

# GENOMIC APPROACHES FOR IMPROVEMENT OF UNDERSTUDIED GRASSES

EDITED BY : Keenan Amundsen, Gautam Sarath and Teresa Donze-Reiner  
PUBLISHED IN : Frontiers in Plant Science



# frontiers

## Frontiers Copyright Statement

© Copyright 2007-2017 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88945-242-2

DOI 10.3389/978-2-88945-242-2

## About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [researchtopics@frontiersin.org](mailto:researchtopics@frontiersin.org)



# GENOMIC APPROACHES FOR IMPROVEMENT OF UNDERSTUDIED GRASSES

Topic Editors:

**Keenan Amundsen**, University of Nebraska-Lincoln, United States

**Gautam Sarath**, United States Department of Agriculture—Agricultural Research Service, United States

**Teresa Donze-Reiner**, West Chester University, United States



Mowed buffalograss in Mead, Nebraska, USA.

Photo by Keenan Amundsen

Grasses are diverse, spanning native prairies to high-yielding grain cropping systems. They are valued for their beauty and useful for soil stabilization, pollution mitigation, biofuel production, nutritional value, and forage quality; grasses encompass the most important grain crops in the world. There are thousands of distinct grass species and many have promiscuous hybridization patterns, blurring species boundaries. Resources for advancing the science and knowledgebase of individual grass species or their unique characteristics varies, often proportional to their perceived value to society. For many grasses, limited genetic information hinders research progress. Presented in this research topic is a brief snapshot of creative efforts to apply modern genomics research methodologies to the study of several minor grass species.

**Citation:** Amundsen, K., Sarath, G., Donze-Reiner, T., eds. (2017). Genomic Approaches for Improvement of Understudied Grasses. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-242-2

# Table of Contents

## Introduction

- 05 Editorial: Genomic Approaches for Improvement of Understudied Grasses**  
Keenan Amundsen, Gautam Sarath and Teresa Donze-Reiner

## Stress tolerance

- 07 Genotypic Variation for Salinity Tolerance in *Cenchrus ciliaris* L.**  
Abdullah J. Al-Dakheel and M. Iftikhar Hussain
- 19 Comparative Analysis of WRKY Genes Potentially Involved in Salt Stress Responses in *Triticum turgidum* L. ssp. durum**  
Fatma-Ezzahra Yousfi, Emna Makhoulfi, William Marande, Abdel W. Ghorbel, Mondher Bouzayen and Hélène Bergès
- 34 De novo Assembly and Characterization of the Transcriptome of Broomcorn Millet (*Panicum miliaceum* L.) for Gene Discovery and Marker Development**  
Hong Yue, Le Wang, Hui Liu, Wenjie Yue, Xianghong Du, Weining Song and Xiaojun Nie
- 47 Water Deficit Affects Primary Metabolism Differently in Two *Lolium multiflorum*/ *Festuca arundinacea* Introgression Forms with a Distinct Capacity for Photosynthesis and Membrane Regeneration**  
Dawid Perlikowski, Mariusz Czyżniewski, Łukasz Marczak, Adam Augustyniak and Arkadiusz Kosmala

## Molecular marker development and applications

- 63 Transcriptome Profiling of Buffalograss Challenged with the Leaf Spot Pathogen *Curvularia inaequalis***  
Bimal S. Amaradasa and Keenan Amundsen
- 75 Validating DNA Polymorphisms Using KASP Assay in Prairie Cordgrass (*Spartina pectinata* Link) Populations in the U.S.**  
Hannah Graves, A. L. Rayburn, Jose L. Gonzalez-Hernandez, Gyoungju Nah, Do-Soon Kim and D. K. Lee
- 87 Implementation of Genomic Prediction in *Lolium perenne* (L.) Breeding Populations**  
Nastasiya F. Grinberg, Alan Lovatt, Matt Hegarty, Andi Lovatt, Kirsten P. Skøt, Rhys Kelly, Tina Blackmore, Danny Thorogood, Ross D. King, Ian Armstead, Wayne Powell and Leif Skøt
- 97 Identification and Characterization of Switchgrass Histone H3 and CENH3 Genes**  
Jiamin Miao, Taylor Frazier, Linkai Huang, Xinquan Zhang and Bingyu Zhao



## Enhanced biomass production

- 109 Identification, Characterization, and Expression Analysis of Cell Wall Related Genes in Sorghum bicolor (L.) Moench, a Food, Fodder, and Biofuel Crop**  
Krishan M. Rai, Sandi W. Thu, Vimal K. Balasubramanian, Christopher J. Cobos, Tesfaye Disasa and Venugopal Mendu
- 128 Integrative analysis and expression profiling of secondary cell wall genes in C<sub>4</sub> biofuel model Setaria italica reveals targets for lignocellulose bioengineering**  
Mehanathan Muthamilarasan, Yusuf Khan, Jananee Jaishankar, Shweta Shweta, Charu Lata and Manoj Prasad
- 149 Proteomic Responses of Switchgrass and Prairie Cordgrass to Senescence**  
Bimal Paudel, Aayudh Das, Michael Long Tran, Arvid Boe, Nathan A. Palmer, Gautam Sarath, Jose L. Gonzalez-Hernandez, Paul J. Rushton and Jai S. Rohila



# Editorial: Genomic Approaches for Improvement of Understudied Grasses

Keenan Amundsen<sup>1\*</sup>, Gautam Sarath<sup>2</sup> and Teresa Donze-Reiner<sup>3</sup>

<sup>1</sup> Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE, United States, <sup>2</sup> Grain, Forage and Bioenergy Research Unit, United States Department of Agriculture—Agricultural Research Service, Lincoln, NE, United States, <sup>3</sup> Department of Biology, West Chester University, West Chester, PA, United States

**Keywords:** biomass yield, genotypic diversity, grasses, stress tolerance, RNA-seq, proteomics, genomics, differential gene expression

## Editorial on the Research Topic

### Genomic Approaches for Improvement of Understudied Grasses

Grasses are diverse, spanning native prairies to high-yielding grain cropping systems. They are valued for their beauty and useful for soil stabilization, pollution mitigation, biofuel production, nutritional value, and forage quality; grasses encompass the most important grain crops in the world. There are thousands of distinct grass species and many have promiscuous hybridization patterns, blurring species boundaries. Resources for advancing the science and knowledgebase of individual grass species or their unique characteristics varies, often proportional to their perceived value to society. For many grasses, limited genetic information hinders research progress. Presented in this research topic is a brief snapshot of creative efforts to apply modern genomics research methodologies to the study of several minor grass species.

Native or naturalized grass species offer unique adaptation advantages and often have better heat, drought, and salinity tolerance than recently introduced species. Genotypes with unique combinations of traits frequently arise and many offer unique and robust sources of stress resistance or desirable production characteristics, but identification of those plants is challenging. Al-Dakheel and Hussain present a novel method for field-evaluating and identifying salinity tolerance in buffelgrass (*Cenchrus ciliaris* L.). Using a field-pot system over three years, 12 salt tolerant, high dry biomass yielding buffelgrass accessions were identified. Salinity tolerance is important in modern agriculture and particularly for many alternative grasses as they are often managed on marginal lands. Yousfi et al. specifically assayed the role of WRKY genes in salinity tolerance in durum wheat (*Triticum turgidum* L.). Five WRKY-containing bacterial artificial chromosomes (BACs) were identified, sequenced, and annotated. Differential response in WRKY genes in salt sensitive vs. salt tolerant germplasm was observed suggesting their role in salt tolerance. Another interesting finding, common to many grasses was the observation that 74.6% of the sequenced BACs contained transposable elements, often found in high copy numbers in grasses and further complicating genetic studies. Yue et al. used a global transcriptome approach to identify transcripts differentially expressed between a waxy, drought sensitive cultivar of broomcorn millet (*Panicum miliaceum* L.) and one that was non-waxy and salt and drought tolerant. Yue et al. reported the first assembly of broomcorn millet and identified 292 differentially expressed transcripts between the studied cultivars. These transcripts may be important in the morphological differences associated with waxiness, and drought and salinity tolerance. In addition to salinity tolerance, drought tolerance is essential for grasses grown on marginal lands. Hybrids can form between Italian ryegrass (*Lolium multiflorum* Lam.) and tall fescue (*Festuca arundinacea* Schreb.) and are exploited to combine

