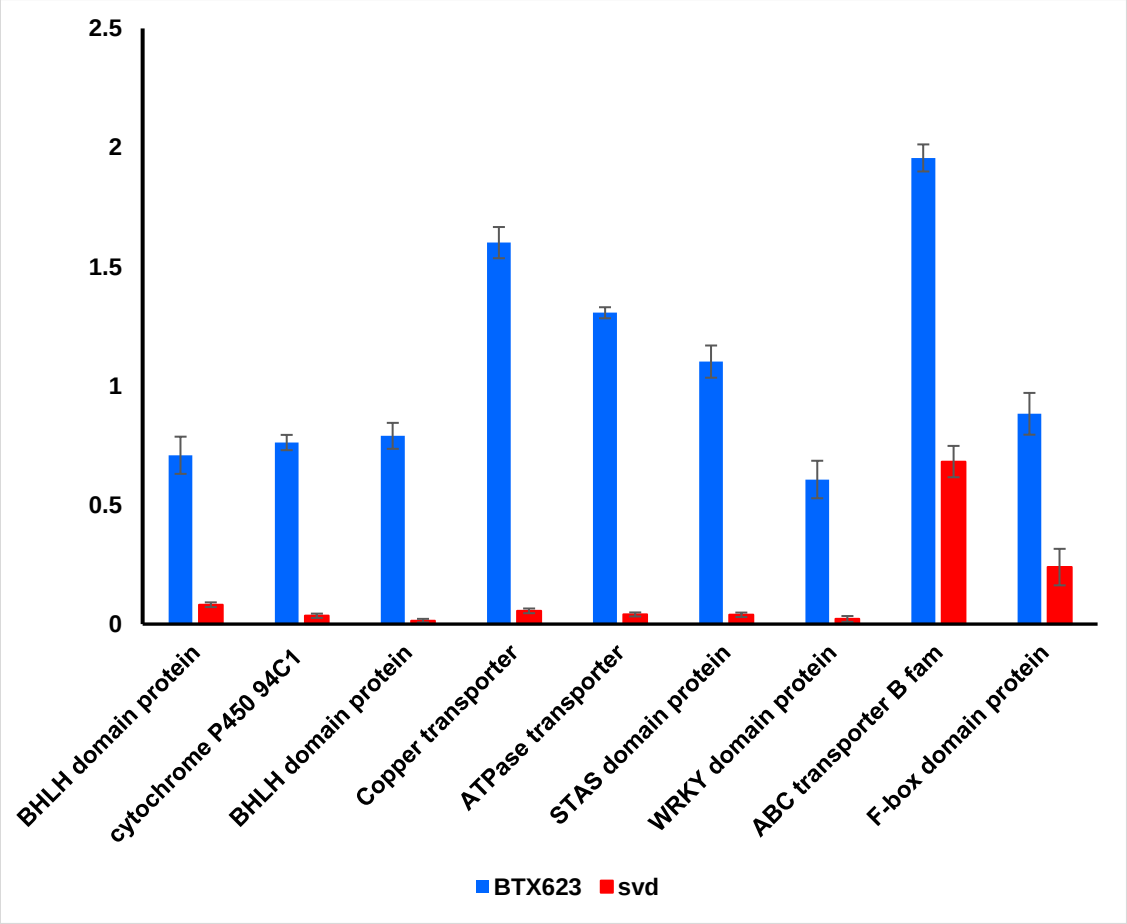

Fig. 33) qPCR analysis of downregulated genes

In general, several important genes in cell regulation genes are also significantly down regulated. These genes include; *SORBI_3002G224500* (BHLH domain protein87), *SORBI_3005G037000* (cytochrome P450 94C1), *SORBI_3007G155300* (BHLH domain

protein), *SORBI_3002G215800* (Copper transporter), *SORBI_006G173800* (ATPase transporter), *SORBI_3010G033800* (STAS domain protein), *SORBI_3003G037500* (WRKY domain protein), *SORBI_3007G042600* ABC transporter B fam), and *SORBI_3006G126500* (F-box domain protein) (Fig. 33). On the other hand, there are also many upregulated genes, and these include; *SORBI_3001G403200 (*Protein Kinases domain containing*), SORBI_3008G114500 (Histone H2A), SORBI_3005G000500* (Kinesin motor domain containing protein), *SORBI_3002G138000* (Histone containing protein H4), *SORBI_3004G317200* (LRR receptor kinase protein), *SORBI_3006G076400* (Histone 3 protein), *SORBI_3001G177900* (Kinesin 12c protein), *SORBI_3002G1142200* (serine/threonine kinase), and *SORBI_3001G001700* (DOF zinc figure nuclease protein) are some of the most upregulated genes (Fig.34). This suggests that SVD is involved in many cellular processes.

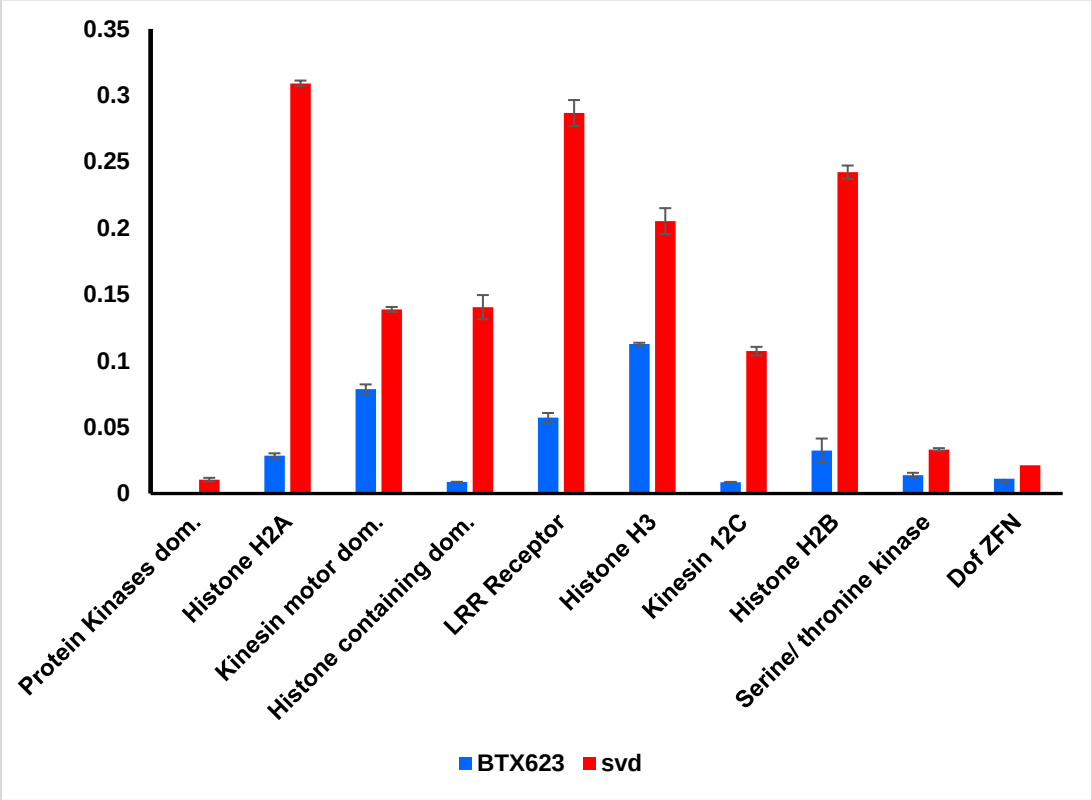

Fig. 34) qPCR analysis of upregulated genes involved in many cellular molecular and physiological processes

4.4.13. KEGG enrichment lists and enrichment pathways

Pathway enrichment analysis identifies significantly enriched metabolic pathways or signal transduction pathways associated with differentially expressed genes compared with the whole genome background. Additionally, the genes were mapped to the KEGG pathway. Among the mapped genes, the upregulated genes in *svd* mutant were mainly enriched in three pathways: Kinases, kinesins and histones. The enriched genes had lower Q values, indicating significant enrichment. However, the KEGG enrichment of downregulated genes in the *svd* mutant were found in 2 groups: plant hormone signal transduction and ABC transporters, but in the case of ABC transporters had higher Q values

indicating less significant enrichment (Fig. 35). The scatter diagram is a graphical display way of KEGG enrichment analysis results. In this plot, enrichment degree of KEGG can be measured through the Rich factor, or Q value and genes counts enriched to this pathway. Rich factor is the ratio of DEGs counts in this pathway to total annotated genes counts. The more the rich factor is the higher the degree of enrichment. Q value is the adjusted p-value after multiple hypothesis testing, and its ranges is between 0 and 1. The more the Q value is close to zero, the more significant is the enrichment. The top 20 most significant enriched pathways are shown in the KEGG scatter plot (Fig. 35). Several DEG's were enriched with different pathways, such as purine metabolism, pyrimidine metabolism, nucleotide excision repair, mismatch repair, homologous recombination, Glutathione metabolism, and plant hormone signal transduction with different levels of enrichment, however, DEG's for DNA replication genes are significantly enriched (Fig.35).