## OPEN ACCESS

### Edited by:

Sergio Lanteri,  
University of Turin, Italy

### Reviewed by:

Sergio Lanteri,  
University of Turin, Italy  
Rieseberg Loren,  
University of British Columbia, Canada

### \*Correspondence:

Keenan Amundsen  
kamundsen2@unl.edu

### Specialty section:

This article was submitted to  
Crop Science and Horticulture,  
a section of the journal  
Frontiers in Plant Science

**Received:** 25 April 2017

**Accepted:** 23 May 2017

**Published:** 09 June 2017

### Citation:

Amundsen K, Sarath G and  
Donze-Reiner T (2017) Editorial:  
Genomic Approaches for  
Improvement of Understudied  
Grasses. *Front. Plant Sci.* 8:976.  
doi: 10.3389/fpls.2017.00976

desirable traits from each species. Perlikowski et al. examined two hybrid forms differing in their photosynthetic capacity during drought stress and their ability for membrane regeneration following removal of the stress, offering insights into the role of metabolic alterations on drought tolerance and membrane recovery.

Amaradasa and Amundsen studied the interaction between a fungal pathogen, *Curvularia inaequalis*, and resistant and susceptible American buffalograss [*Buchloë dactyloides* (Nutt.) Engelm. syn. *Bouteloua dactyloides* (Nutt.) Columbus]. Their analysis led to the development of RNA-based markers that have potential to screen and identify sources of host resistance in the absence of an *in vivo* assay. These markers have the added advantage of being gene-based, avoiding some of the challenges associated with genetic studies in complex polyploid genomes. Genetic marker development and testing is an important early step in working with grass species lacking genomic resources. As a part of the Yue et al. study, they identified 35,216 simple sequence repeat sequences in broomcorn millet that could be developed into molecular markers. Graves et al. used single nucleotide polymorphic markers (SNPs) to develop KASP assays. The KASP assays were able to discriminate hybridization and self-fertilization events in populations of prairie cordgrass (*Spartina pectinata* Link), which is challenging in a species with diverse and complex inheritance patterns. The ability to exploit important traits from breeding populations is critically important in order to maximize their value. Grinberg et al. used perennial ryegrass (*Lolium perenne* L.) breeding populations as a model to compare different predictive models in a genomic prediction framework with a goal to ultimately improve several biomass, yield, and nutritional value traits. Miao et al. examined the role of histones, another important conserved protein family essential for genome stability in switchgrass (*Panicum virgatum* L.) by exploiting genome specific markers and transforming tobacco to confirm their functional role. Miao et al. confirmed that the histone genes being investigated could trigger cell death and their nuclear localization was critical for their function.

Biomass is critically important for grasses destined for use by the biofuel industry. Lignin, cellulose, and pectin are cell wall constituents that influence how grasses can be used for biofuel production. Rai et al. conducted a genome-wide analysis to identify genes associated with cell wall composition of sorghum [*Sorghum bicolor* (L.) Moench]. By physically

mapping those genes to the sorghum genome, researchers can use that information to alter cell wall composition through traditional plant breeding methods. Muthamilarasan et al. also examined gene expression profiles in cell wall associated genes in foxtail millet (*Setaria italica* L.) and identified genes differentially expressed in response to abiotic stress and exogenous hormone applications. Muthamilarasan et al. further describe the importance of synteny among grasses and conservation among cell wall genes. The study of Paudel et al. also highlights conservation among the grasses by studying expressed proteins during senescence in switchgrass and prairie cordgrass. Early senescence reduces biomass and therefore is an important process to understand in perennial grasses. By comparing early senescence genotypes with late senescence genotypes in both species, proteins intimately involved in senescence were found and could be exploited in future studies and breeding programs to develop germplasm with delayed senescence.

Here we highlight 11 understudied grass species, relative to the major cereal grasses, which is only a small fraction of the thousands of known species. As land resources become scarce and demand for highly productive arable land increases, identification of understudied grasses and their desirable traits can fast-forward their suitability for use in marginal lands. The papers presented in this research topic demonstrates novel approaches or new applications for proven methods to improve our understanding of perennial grasses and their function and illustrates important steps toward use of these understudied species in modern agriculture.

## AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Amundsen, Sarath and Donze-Reiner. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Genotypic Variation for Salinity Tolerance in *Cenchrus ciliaris* L.

Abdullah J. Al-Dakheel and M. Iftikhar Hussain \*

Crop Diversification and Genetic Improvement Section, International Center for Biosaline Agriculture, Dubai, United Arab Emirates

## OPEN ACCESS

### Edited by:

Teresa Donze,  
University of Nebraska–Lincoln, USA

### Reviewed by:

Caiguo Zhang,  
University of Colorado Denver, USA

Zeran Li,  
Washington University, USA

### \*Correspondence:

M. Iftikhar Hussain  
m.iftikhar@biosaline.org.ae;  
mih786@gmail.com

### Specialty section:

This article was submitted to  
Crop Science and Horticulture,  
a section of the journal  
Frontiers in Plant Science

**Received:** 02 April 2016

**Accepted:** 11 July 2016

**Published:** 28 July 2016

### Citation:

Al-Dakheel AJ and Hussain MI (2016)  
Genotypic Variation for Salinity  
Tolerance in *Cenchrus ciliaris* L.  
Front. Plant Sci. 7:1090.  
doi: 10.3389/fpls.2016.01090

Scarcity of irrigation water and increasing soil salinization has threatened the sustainability of forage production in arid and semi-arid region around the globe. Introduction of salt-tolerant perennial species is a promising alternative to overcome forage deficit to meet future livestock needs in salt-affected areas. This study presents the results of a salinity tolerance screening trial which was carried out in plastic pots buried in the open field for 160 buffelgrass (*Cenchrus ciliaris* L.) accessions for three consecutive years (2003–2005). The plastic pots were filled with sand, organic, and peat moss mix and were irrigated with four different quality water (EC 0, 10, 15, and 20 dS m<sup>-1</sup>). The results indicate that the average annual dry weights (DW) were in the range from 122.5 to 148.9 g/pot in control; 96.4–133.8 g/pot at 10 dS m<sup>-1</sup>; 65.6–80.4 g/pot at 15 dS m<sup>-1</sup>, and 55.4–65.6 g/pot at 20 dS m<sup>-1</sup>. The highest DW (148.9 g/pot) was found with accession 49 and the lowest with accession 23. Principle component analysis shows that PC-1 contributed 81.8% of the total variability, while PC-2 depicted 11.7% of the total variation among *C. ciliaris* accessions for DW. Hierarchical cluster analysis revealed that a number of accessions collected from diverse regions could be grouped into a single cluster. Accessions 3, 133, 159, 30, 23, 142, 141, 95, 49, 129, 124, and 127 were stable, salt tolerant, and produced good dry biomass yield. These accessions demonstrate sufficient salinity tolerance potential for promotion in marginal lands to enhance farm productivity and reduce rural poverty.

**Keywords:** buffelgrass, biomass yield, multivariate analysis, salt tolerance, genotypic diversity

## INTRODUCTION

Salt-affected soils and millions of hectares of marginal lands have limited the scope for crop production (Wang et al., 2012). According to the Food and Agriculture Organization (FAO), 34 million hectares (11% of the total irrigated area of the world) are affected by different levels of salinization (Food Agriculture Organization of the United Nations, 2012). Moreover, the world loses 0.25–0.5 M ha of agricultural land annually because of salt buildup, which mainly results from irrigation, especially in arid and semiarid areas (Peng et al., 2008; Qadir et al., 2014). Soil salinity reduces the productivity of most crops, although to a varying extent depending on species (Roy et al., 2014; Hussain et al., 2016). Besides improving water management practices to reduce salt accumulation in the root zone, there is a need to improve salinity tolerance of strategically important crops. The use of low quality saline water for plant production is an option to conserve limited freshwater resources, particularly for the water-scarce regions of Arabian and African peninsula.