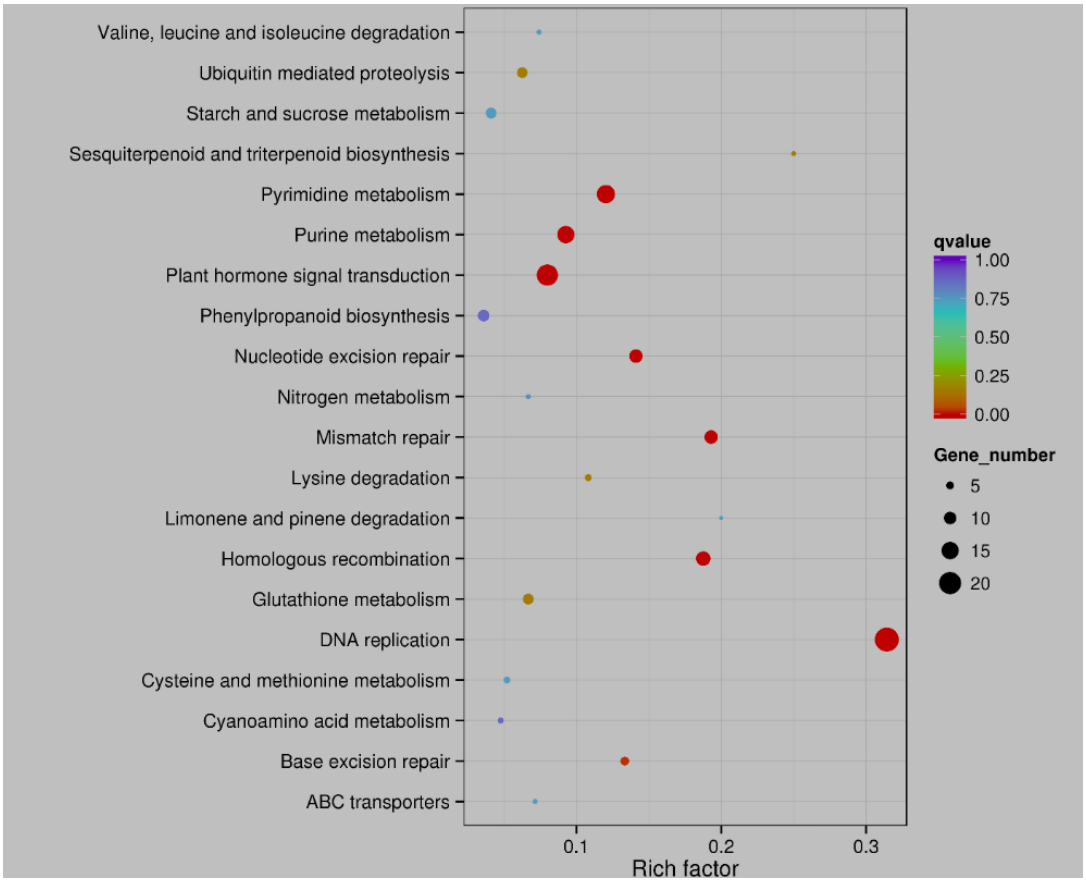

Fig. 35) Statistics of Pathway enrichment

KEGG scatter plot of 20 extremely enriched DEGs. The x-axis represents Rich factor; the y-axis represents name of pathway; the plot size represents the expression number of DEGs in pathways; the color of plot represents corresponding Q value

4.5. Discussion

The development of easy and ultra-fast next generation sequencing approaches made access to large genome data of different plants or other organisms easily accessible. However, the biggest bottleneck to put these data into application is the sophisticated data processing approach. Appropriate sampling and replications with a defined experimental design, quality control of the total RNA or DNA extracted from the samples and quality

control after library preparation were the three key stages for the successful completion of the RNA-seq project. Nonetheless, other sequencing quality parameters can give crucial knowledge about the precision of each phase in the procedures, including library preparation, base calling, read alignment, and variant calling are some of the most important steps that can avoid experimental bias and inappropriate conclusions. The analysis stages QC(quality control) is the key but often a neglected step (Sheng et al., 2017). There are different quality control methods of RNA-seq data and the *phred* base calling system is one of the methods used (Ewing et al., 1998). In this analysis, all steps of quality control measures were performed, *phred* value of Q30 values of 93 and more, which is the standard for next step analysis, mapping to the reference (Table 2).

After the QC control different alignment and mapping algorisms in whole genome sequencing projects are key tools in identifying genetic variants and differentially expressed genes. Tophat2 and cufflinks, HISAT2 and HISAT are some of these packages. Algorithm for mapping sequences the HISAT2 package was selected to map the filtered sequenced reads to the reference genome. This is because the HISAT2 supports genomes of any size, including those larger than 4 billion bases and most of the parameters are set to default. Spliced reads of RNA-seq data can be effectively aligned using HISAT2, which is the fastest system currently available, with equal or better accuracy than any other method. In this case, *Sorghum bicolor* has a size of ~730Mb (relatively smaller) and the reference sequence Sbicolor_454_V3.1.1.gene.gff3 was used for mapping.

RNA-seq was performed for *svd* and BTx623 as a wild type control using next generation sequencing system to comprehend the transcriptome changes in the *svd* mutant verses the WT. The observations suggest that many forms of membrane transport,

movement along the microtubule filaments, cell proliferation, stress signal transduction, and chromatin conformation may be heavily affected in the *svd* mutant.

The interactions of multiple genes may be involved in certain biological functions. KEGG (Kyoto Encyclopedia of Genes and Genomes) is a collection of manually curated databases dealing with genomes, and their functional information including biological pathways, diseases, drugs, and chemical substances. KEGG is utilized for bioinformatics research and education, including data analysis in genomics, metagenomics, metabolomics, and other omics studies. KEGG (http://www.genome.jp/kegg/) database, a popular tool for linking genomes to life and the environment, helps to understand a high-level function cellular processes and organismal deeds (Wu et al., 2006; Kanehisa et al., 2008). KEGG orthologue Based Annotation System (KOBAS) server was used to blast nucleotide or amino acid sequences or by sequence identifiers in popular databases and can annotate the input with KO terms and KEGG pathways by BLAST sequence similarity or directly ID mapping to genes with known annotations (Mao et al., 2005; Wu et al., 2006).

On the other hand, gene ontology (GO), http://www.geneontology.org/ is a major bioinformatics initiative to unify the presentation of gene and gene product attributes across all species. GO enrichment analysis is used by GO-seq, which is based on Wallenius non-central hyper-geometric distribution (Oshlack et al., 2010; Young et al., 2010a). Its characteristics are the probability of drawing an individual from a certain category is different from that of drawing it from outsides of the category, and this difference is obtained from estimating of the preference of gene length. The gene ontology (GO) terms are widely used to describe cellular component, molecular function, and biological process of genes. A GO enrichment bar chart is used to illustrate the differentially expressed genes

enriched GO terms and the counts of genes for each GO terms. By using a program GOSeq to study the gene ontology reduce complexity and highlight biological processes in genome-wide expression studies (Young et al., 2010b) and the results showed that different enrich ontologies for different biological processes including cell cycle process and DNA replication, microtubule movement, DNA conformation and DNA metabolic processes were upregulated . Consequently, many transporters, transmembrane proteins and different metabolic processes were down regulated. These results suggest that the *svd* mutant undergoes many cellular, molecular, and biological changes and develops stress because of loss-of-function of the *SVD* gene.