The cultivation of forages, biomass crops, and perennial grasses as feedstock for energy, biomaterials, and livestock rearing have been promoted as an opportunity to improve sustainability of forage supply, energy security, and contributing to the rural development (Ahmad, 2010). Therefore, demand for sustainable biomass production for livestock and energy use has raised the interest in perennial crops like *Cenchrus ciliaris* L. Buffelgrass (*C. ciliaris* L.) is a perennial ( $C_4$ ) forage grass (family poaceae), sometimes produces rhizomes and is native to the Arabian Peninsula. The *C. ciliaris* is dominant in natural grazing zones of Ethiopia (Angassa and Baars, 2000), Australia (Bulldgen and Francois, 1998), and North Africa (Mseddi et al., 2004). Buffelgrass has proved useful for pasture and soil retention in a wide range of environments due to its drought tolerance, high biomass, deep roots, rapid response to summer rains, and resistance to overgrazing. With extensive belowground systems, cultivation of perennial grasses present high efficiencies in the use of nutrient and water resources and control of soil erosion, carbon sequestration with the restoration of soil properties (fertility, structure, organic matter). Compared with annual systems, herbaceous perennial crops have the advantages of erodibility, and crop management options, such as pesticides and fertilizers inputs (Zhang et al., 2011; Fernando et al., 2012). The salt tolerance of different *C. ciliaris* genotypes need to be evaluated to test their suitability for marginal environments to offer a more practical solution for effective utilization of salt affected soils. Among buffelgrass, accession from North America, “Texas 4464” has been reported as drought tolerant (Ayerza, 1981), and “Biloela” as salt-tolerant (Graham and Humphreys, 1970). Therefore, strategies for mitigating salinity problems in crop production include both development of management options (Shannon, 1997) and genetic improvement of current cultivars (Krishnamurthy et al., 2007).

Germplasm of a specific crop collected from the diverse sources offers greater genetic diversity and may furnish useful traits to widen the genetic base of crop species. The collection, screening and description of the existing variability among the forage crops are the first step in the performance evaluation and selection process (Ponsens et al., 2010). Knowledge about germ-plasm diversity and salinity tolerance evaluation will be an excellent tool to screen and select high yielding accessions for further evaluation under field conditions. Screening large numbers of genotypes for salinity tolerance in the field is notoriously difficult because of the variability of salinity within fields (Daniells et al., 2001). Moreover, it would be difficult to determine the critical parameters under field conditions since any environmental change could result in dramatic change in the plant's response to salinity (Shannon, 1997). Although *C. ciliaris* response to salinity stress has been a topic of many researchers (Arshad et al., 2000; Hacker and Waite, 2001; Jorge et al., 2008; Ksiksi and El-Shaigy, 2012); to best of our knowledge, no study has evaluated and characterized the *C. ciliaris* genotypes in terms of agromorphological attributes and dry matter yield responses so far. This study evaluates the morphological and biomass yield responses of *C. ciliaris* genotypes to water salinity in pot culture trial.

In a study of salt stress on buffelgrass and its effects on productivity decline; Lanza Castelli et al. (2010) has found that *C. ciliaris* accession, Texas 4464, is susceptible to salt stress at 300 mM NaCl concentrations at the seedling stage, while Americana showed tolerance against salinity. The fresh weight, root length, and plant height of these accessions were least affected by salinity. However, screening and selection of large numbers of buffelgrass genotypes for salinity tolerance is lacking. Therefore, the current research was undertaken with the aim to identify superior genotypes for forage production under hot arid conditions of UAE using naturally available low quality saline water. The results of this research are expected to provide useful information that can be used to group accessions for relative comparison and evaluation of biomass yield in areas where problems of soil and water salinity are increasing.

## MATERIALS AND METHODS

### Site Description

The field trials were conducted for three growing seasons during 2003–2005 on the experimental farm of the International center for Biosaline Agriculture (ICBA), located on the eastern side of Dubai between  $25^{\circ}5' N$  and  $55^{\circ}23' E$  with an elevation of 30 m above mean sea level. The soil of the experimental field is Carbonatic, Hyperthermic Typic Torripsamment having a negligible level of inherent soil salinity ( $0.2 \text{ dS m}^{-1}$ ). The study area is characterized by very hot, dry days in summer (April–October), when temperatures can reach  $45^{\circ}\text{C}$ , while the winter days (December–February) are mainly cooler and dry with low average night-time temperatures ( $10^{\circ}\text{C}$ ). Because of the aridity, and relatively cloudless skies, there are great extremes of temperature variation, but there are also wide variations between the seasons. Air temperature follows a regular seasonal trend, with a minimum in January and a maximum in July. According to the ICBA weather station data the annual average temperature of the study area was  $10^{\circ}\text{C}$  during the month of January while average maximum temperature during summer, July, was  $45^{\circ}\text{C}$ . The average annual rainfall is  $<80 \text{ mm}$  and falls in short, torrential bursts during the summer months.

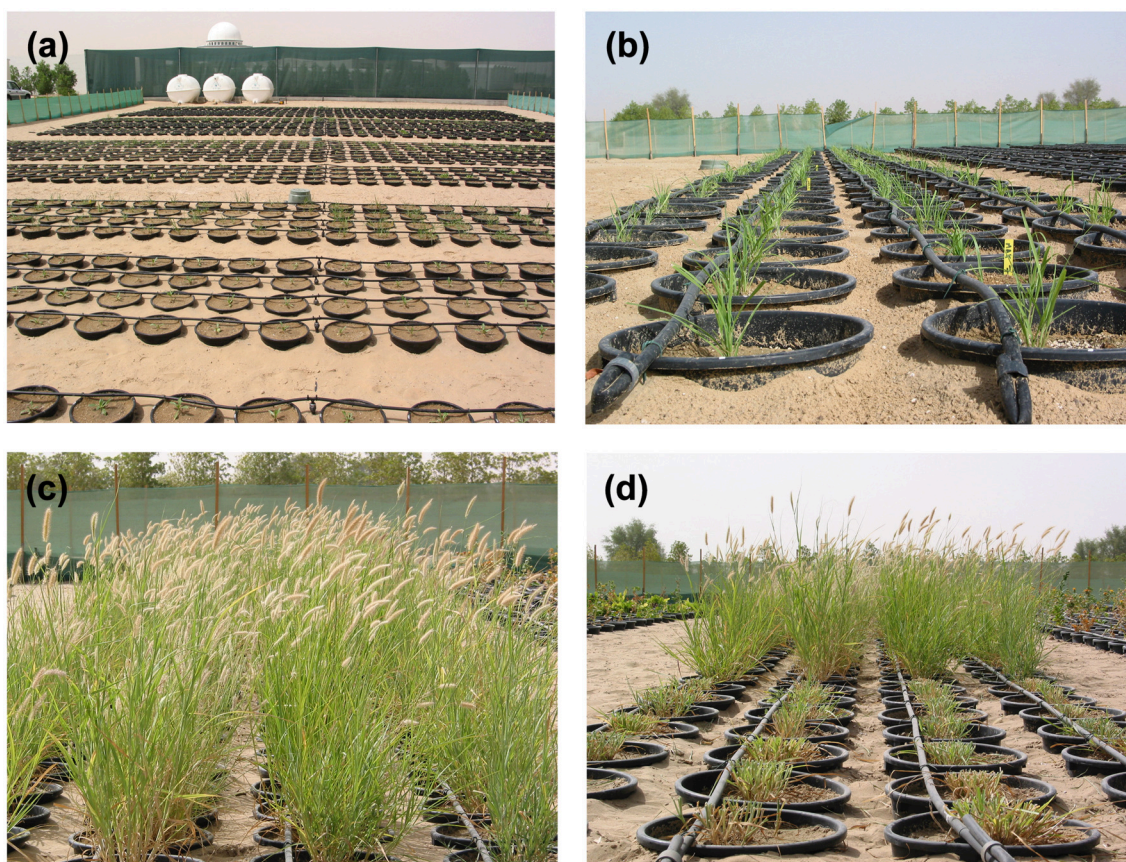
### Plant Material, Experimental Design, and Management Practices

In total, 160 accessions of buffelgrass (*C. ciliaris* L.) were used in this experiment as shown in **Table 1**. The seeds of *C. ciliaris* L. were received from United States Department of Agriculture (USDA); originated from Asia, Africa, Australia, India, USA and some local landraces and commercial cultivars were also included in the study. Seeds of *C. ciliaris* genotypes (three per pot) were sown in poly vinyl chloride pots ( $30 \times 30 \text{ cm}$ ) on 2nd March 2003. The pots were buried in the open field (**Figure 1**), near the net house at experimental research station (ICBA, Dubai, UAE) for three consecutive years (2003–2005). After uniformity of emergence, seedlings were thinned to one plant per pot to maintain good stand establishment. Experimental soil achieved field capacity and permanent wilting point at 0.1 bar and 15 bars,

TABLE 1 | Buffel grass (*Cenchrus ciliaris* L.) germplasm collection with genebank entry numbers and country of origin.