4.6. References

Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Rudnev, D., Evangelista, C., Kim, I.F., Soboleva, A., Tomashevsky, M., and Edgar, R. (2007). NCBI GEO: mining tens of millions of expression profiles--database and tools update. Nucleic Acids Res **35,** D760-765.

Buchanan, C.D., Lim, S., Salzman, R.A., Kagiampakis, I., Morishige, D.T., Weers, B.D., Klein, R.R., Pratt, L.H., Cordonnier-Pratt, M.M., Klein, P.E., and Mullet, J.E. (2005). Sorghum bicolor's transcriptome response to dehydration, high salinity and ABA. Plant Mol Biol **58,** 699-720.

Davidson, R.M., Gowda, M., Moghe, G., Lin, H., Vaillancourt, B., Shiu, S.H., Jiang, N., and Robin Buell, C. (2012). Comparative transcriptomics of three Poaceae species reveals patterns of gene expression evolution. Plant J **71,** 492-502.

Druka, A., Muehlbauer, G., Druka, I., Caldo, R., Baumann, U., Rostoks, N., Schreiber, A., Wise, R., Close, T., Kleinhofs, A., Graner, A., Schulman, A., Langridge, P., Sato, K., Hayes, P., McNicol, J., Marshall, D., and Waugh, R. (2006). An atlas of gene expression from seed to seed through barley development. Funct Integr Genomics **6,** 202-211.

Du, Z., Zhou, X., Ling, Y., Zhang, Z., and Su, Z. (2010). agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res **38,** W64-70.

Duo, A., Robinson, M.D., and Soneson, C. (2018). A systematic performance evaluation of clustering methods for single-cell RNA-seq data. F1000Res **7,** 1141.

Ewing, B., Hillier, L., Wendl, M.C., and Green, P. (1998). Base-Calling of Automated Sequencer Traces Using Phred. I. Accuracy Assessment. Genome Res **175**.

Finotello, F., and Di Camillo, B. (2015). Measuring differential gene expression with RNA-seq: challenges and strategies for data analysis. Brief Funct Genomics **14,** 130-142.

Gilbert, K.B., Fahlgren, N., Kasschau, K.D., Chapman, E.J., Carrington, J.C., and Carbonell, A. (2014). Preparation of Multiplexed Small RNA Libraries From Plants. Bio Protoc **4**.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., and Rokhsar, D.S. (2012). Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res **40,** D1178-1186.

Hanriot, L., Keime, C., Gay, N., Faure, C., Dossat, C., Wincker, P., Scote-Blachon, C., Peyron, C., and Gandrillon, O. (2008). A combination of LongSAGE with Solexa sequencing is well suited to explore the depth and the complexity of transcriptome. BMC Genomics **9,** 418.

Hansen, K.D., Irizarry, R.A., and Wu, Z. (2012). Removing technical variability in RNA-seq data using conditional quantile normalization. Biostatistics **13,** 204-216.

Hruz, T., Wyss, W., Docquier, M., Pfaffl, M.W., Masanetz, S., Borghi, L., Verbrugghe, P., Kalaydjieva, L., Bleuler, S., Laule, O., Descombes, P., Gruissem, W., and Zimmermann, P. (2011). RefGenes: identification of reliable and condition specific reference genes for RT-qPCR data normalization. BMC Genomics **12**.

Jiao, Y., Tausta, S.L., Gandotra, N., Sun, N., Liu, T., Clay, N.K., Ceserani, T., Chen, M., Ma, L., Holford, M., Zhang, H.Y., Zhao, H., Deng, X.W., and Nelson, T.

(2009). A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies. Nat Genet **41,** 258-263.

Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T., and Yamanishi, Y. (2008). KEGG for linking genomes to life and the environment. Nucleic Acids Res **36,** D480-484.

Kogenaru, S., Yan, Q., Guo, Y., and Wang, N. (2012). RNA-seq and microarray complement each other in transcriptome profiling. BMC Genomics **12**.

Komili, S., and Silver, P.A. (2008). Coupling and coordination in gene expression processes: a systems biology view. Nat Rev Genet **9,** 38-48.

Libault, M., Farmer, A., Joshi, T., Takahashi, K., Langley, R.J., Franklin, L.D., He, J., Xu, D., May, G., and Stacey, G. (2010). An integrated transcriptome atlas of the crop model Glycine max, and its use in comparative analyses in plants. Plant J **63,** 86-99.

Mao, X., Cai, T., Olyarchuk, J.G., and Wei, L. (2005). Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. Bioinformatics **21,** 3787-3793.

Marquez, Y., Brown, J.W., Simpson, C., Barta, A., and Kalyna, M. (2012). Transcriptome survey reveals increased complexity of the alternative splicing landscape in Arabidopsis. Genome Res **22,** 1184-1195.

Martinez, N.J., and Walhout, A.J. (2009). The interplay between transcription factors and microRNAs in genome-scale regulatory networks. Bioessays **31,** 435-445.

Mattick, J.S., Amaral, P.P., Dinger, M.E., Mercer, T.R., and Mehler, M.F. (2009). RNA regulation of epigenetic processes. Bioessays **31,** 51-59.

Meyer, E., Logan, T.L., and Juenger, T.E. (2012). Transcriptome analysis and gene expression atlas for Panicum hallii var. filipes, a diploid model for biofuel research. Plant J **70,** 879-890.