S. No.	Accessions	Species	Country of origin	S. No.	Accessions	Species	Country of origin	S. No.	Accessions	Species	Country of origin	S. No.	Accessions	Species	Country of origin
1	ICBA 22	<i>C. ciliaris</i>	UAE	41	PI 299521	<i>C. ciliaris</i>	South Africa	81	PI 409263	<i>C. ciliaris</i>	South Africa	121	PI 409528	<i>C. ciliaris</i>	South Africa
2	ICBA 38	<i>C. ciliaris</i>	UAE	42	PI 307619	<i>C. ciliaris</i>	Australia	82	PI 409264	<i>C. ciliaris</i>	South Africa	122	PI 409556	<i>C. ciliaris</i>	South Africa
3	PI 153671	<i>C. ciliaris</i>	Kenya	43	PI 308595	<i>C. ciliaris</i>	Australia	83	PI 409265	<i>C. ciliaris</i>	South Africa	123	PI 409581	<i>C. ciliaris</i>	South Africa
4	PI 156546	<i>C. ciliaris</i>	Zimbabwe	44	PI 315681	<i>C. ciliaris</i>	United States	84	PI 409266	<i>C. ciliaris</i>	South Africa	124	PI 409585	<i>C. ciliaris</i>	South Africa
5	PI 161631	<i>C. ciliaris</i>	South Africa	45	PI 339892	<i>C. ciliaris</i>	Australia	85	PI 409267	<i>C. ciliaris</i>	South Africa	125	PI 409653	<i>C. ciliaris</i>	South Africa
6	PI 161633	<i>C. ciliaris</i>	South Africa	46	PI 365650	<i>C. ciliaris</i>	Tanzania	86	PI 409270	<i>C. ciliaris</i>	South Africa	126	PI 409669	<i>C. ciliaris</i>	South Africa
7	PI 161637	<i>C. ciliaris</i>	South Africa	47	PI 365651	<i>C. ciliaris</i>	Tanzania	87	PI 409272	<i>C. ciliaris</i>	South Africa	127	PI 409689	<i>C. ciliaris</i>	South Africa
8	PI 185564	<i>C. ciliaris</i>	South Africa	48	PI 365720	<i>C. ciliaris</i>	Tanzania	88	PI 409273	<i>C. ciliaris</i>	South Africa	128	PI 409692	<i>C. ciliaris</i>	South Africa
9	PI 193445	<i>C. ciliaris</i>	Australia	49	PI 385321	<i>C. ciliaris</i>	Tanzania	89	PI 409274	<i>C. ciliaris</i>	South Africa	129	PI 409704	<i>C. ciliaris</i>	South Africa
10	PI 203363	<i>C. ciliaris</i>	South Africa	50	PI 409174	<i>C. ciliaris</i>	South Africa	90	PI 409280	<i>C. ciliaris</i>	South Africa	130	PI 409718	<i>C. ciliaris</i>	South Africa
11	PI 216374	<i>C. ciliaris</i>	United States	51	PI 409185	<i>C. ciliaris</i>	South Africa	91	PI 409281	<i>C. ciliaris</i>	South Africa	131	PI 409720	<i>C. ciliaris</i>	South Africa
12	PI 225012	<i>C. ciliaris</i>	Ghana	52	PI 409188	<i>C. ciliaris</i>	South Africa	92	PI 409282	<i>C. ciliaris</i>	South Africa	132	PI 414447	<i>C. ciliaris</i>	South Africa
13	PI 225583	<i>C. ciliaris</i>	South Africa	53	PI 409194	<i>C. ciliaris</i>	South Africa	93	PI 409292	<i>C. ciliaris</i>	South Africa	133	PI 414452	<i>C. ciliaris</i>	South Africa
14	PI 226090	<i>C. ciliaris</i>	Kenya	54	PI 409198	<i>C. ciliaris</i>	South Africa	94	PI 409293	<i>C. ciliaris</i>	South Africa	134	PI 414499	<i>C. ciliaris</i>	South Africa
15	PI 243198	<i>C. ciliaris</i>	Zimbabwe	55	PI 409201	<i>C. ciliaris</i>	South Africa	95	PI 409295	<i>C. ciliaris</i>	South Africa	135	PI 414508	<i>C. ciliaris</i>	South Africa
16	PI 243199	<i>C. ciliaris</i>	Zimbabwe	56	PI 409205	<i>C. ciliaris</i>	South Africa	96	PI 409298	<i>C. ciliaris</i>	South Africa	136	PI 414513	<i>C. ciliaris</i>	South Africa
17	PI 245375	<i>C. ciliaris</i>	India	57	PI 409212	<i>C. ciliaris</i>	South Africa	97	PI 409301	<i>C. ciliaris</i>	South Africa	137	PI 414521	<i>C. ciliaris</i>	South Africa
18	PI 253261	<i>C. ciliaris</i>	South Africa	58	PI 409214	<i>C. ciliaris</i>	South Africa	98	PI 409302	<i>C. ciliaris</i>	South Africa	138	PI 414529	<i>C. ciliaris</i>	South Africa
19	PI 253263	<i>C. ciliaris</i>	South Africa	59	PI 409216	<i>C. ciliaris</i>	South Africa	99	PI 409309	<i>C. ciliaris</i>	South Africa	139	PI 442096	<i>C. ciliaris</i>	Japan
20	PI 253269	<i>C. ciliaris</i>	South Africa	60	PI 409217	<i>C. ciliaris</i>	South Africa	100	PI 409327	<i>C. ciliaris</i>	South Africa	140	PI 443507	<i>C. ciliaris</i>	Mexico
21	PI 271198	<i>C. ciliaris</i>	India	61	PI 409221	<i>C. ciliaris</i>	South Africa	101	PI 409348	<i>C. ciliaris</i>	South Africa	141	PI 516516	<i>C. ciliaris</i>	Morocco
22	PI 271203	<i>C. ciliaris</i>	India	62	PI 409222	<i>C. ciliaris</i>	South Africa	102	PI 409352	<i>C. ciliaris</i>	South Africa	142	Grif 1619	<i>C. ciliaris</i>	Australia
23	PI 271206	<i>C. ciliaris</i>	India	63	PI 409223	<i>C. ciliaris</i>	South Africa	103	PI 409361	<i>C. ciliaris</i>	South Africa	143	Grif 1621	<i>C. ciliaris</i>	Kenya
24	PI 271208	<i>C. ciliaris</i>	India	64	PI 409224	<i>C. ciliaris</i>	South Africa	104	PI 409365	<i>C. ciliaris</i>	South Africa	144	Grif 1622	<i>C. ciliaris</i>	Botswana
25	PI 271209	<i>C. ciliaris</i>	India	65	PI 409226	<i>C. ciliaris</i>	South Africa	105	PI 409367	<i>C. ciliaris</i>	South Africa	145	Grif 1637	<i>C. ciliaris</i>	Tanzania
26	PI 271214	<i>C. ciliaris</i>	India	66	PI 409228	<i>C. ciliaris</i>	South Africa	106	PI 409375	<i>C. ciliaris</i>	South Africa	146	Grif 1639	<i>C. ciliaris</i>	Pakistan
27	PI 271219	<i>C. ciliaris</i>	India	67	PI 409229	<i>C. ciliaris</i>	South Africa	107	PI 409379	<i>C. ciliaris</i>	South Africa	147	Grif 1640	<i>C. ciliaris</i>	Australia
28	PI 271593	<i>C. ciliaris</i>	India	68	PI 409230	<i>C. ciliaris</i>	South Africa	108	PI 409381	<i>C. ciliaris</i>	South Africa	148	MAF 74	<i>C. ciliaris</i>	UAE
29	PI 275104	<i>C. ciliaris</i>	India	69	PI 409231	<i>C. ciliaris</i>	South Africa	109	PI 409385	<i>C. ciliaris</i>	South Africa	149	ICBA 49	<i>C. ciliaris</i>	UAE
30	PI 279596	<i>C. ciliaris</i>	Philippines	70	PI 409232	<i>C. ciliaris</i>	South Africa	110	PI 409398	<i>C. ciliaris</i>	South Africa	150	C MAK 1	<i>C. ciliaris</i>	UAE
31	PI 284831	<i>C. ciliaris</i>	South Africa	71	PI 409234	<i>C. ciliaris</i>	South Africa	111	PI 409399	<i>C. ciliaris</i>	South Africa	151	C MAK 2	<i>C. ciliaris</i>	UAE
32	PI 284834	<i>C. ciliaris</i>	Morocco	72	PI 409235	<i>C. ciliaris</i>	South Africa	112	PI 409400	<i>C. ciliaris</i>	South Africa	152	C MAK 3	<i>C. ciliaris</i>	UAE
33	PI 284835	<i>C. ciliaris</i>	India	73	PI 409236	<i>C. ciliaris</i>	South Africa	113	PI 409401	<i>C. ciliaris</i>	South Africa	153	C MAK 4	<i>C. ciliaris</i>	UAE
34	PI 292637	<i>C. ciliaris</i>	Argentina	74	PI 409237	<i>C. ciliaris</i>	South Africa	114	PI 409402	<i>C. ciliaris</i>	South Africa	154	C MAK 5	<i>C. ciliaris</i>	UAE
35	PI 293325	<i>C. ciliaris</i>	Argentina	75	PI 409238	<i>C. ciliaris</i>	South Africa	115	PI 409408	<i>C. ciliaris</i>	South Africa	155	C MAK 6	<i>C. ciliaris</i>	UAE
36	PI 294595	<i>C. ciliaris</i>	Australia	76	PI 409240	<i>C. ciliaris</i>	South Africa	116	PI 409426	<i>C. ciliaris</i>	South Africa	156	C MAK 7	<i>C. ciliaris</i>	UAE
37	PI 295655	<i>C. ciliaris</i>	Zimbabwe	77	PI 409241	<i>C. ciliaris</i>	South Africa	117	PI 409429	<i>C. ciliaris</i>	South Africa	157	C MAK 8	<i>C. ciliaris</i>	UAE
38	PI 295659	<i>C. ciliaris</i>	Zimbabwe	78	PI 409250	<i>C. ciliaris</i>	South Africa	118	PI 409438	<i>C. ciliaris</i>	South Africa	158	Gaydah	<i>C. ciliaris</i>	AUS
39	PI 298980	<i>C. ciliaris</i>	Zimbabwe	79	PI 409254	<i>C. ciliaris</i>	South Africa	119	PI 409442	<i>C. ciliaris</i>	South Africa	159	Bitola	<i>C. ciliaris</i>	AUS
40	PI 299505	<i>C. ciliaris</i>	South Africa	80	PI 409262	<i>C. ciliaris</i>	South Africa	120	PI 409444	<i>C. ciliaris</i>	South Africa	160	Commercial	<i>C. ciliaris</i>	USA