Moussa, M., and Mandoiu, II. (2018). Single cell RNA-seq data clustering using TF-IDF based methods. BMC Genomics **19,** 569.

Oshlack, A., Robinson, M.D., and Young, M.D. (2010). From RNA-seq reads to differential expression results. Genome Biol **11,** 220.

Richterich, P. (1998). Estimation of Errors in ‘‘Raw’’ DNA Sequences: A Validation Study. Genome Res **251**.

Salzman, R.A., Brady, J.A., Finlayson, S.A., Buchanan, C.D., Summer, E.J., Sun, F., Klein, P.E., Klein, R.R., Pratt, L.H., Cordonnier-Pratt, M.M., and Mullet, J.E. (2005). Transcriptional profiling of sorghum induced by methyl jasmonate, salicylic acid, and aminocyclopropane carboxylic acid reveals cooperative regulation and novel gene responses. Plant Physiol **138,** 352-368.

Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Scholkopf, B., Weigel, D., and Lohmann, J.U. (2005). A gene expression map of Arabidopsis thaliana development. Nat Genet **37,** 501-506.

Sekhon, R.S., Lin, H., Childs, K.L., Hansey, C.N., Buell, C.R., de Leon, N., and Kaeppler, S.M. (2011). Genome-wide atlas of transcription during maize development. Plant J **66,** 553-563.

Shakoor, N., Nair, R., Crasta, O., Morris, G., Feltus, A., and Kresovich, S. (2014). A Sorghum bicolor expression atlas reveals dynamic genotype-specific expression profiles for vegetative tissues of grain, sweet and bioenergy sorghums. BMC Plant Biol **14**.

Sheng, Q., Vickers, K., Zhao, S., Wang, J., Samuels, D.C., Koues, O., Shyr, Y., and Guo, Y. (2017). Multi-perspective quality control of Illumina RNA sequencing data analysis. Brief Funct Genomics **16,** 194-204.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol **28,** 511-515.

Vanderlip, R.L., and Revees, H.E. (1972). Growth stages of sorghum (Sorghum Bicolor L.). Agronomy Journal **64**.

Wang, L., Feng, Z., Wang, X., Wang, X., and Zhang, X. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics **26,** 136-138.

Weiss, R.A. (1998). Viral RNA-dependent DNA Polymerase RNA-dependent DNA Polymerase in Virions of Rous Sarcoma Virus. Medical Virology **8,** 3-11.

Wu, J., Mao, X., Cai, T., Luo, J., and Wei, L. (2006). KOBAS server: a web-based platform for automated annotation and pathway identification. Nucleic Acids Res **34,** W720-724.

Young, M.D., Wakefield, M.J., Smyth, G.K., and Oshlack, A. (2010a). Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol **11,** R14.

Young, M.D., Wakefield, M.J., Smyth , G.K., and Oshlack, A. (2010b). Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol **11**.

Zhang, J.Y., Lee, Y.C., Torres-Jerez, I., Wang, M., Yin, Y., Chou, W.C., He, J., Shen, H., Srivastava, A.C., Pennacchio, C., Lindquist, E., Grimwood, J., Schmutz, J., Xu, Y., Sharma, M., Sharma, R., Bartley, L.E., Ronald, P.C., Saha, M.C., Dixon, R.A., Tang, Y., and Udvardi, M.K. (2013). Development of an integrated transcript sequence database and a gene expression atlas for gene discovery and analysis in switchgrass (Panicum virgatum L.). Plant J **74,** 160-173.

Zimmermann, P., Bleuler, S., Laule, O., Martin, F., Ivanov, N.V., Campanoni, P., Oishi, K., Lugon-Moulin, N., Wyss, M., Hruz, T., and Gruissem, W. (2014). ExpressionData - A public resource of high quality curated datasets representing gene expression across anatomy, development and experimental conditions. BioMed Centeral **7**.

CHAPTER V

5. CONCLUSION

Global population growth and the changing climate are the rigorous pressures to the world, while the first one increases the demand for food, the second one creates a problem in the production of enough produces to the required supply. Either way, they have a significant combined effect. Crop production improvement shows tremendous progress, however, the available resources in different countries and the technological advances of countries are key determinant factors. Therefore, the need for a universal strategy of advanced and modernized crop production is immense. Plant improvement program is one of the key components of advanced agriculture. These include breeding of adaptable and high yielding crop varieties with good quality and safety. The variation between gene pool is the most important resource for successful breeding and for getting desired traits. There are different methods of getting variation, such as naturally existing variation and induced variation.

The slow and random mutation frequency of natural variation, which greatly varies based on the nature of the species and the genome becomes the greatest bottleneck to accelerate the breeding programs. This leads to a need to develop induced mutation. There are different mutation methods depending on the species, plant life cycle, the nature of pollination, and the genome size. Chemical mutagenesis, irradiation mutagenesis, insertional mutagenesis are the common methods.

Despite one of the oldest methods, fast neutron radiation is becoming an indispensable tool for the generation of mutants in plants. Arabidopsis, maize, rice, sorghum, soybean, and other plants were all subjected to fast neutron radiation mutation. Optimization of the appropriate doses and appropriate screening methods are the key requirements for successful mutation of plants including sorghum.

This study is about sorghum mutagenesis and description of a specific mutant (*svd* for shriveled seed) in sorghum. Firstly, the study of *svd* mutant in itself is a crucial finding to associate more sorghum accession collections globally and further study the link between SVD gene and sorghum seed shape and size. Secondly, the overall approach of the process from mutant phenotype crossing back and segregation analysis followed by whole genome sequencing to find the cause of the mutant phenotype is much easier than the previous methods which requires more time and resource. However, sequencing of the three mutants is not enough to guarantee a complete approach and requires more mutants to be sequenced and studied for the development of a feasible gene identification for sorghum and to create a sequence indexed mutants' database.

SVD was studied using different approaches including phenotypic characterization, molecular characterization and biochemical characterization. SVD protein is primary localized in plasma membrane. However, it would be more informative if it can be examined using different subcellular markers including the lysosomes, Golgi complex, vacuoles and other subcellular organelles. This will give a gives a complete insight if SVD is involved in trafficking of copper within the cell between different organelles in addition to the detoxification work. Furthermore, overexpression of SVD in sorghum and knockout of SVD from sorghum using CRISPR based gene editing will help for a complete understanding of the functions of SVD. The knockout approach will give a clear understanding by targeting the specific metal binding sites and knocking out from the wildtype reference genome and on the other hand overexpression of SVD in the WT (BTx623) will give a clue on the quantitative significance of SVD on detoxification or loading and transport of copper in the cell.

APPENDICES

6. Accession numbers used in the study

All sorghum sequences are retrieved from Phytozome V13, SbHMA5/SVD: Sobic.006G173800.1, SbHMA3: Sobic.002G083100.1, SbHMA6: Sobic.004G057400.1 , SbHMA5-like:Sobic.006G173700.1 , SbHMA4:Sobic.004G079900.1 , SbHMA9:Sobic.010G220600.1 , SbHMA7:Sobic.007G225000.2, SbHMA8:Sobic.001G484400.1 , SbHMA2:Sobic.010G250900.3, SbHMA3-like:Sobic.002G083000.1, SbHMA1:Sobic.010G240500.2, and all rice sequences are retrieved from rice genome annotation project database : (rice.plantbiology.msu.edu/analyses_search_blast.shtml), OsHMA1: LOC_Os06g47550.1, OsHMA2: LOC_Os06g48720.1, OsHMA3: LOC_Os07g12900.1, OsHMA4: LOC_Os02g10290.1, OsHMA5: LOC_Os04g46940.1: , OsHMA6:LOC_Os02g07630.1, OsHMA7:LOC_Os08g37950.1, OsHMA8:LOC_Os03g08070.1, OsHMA9: LOC_Os06g45500.1, and all Arabidopsis sequences are retrieved from TAIR BLAST 2.9.0^{+}; AtHMA5 : AT1G63440.1, AtHMA7: AT5G44790.1, AtHMA8: AT5G21930.1, AtHMA6: AT4G33520.2, AtHMA2:AT4G30110.1, AtHMA4 : AT2G19110.1, AtHMA3: AT4G30120.1, AtHMA1: AT4G37270.1

Table 7: All Primer used in the study

987-1-3 F Transcript	TTGACCTTTCCACTCCGAT
556-1-4A F	ATGTCGAAGCCCGCGGA
556-1-4A R	TCACCCCCAGAACCTGCTCA
556-1-4B F	ATGGGTATCTCGCGTGACT
556-1-4B R	CTAAGATGCAGCACCCTTG
556-1-4C F DNA primer	GTGTATCATCTTCAATCATGC
556-1-4C R DNA Primer	AGCATGAGGCCGTGGATGT
556-1-4D F DNA primer	TCCTCAGGCCCATGTACTCTG
556-1-4D R DNA primer	CGGTCTTGAGGAGGTCAATTTCCT
556-1-4E F Transcript	ACAAGGCATATGGCGGTG
556-1-4E R Transcript	ATGAGGTAGTGGGTGAACCCG
Cu exp ATPase attb1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGAATATGAGTCAGCTGCCCTTT
Cu exp ATPase attb2	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTTCCAGCTGTGTCATCAACT
987-1-3 attb1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCTTGACCTTTCCACTCCGAT
987-1-3attb2	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTACATCAAGCGCTTGTCT
pGWB521-seq-F	acttgaacggtagcgctg
pGWB521-seq-R	tcggggaaattcgagctct
Olige dT-18	TTTTTTTTTTTTTTTTT
Cu exporting Transc- F	GAATATGAGTCAGCTGCCCTTT
Cu exporting Transc- R	CTTCCAGCTGTGTCATCAACT
Cu- ATPase 2997 F	ATGGTGGCGACTACTCGAGCTCT
Cu- ATPase 2997 R	TCAACTGCCAACAATCTTTGGAGCCT
Cu ATPase 2997 Tran F	TGTAGTGTTGCGTCTCCTTGCT
Cu ATPase 2997 Tran R	CAAATGATATGCCCGTCTTCCAGC
Cu- ATPase 2997 F	ATGGTGGCGACTACTCGAGCTCT
Cu- ATPase 2997 R	TCAACTGCCAACAATCTTTGGAGCCT