**FIGURE 1 | Screening *Cenchrus ciliaris* for salt tolerance in pots buried in open field at Research Station, ICBA, Dubai, UAE. (A) Germination, (B) growth and tillering, (C) heading stage, and (D) harvesting.**

respectively. All pots were irrigated to upper limit of field capacity before planting.

The soil was collected from an open area of ICBA that had not been previously irrigated with saline water. Textural analysis showed that the resulting soil had 98% sand, 0.5% silt, and 1% clay, which would be classified as sandy soils. The plastic pots were filled with soil mix (20 kg/pot) containing 50% sand, 25% organic fertilizer, and 25% peat moss. The fertilizer NPK was mixed periodically at the start of the trial and after each harvest. The weeds were eradicated by hand weeding.

## Irrigation and Saline Water Treatment Application

The *C. ciliaris* accessions were arranged in the main plot while saline water treatment (0, 10, 15, and 20 dS m<sup>-1</sup>) was randomized in RCBD split plot design. Initially, the pots were irrigated with fresh water up till 1 month to facilitate the germination and seedling establishment. One pot was assigned to each accession per replication and four replications per treatment were maintained throughout the trial. The saline water treatments were applied 1 month after sowing and each salinity treatment

was made by diluting ground water (EC: 25 dS m<sup>-1</sup>) with fresh water (1.5 dS m<sup>-1</sup>) in a separate tank to achieve the target salinities and then delivered to the pots via drip irrigation system. Irrigations were applied on daily basis at rates equivalent to ET<sub>0</sub> plus 20% for leaching requirements. The four salinity levels were maintained constantly throughout the growth period during all the years.

The field experiment was equipped with a drip system (pressure compensating (PC), micro flapper with 4 L hr<sup>-1</sup> flow rate) and 0.5 m distance between rows and 0.25 m between drippers. Each pot has one dripper; the pot area was 0.12 m<sup>2</sup> which means that for each 4 liters the depth would be 33 mm hr<sup>-1</sup>. Irrigation monitoring, scheduling and salinity management were achieved using Decagon<sup>®</sup> sensors. During the summer season (9 months from March–November) the pots were irrigated 30 min per day which means an irrigation depth of 16 mm per pot with the leaching fraction, but during mild winter season (3 months from December–February) the irrigation duration was 20 min which means an irrigation depth of 11 mm per day with the leaching fraction. The drippers were tested biweekly to check the distribution uniformity and coefficient using the method of the low quarter average divided by the average of the

whole readings, the data (not shown) that the distribution uniformity was not <85% with a distribution coefficient of 80%.

## Harvesting and Biomass Yield Measurements

The plants were harvested at heading stage and weighed with an analytical balance. Table 2 demonstrates the harvesting dates

**TABLE 2 | Sequence of harvesting schedule, stage, cut date, and number of cuts in perennial crop.**

Harvesting schedule	Harvesting stage	Cut date	Year	Number of cuts
1st cut	Heading stage	06-05-03	2003	4
2nd cut	//	14-07-03		
3rd cut	//	29-09-03		
4th cut	//	15-12-03		
5th cut	//	01-03-04	2004	4
6th cut	//	15-06-04		
7th cut	//	06-09-04		
8th cut	//	13-12-04		
9th cut	//	08-03-05	2005	4
11th cut	//	16-06-05		
12th cut	//	27-08-05		
13th cut	//	12-12-05		