Cu ATPase 2997 Tran F	TGTAGTGTTGCGTCTCCTTGCT
Cu ATPase 2997 Tran R	CAAATGATATGCCCGTCTTCCAGC
ATPase 543F	ATAATGGTGACTGGCGACAACTGG
ATPase 543 R	GGCAAATGATATGCCCGTCTTCCA
ATPase 2529	ATGGCCTGCAAGTATTGCACA
ATPase T-2 F (op)	ATAATGGTGACTGGCGACAACTGG
ATPase T-2 R (op)	GGCAAATGATATGCCCGTCTTC
ATPase T-3 F (op)	AATGGTGACTGGCGACAACT
ATPase T-3 R (op)	GGCAAATGATATGCCCGTCTTCCA
ATPase 2529	TCAACTGCCAACAATCTTTGGAGC
Cu- ATPase 2997 attb1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGTGGCGACTACTCGAGCTCT
Cu- ATPase 2997 attb2	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAACTGCCAACAATCTTTGGAGCCT
Cu- ATPase 2997 F	aagcttATGGTGGCGACTACTCGAGCTCT
Cu- ATPase 2997 R	ggatccTCAACTGCCAACAATCTTTGGAGCCT
Cu- ATPase 2997 F	aagcttATGGTGGCGACTACTCGAGCTCT
Cu- ATPase 2997 R	ggatccTCAACTGCCAACAATCTTTGGAGCCT
SbFRN3 attb1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGGCATCCTGGACCACT
SbFRN3 attb2	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCACATGACCGTGCATGCGT
SbCuBL attb1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGACGATCGTAGAGATGC
SbCuBL attb2	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTACATGATCGAGCAAGCA
SbHMT1 attb1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGCTTAAGGAGACCATTCTCAC
SbHMT1 attb2	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTACATCGACGAGCAGATGTCA
SbATX1 attb1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCGGCCCAGACTGTCG
SbATX1 attb2	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGGCAGTAGTAGCATCAGCA
SbB0811B10.6 F	ATGGCCGCCAAGAAGGTG
SbB0811B10.6 R	TCAGCAGATGACGCAGGAGT
SbB0811B10.6 attb1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCCGCCAAGAAGGTG
SbB0811B10.6 attb2	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGCAGATGACGCAGGAGT
SbATOX1-Like attb1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGAGGGCGGGCGGGATGCTGT
SbATOX1-Like attb2	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAGGCGAGGAGCTCGGCCTTCT

Sbipp23 attb1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGGAGGCGTGGCATCGGA
Sbipp23 attb2	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCACATAAGGGAGCAGGCGTTG
ATPase qPCR primers	
HMA5-L qPCR 1 F	TCACGGCCAGTAGCAAAGAG
HMA5-L qPCR 1R	TTTGCCATTTGCGCTGACTC
HMA5-L qPCR 2 F	TTGTGGCCATCGACCAAGAA
HMA5-L qPCR 2 R	GCATTAGCTGTCCCCCAGTT
HMA5-L qPCR 3 F	ATGTGATCACCGCCATCGAC
HMA5-L qPCR 3 R	GACGACACTGACGGACGAAG
Cu- ATPase 2997 attb1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGTGGCGACTACTCGAGCTCT
ATPase 2997 nostop	GGGGACCACTTTGTACAAGAAAGCTGGGTCACTGCCAACAATCTTTGGAGCCT
Attb sequences	
attb1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTC
attb2	GGGGACCACTTTGTACAAGAAAGCTGGGTC
Yeast two hybrid interaction	
SbDETO1 attb1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCCTTCGCACCTTCTTCCATCT
SbDETO2 attb2	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCACATGACGCTGCACGC
SbDDCP attb1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGAGCAAGGAGGACGTG
SbDDCP attb2	GGGGACCACTTTGTACAAGAAAGCTGGGTCTCACATCACGGTGCAGCTGTT
SbJNB attb1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCAGATCGATCGATCGGTACACCA
SbJNB attb2	GGGGACCACTTTGTACAAGAAAGCTGGGTCAGCGTCTCCATCCATGCATCG
SbCCH attb1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATGGCTCAGCAAAAGGTGG
SbCCH attb2	GGGGACCACTTTGTACAAGAAAGCTGGGTCTAGGTGAGGCCATGTGATG
HMA5pro attb1	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGTAAGGGTACGTACAGTTGGA
HMA5pro attb2	GGGGACCACTTTGTACAAGAAAGCTGGGTCACTGAAAAGGGCAGCTGACTC
Realtime primers qPCR primers	
SbIMK3-F	CTACCTCCACAACAACCGCTTC

SbIMK3-R	GCTTTGTGGCGTTGCCAAT
SbKIN-5A-F	TGGTGTCATTCCAAGGGCTGTT
SbKIN-5A-R	GTGAGCGACTGCTTTGCTTGTT
SbH2AXb-F	ACCTCAAGGCCGGCAAGTA
SbH2AXb-R	GCAGCGTCTGGTGGATGTT
SbMit-NPK1-F	GGTGAATTGGCCAAACATGCTG
SbMit-NPK1-R	AAGTATCGCATGCACCTCCACA
SbKIN-12C-F	AGAACAGAAGGCTCCGCATTCA
SbKIN-12C-R	TCAATGGAGCATCGCACGAGTA
SbLRR-ERECTA-F	ATGGCAGCCTCTGGGACATTTT
SbLRR-ERECTA-R	TTTCTCGTTGAGCCGGGATGTT
SbKi-Nek1-F	GGAGATGCAGCTCATTGCTACA
SbKi-Nek1-R	AGCAAGCCCAAAGTCACCAAGT
SbCKB1-1-F	TTCACCGGGATTTGAAGCCA
SbCKB1-1-R	GCCAATCCCTCAATTCACTCAC
SbLRR-Atg37250-F	AACGTGAAGCCGAGCAACAT
SbLRR-Atg37250-R	AGCACGCCGAACGAGTACA
SbH2A-F	GTCAGTGTCGAGGTCGGTGAA
SbH2A-R	AGCAGAACCGGGTTGATGTTG
SbH4-F	AGGTGCTCCGCGACAACAT
SbH4-R	TTAGCCGCCGAAACCGTAGA
SbH2B-F	AGAAGCAGAAGCAGAAGCAG
SbH2B-R	TCGAACATGTCGGCCATCA
SbMito-kin14L-F	ACGCAAGCAGTAACCTGCAGAA
SbMito-kin14L-R	TCGTCCAGGGATCTTGAAAGCA
SbH3-F	GTGAGATCCGCAAGTACCAGAA
SbH3-R	TGGATGTCCTTGGGCATGAT
SbABC14-F	GCTGGCAAGTTGCAATGCTTTC
SbABC14-R	ACTGCCCCAATCCAAACCATCA
SbRPR12-F	ACAACTTCTCAGGCACGCTCAA

SbRPR12-R	AGATGGAATCGGTCCCGTGAAT
SbbHLH87-F	TCGATCACCTCCAGGACTGCTA
SbbHLH87-R	GACGCCGTGTCCATCTTGGT
SbZFNPBF-F	CAGGCTTATTCCCCAATGTGCT
SbZFNPBF-R	CATGCCCATGCATTGCACCATA
SbGST3-F	TGCGGCTCCGTCAACATTCT
SbGST3-R	GGCACTCCACGTCAAGATCAAA
SbCa-unipr2-F	TCCAGCTCGACGGGCTTAT
SbCa-unipr2-R	ATGGCCTTCGCGACCATGT
SbCytP450-F	CACCATCTGCAAGATCTCCTTC
SbCytP450-R	ATGTCGCGGAGGTACTTGT
SbF-boxAtg67480-F	AATTCATCGCCGTTCGGAAGGA
SbF-boxAtg67480-R	CGCAACATTCAGCTTGGCAAGT
SbNAC21/22-F	CACTGTCCAGCGAGCTGAT
SbNAC21/22-R	AACGGGTTCCACGTCTGCT
SbKKK2-F	AGAGCTGGAGCATCAATTCGTC
SbKKK2-R	CATATGCACATGAGGTGTACGG
SbbHLH69-F	TCGGTGCCTCATCACATGTCTT
SbbHLH69-R	ATCTCCGGGTCGATGTTCACG
SbMan3-F	AACCGCACCCAGAGGGACTA
SbMan3-R	TGCGTACGCGTATCTTCCTGTT
SbNADPH1-F	AAGCTCGTGCCGTTCGACAT
SbNADPH1-R	AACCTCAACGCTACCTGAGCAA
SbTIFFY11-F	ATGCCCACGGACAGCATGA
SbTIFFY11-R	TTCCTGGCGATGGGCATGT
SbRAP2-13-F	AAGCTCCACGCCATCTGCAA
SbRAP2-13-R	TCAGGATGGAGTCCCAGTCGAT
SbPolyol1-F	AGCGCCTGCTCGACATCAA
SbPolyol1-R	AGCGTCGAGATGGGTATGAAGA
SbWRKY72-F	GCTCTCCCTCGGTCTTGGTTT

SbWRKY72-R	CATCGTGGGAGTGTCGCATTTG
SbSul-tra3.4-F	ACGAGTGGAAATGGCAGACGAT
SbSul-tra3.4-R	TTGATGGATGCGAAGGTTCTGC
SbCu-Like5.1-F	ACGTCGGCCACGATCCTATT
SbCu-Like5.1-R	TTGAACGACATGACGGAGAGCA
SbZntra2-like-F	TCTGCTTCCACTCCGTCTTCG
SbZntra2-like-R	AGGTGGCTGATGGAGACGTAGA
SbRHA1B-F	TGATGGGATTCCCCATGGGCTA
SbRHA1B-R	TGGAAGTCGCCCAGGCAGA
SbSKIP14-F	GCAAGTCACTACGTTCAGCTGT
SbSKIP14-R	TGCTTCTGATGCAGTGACACGA
SbG3PT1-F	GTATGTTGCATTGGCCCAGTCA
SbG3PT1-R	ACACCATCTGCATGCCCAGAAA
SbPAP15-F	ACCGACTGCTATTCCTGCTCAT
SbPAP15-R	TACTGCGCACCTGATCTGTT

VITA

DIMIRU TILAHUN TADESSE

Candidate for the Degree of

Doctor of Philosophy

Dissertation*:* *SHRIVELED SEED* (*SVD*), A P_{1B} TYPE HMA5 TRANSPORTER ATPASE, IS INVOLVED IN COPPER DETOXIFICATION TO CONTROL THE GROWTH AND DEVELOPMENT OF SORGHUM

Major Field: Crop Science / Concentration: Molecular Biology

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Crop Science at Oklahoma State University, Stillwater, Oklahoma in December 2020.

Completed the requirements for the Master of Science in Biology/ Genetics at Wollo University, Dessie, Ethiopia in November 2014.

Completed the requirements for the Bachelor of Science in Plant Science at Bahir Dar University, Bahir Dar, Ethiopia in October 2011.

Experience:
Dimiru worked as an extension worker (2007-08), as a supervisor (2008-2010), in governmental agriculture offices. He also served as a technical teaching assistant (2010-2011), as a graduate teaching assistant (2012-2014) and as a lecturer (2014-2016) in Wollo University.

Professional memberships:
1. Member of Plant and Soil Sciences Graduate Student Organization (PaSS-GSO) of Oklahoma State University (2016- 2020).
2. ASPB- American Society of Plant Biologists member.
3. ASA-CSSA-SSSA member.
4. AAAS- American association for the advancement of science.