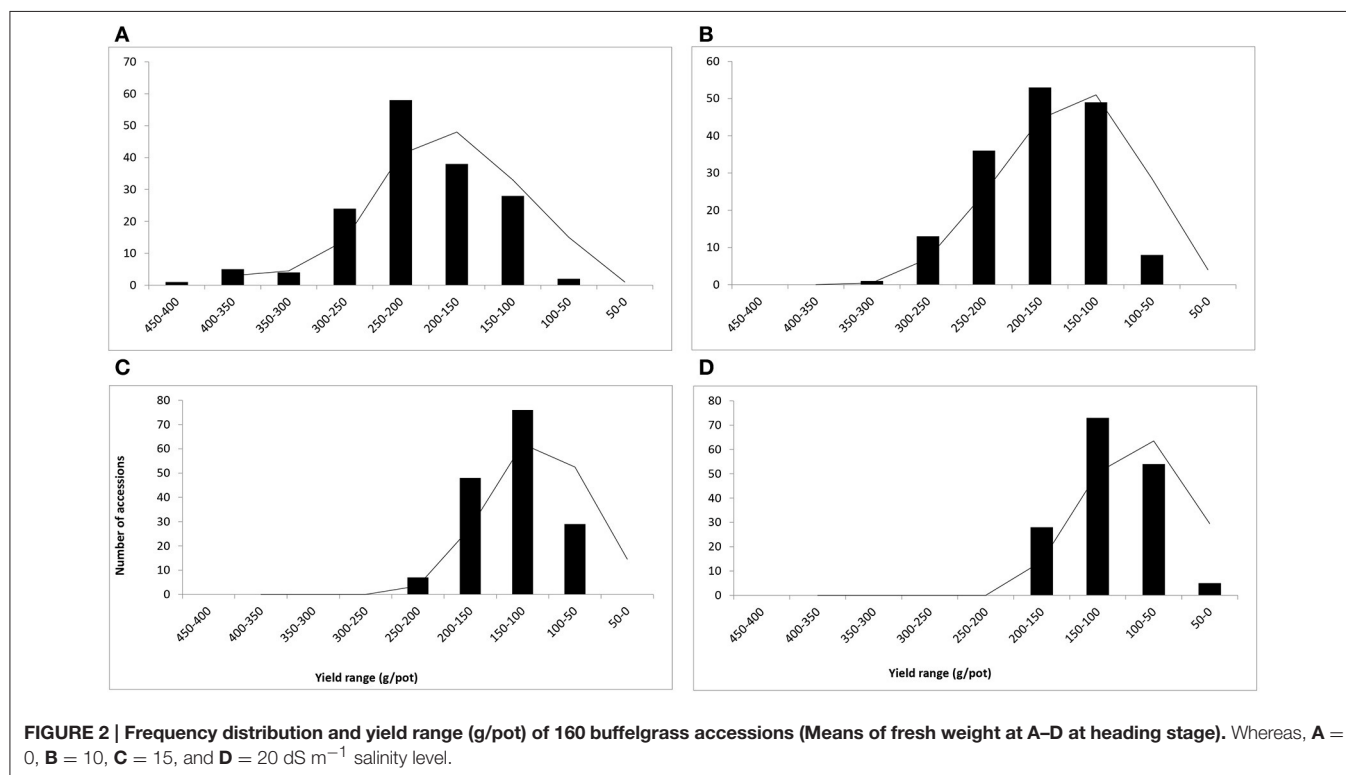
*Cenchrus ciliaris* for salinity tolerance comparison trial for year 2003–2005.

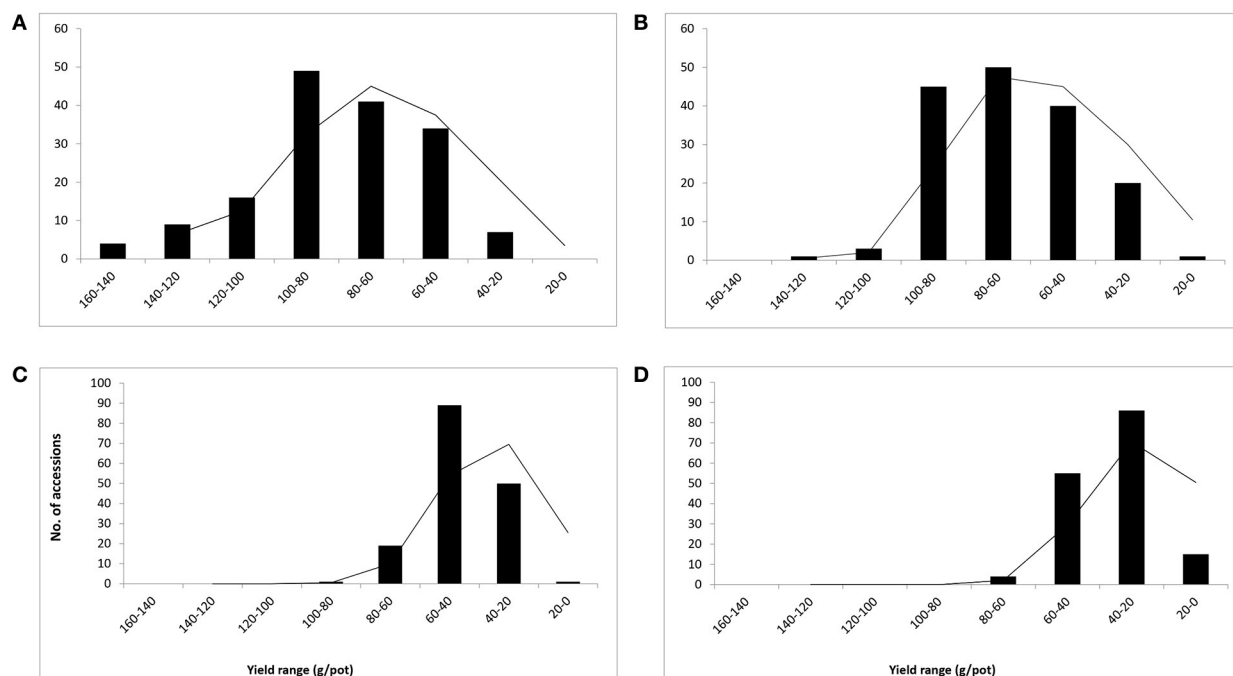
and number of harvests (cuts) obtained during each year. For each harvest, the total fresh weight of the collected biomass samples was weighed in g/pot. The samples were dried in a forced air oven at 60°C for 72 h and dry matter yield (DW) was determined.

## Data and Statistical Analysis

Statistical analyses were performed on fresh biomass (FW) and dry weight (DW) in two steps:

- (1) The experimental data for each cut (fresh weight, dry weight) of 160 genotypes were analyzed using a general linear model with salinity treatment and genotypes as fixed factor and year as a random factor using SPSS for Windows version 17.0 (SPSS Inc., Chicago, IL, USA). Difference between treatments means were compared using Tukey's HSD test.
- (2) Principal component analysis (PCA) was performed on genotypic trait means (FW, DW), to partition the performance of accessions under salt stress conditions. The variate stands for any response variable and "individual" stands for any entry in this analysis. PCA helps the selection of entries, at 25% intensity of selection, based on entries loadings on the first and the second components at each salinity level. The accessions were selected and ability of each entry was determined using its loadings on two components of PCA.
- (3) Hierarchical cluster analysis of 160 buffelgrass accessions using JMP 12.0 software was carried out taking fresh biomass yield and dry weight taking into account means of all years at overall salinity. The between-groups linkage method of clustering was adopted using Euclidean distances.





**FIGURE 3 |** Frequency distribution and yield range (g/pot) of 160 buffelgrass accessions (Mean of dry weight at A–D at heading stage). Whereas, **A** = 0, **B** = 10, **C** = 15, and **D** = 20 dS m<sup>-1</sup> salinity level.

## RESULTS

### Impact of Salinity on Average Biomass Yield of *C. ciliaris*

The salinity treatments significantly influenced the growth and biomass yield of various *C. ciliaris* accessions. The average biomass yield (fresh and dry weight) were calculated by taking the average of 3 years' data for each salinity level (Figures 2, 3). Fresh weight (FW) ranged between 50 and 450 g/pot in control, 50–350 g/pot at low salinity (10 dS m<sup>-1</sup>), 50–300 g/pot at medium high salinity (15 dS m<sup>-1</sup>) and 50–200 g/pot at high salinity (20 dS m<sup>-1</sup>) (Figures 2A–D). The curve was bell shaped in control pots and right skewed distribution when the values clustered more toward lower production groups at high salinity (Figure 2D). Accession 127 proved to be stable and higher yielder that produced the maximum FW (400.8 g/pot) in control and 348.9 g/pot fresh biomass at 10 dS m<sup>-1</sup>. However, at medium and higher salinity (15, 20 dS m<sup>-1</sup>), its biomass yield decreased significantly and produced 180.1 and 122.3 g/pot, respectively. Accession 49 proved to be stable genotype that produced the 2nd highest biomass yield (351.1 g/pot) in the control pots while 292.4 g/pot at 10 dS m<sup>-1</sup>, 219.3 g/pot at 15 dS m<sup>-1</sup>, and 181.7 g/pot at 20 dS m<sup>-1</sup>, respectively. At medium and higher salinity (15 and 20 dS m<sup>-1</sup>), accession 30 produced 233.8 g and 154.2 g/pot FW and demonstrated as medium salt tolerant genotypes.

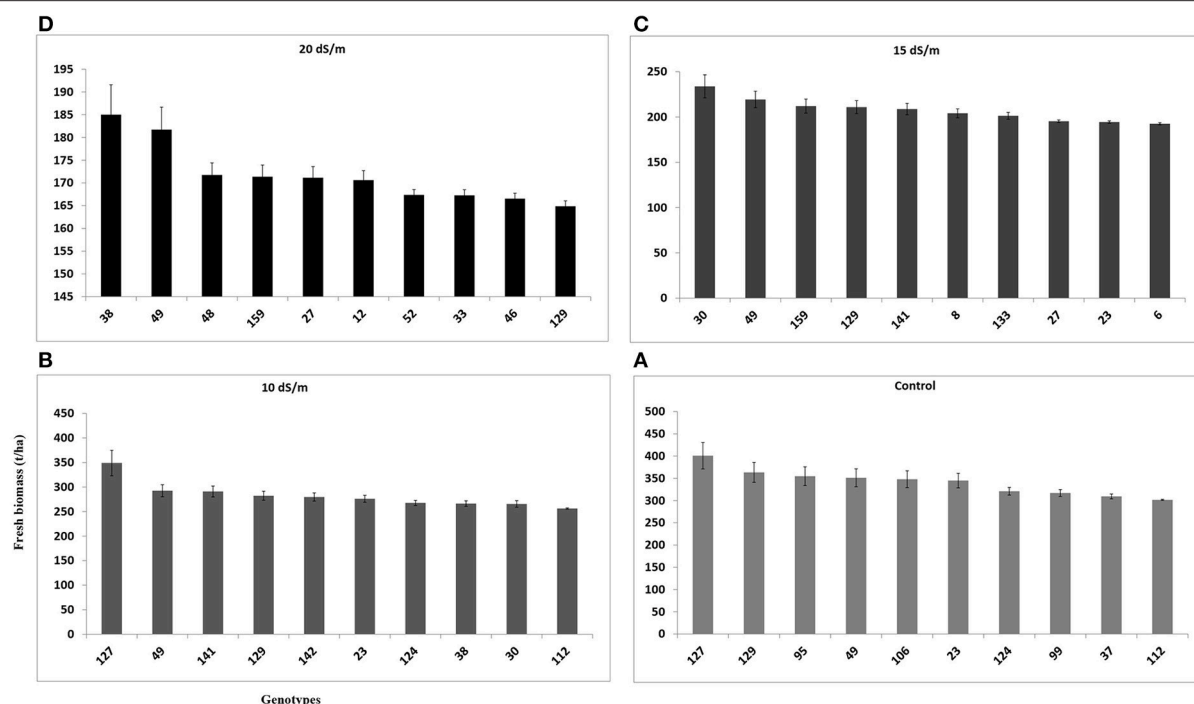
The dry weight range was in the range of 20–160 g/pot in control, 20–140 g/pot at 10 dS m<sup>-1</sup>, 100–20 g/pot at 15 dS m<sup>-1</sup>,

and 20–80 g/pot at 20 dS m<sup>-1</sup> respectively (Figures 3A–D). The curve distribution was bell shaped in control pots and right skewed at high salinity (20 dS m<sup>-1</sup>) (Figure 3D). The range of variation reduced with an increase in salinity levels and ranged between 20 and 160 g/pot when salinity varies from 0 to 20 dS m<sup>-1</sup>. The reduction in dry biomass was 87.5, 85.7, 80, and 75% at 10, 15, and 20 dS m<sup>-1</sup>, respectively (Figures 3A–D). The average mean DW was 148.9 g/pot recorded for accession 49 in control while accession 127 produced the highest dry biomass (141.4 g/pot), in control pots, respectively (data not shown). The accession 124 produced the maximum DW (80.3 g/pot) at 20 dS m<sup>-1</sup> salinity and proved to be stable and higher yield genotype. At highest salinity (20 dS m<sup>-1</sup>), accession 49 produced the maximum dry biomass (65.6 g/pot) (data not shown).

### Effect of Salinity on Total Annual Yield of *C. ciliaris*

Fresh biomass (total annual yield) ranged from 301.6 to 400.8 g/pot in control, 256.2–348.9 g/pot at 10 dS m<sup>-1</sup>, 192.5–233.8 g/pot at 15 dS m<sup>-1</sup>, and 164.8–185.0 g/pot at 20 dS m<sup>-1</sup> among the top 10 accessions of *C. ciliaris* (Figure 4). The results illustrated that FW in 112, 37, 30, 23, and 46 accessions were more affected by the salinity at 10, 15, and 20 dS m<sup>-1</sup>, respectively. The accessions 49, 159, 129, 30, 38, were the least affected by increasing the salinity level from 0 to 20 dS m<sup>-1</sup>. However, accession 129 produced good fresh biomass yield in control, 10 dS m<sup>-1</sup>, and 15 dS m<sup>-1</sup> salinity while it was not a stable genotype at higher salinity (20 dS m<sup>-1</sup>) that produced less FW as compared





**FIGURE 4 | Effect of different irrigation water salinity levels on fresh biomass yield (total annual yield) of TOP 10 *Cenchrus ciliaris* genotypes.** Error bars represent standard deviations of the mean of four replicates. Whereas, (A) = 0, (B) = 10, (C) = 15, and (D) = 20 dS m<sup>-1</sup> salinity level.

to other top 10 genotypes (Figure 4). The salinity tolerance values in DW from all buffelgrass accessions show a high degree of variability under the various salinity levels. Average annual DW yield ranged from 122.5 to 148.9 g/pot in control; 96.4–133.7 g/pot at 10 dS m<sup>-1</sup>; from 65.6 to 80.3 g/pot at 15 dS m<sup>-1</sup> and 55.4–65.6 g/pot at higher salinity (20 dS m<sup>-1</sup>). The highest DW (148.9 g/pot) was produced by accession 49 in control and the lowest one (55.4 g/pot) was recorded in accession 23 following treatment at the highest salinity level (20 dS m<sup>-1</sup>) (Figure 5).

## Evaluation of Performance of *C. ciliaris* Accessions through Multivariate Analysis

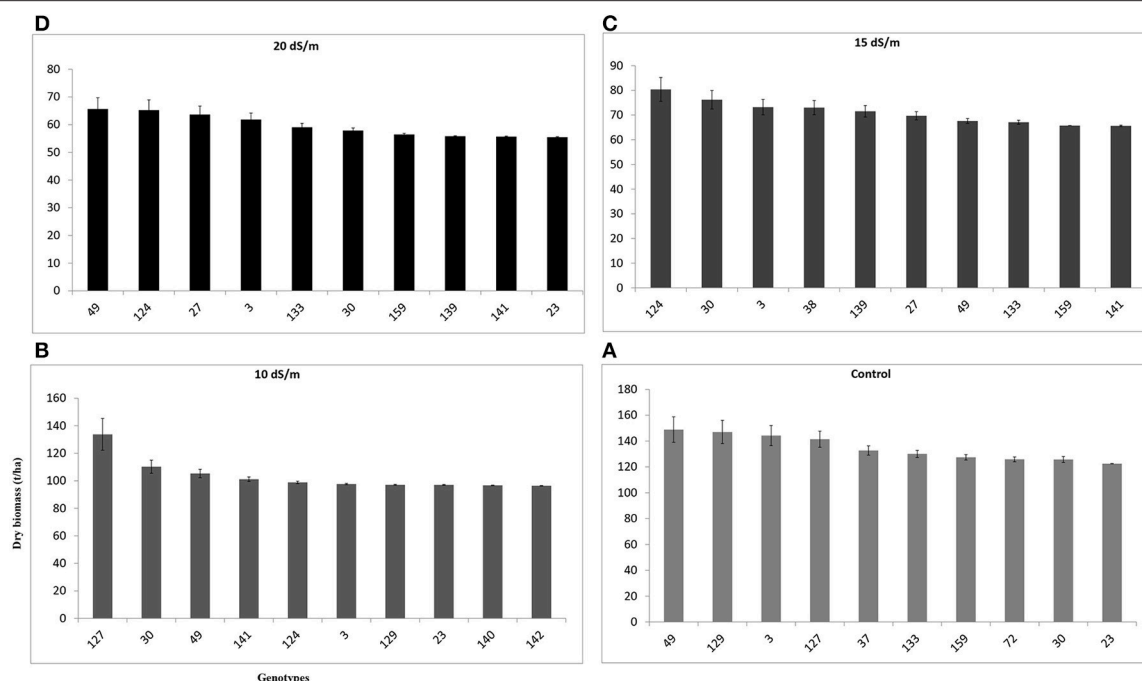
### Principle Component Analysis

The results of principal components analysis show that out of four components, only 1 had extracted Eigen values over 1. This is based on Chatfield and Collin (1980) assumption which stated that components with an Eigen value of <1 should be eliminated. The extracted component was subsequently rotated according to Varimax rotation in order to make interpretation easier and fundamental significance of extracted component to the irrigation water salinity. Principle component PC 1 was extracted having Eigen value >1 and contributed 83.5% of the total variability among the *C. ciliaris* genotypes assessed for fresh biomass (Figure 6). However, the PC 2 depicted 10.7% of the total variation for the same parameter. Furthermore, PC1 contributed 81.8% of the total variability among the 160 genotypes and PC2 only depicted 11.7% of the total variation for dry weight (DW). Salinity applies a gradual selection pressure

on the entries that was higher for dry biomass than for fresh biomass. PCA of *C. ciliaris* genotypes revealed diverse grouping pattern. The separation on the basis of PC1 and PC2 revealed that the genotypes were scattered in all the quarters, which show the high level of genotypic variation among the accessions (Figure 6). The first principal component correlated with five of the original variables (DWS2, FW2, FWS3, DWS1, and FWS4) and more strongly with the DWS2. In fact, we could state that based on this correlation, principal component is primarily a measure of the DWS2 (data not shown). In the first principal components, accessions 49, 22, 16, 42, 8, 149, showed more genotypic variation based on FW and are located close to axis line. In the second principal components, accessions 49, 133, 23, 3, 30, 6, showed more genotypic variation based on dry weight (DW) trait and are also located close to axis line.

### Hierarchical Cluster Analysis

Germplasm evaluation for quantitative traits may help to work out the relative importance of various traits within each cluster. Hierarchical cluster analysis based on agronomical traits (fresh and dry biomass) divided the 160 accessions into 6 main clusters. Maximum number of accessions (42) were present in group I, followed group II (27), group III (26), group IV (19), group V (17), group VI (17), and group VII (12) (Figure 7). Hierarchical cluster analysis showed that some of accessions originated from various geographical areas were grouped into the same cluster, while many others fell into different clusters. Group VII consisted of highly salt tolerant genotypes (3, 133, 159, 30, 23, 142, 141, 95, 49, 129, 124,



**FIGURE 5 | Effect of different irrigation water salinity levels on dry weight (total annual dry biomass) of TOP 10 *Cenchrus ciliaris* genotypes.** Error bars represent standard deviations of the mean of four replicates. Whereas, (A) = 0, (B) = 10, (C) = 15, and (D) = 20 dS m<sup>-1</sup> salinity level.

and 127) that acquired maximum biomass yield at all target salinities and also among the top 10 accessions. Grouping pattern indicated that the clusters were heterogeneous with regard to the geographical origin and some entries from different geographic regions were pooled in the same cluster (Figure 7).

### Performance of Commercial Cultivars of *C. ciliaris*

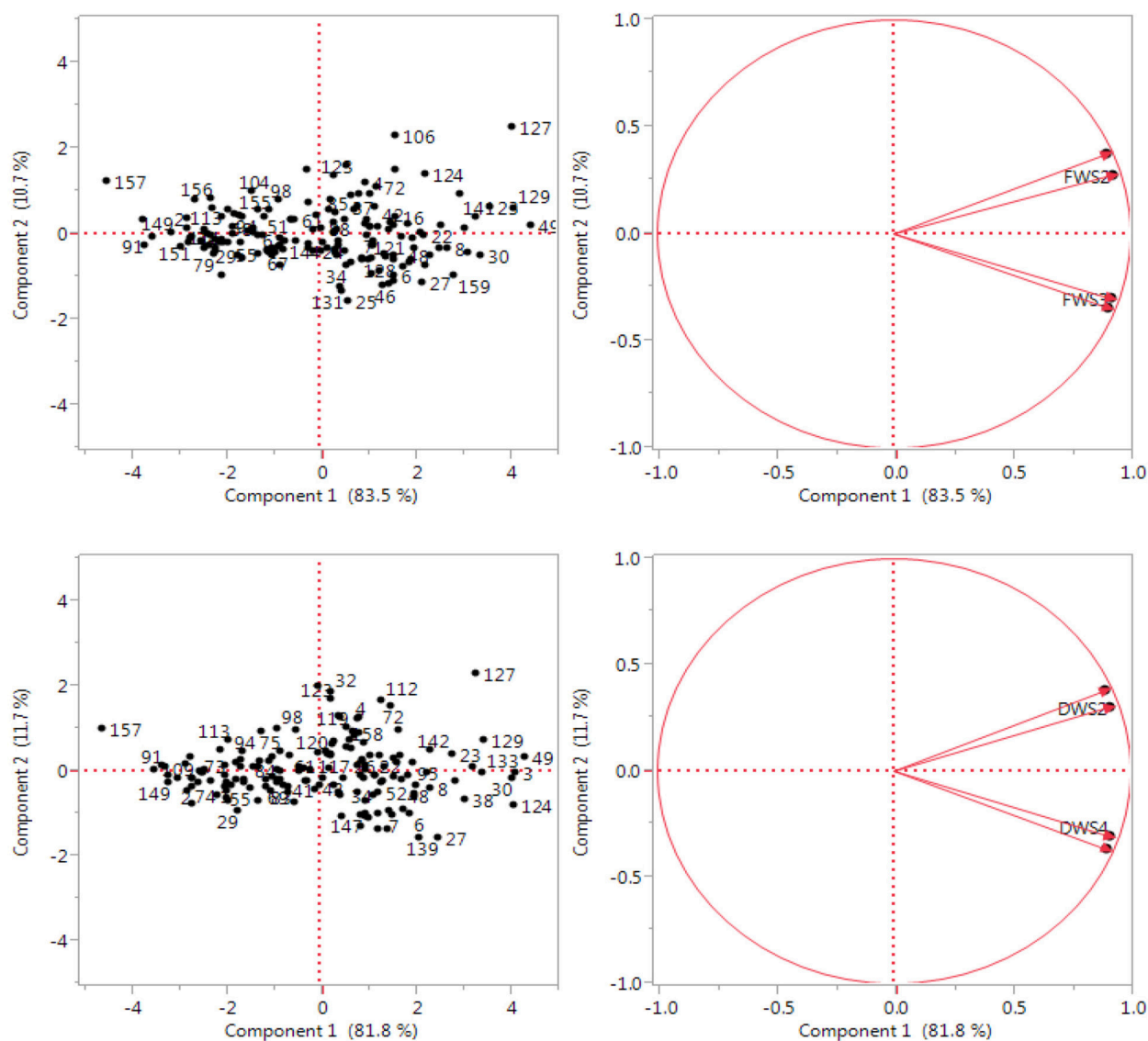
The observations of present study were largely consistent with the available descriptions of commercial cultivars. The genotype 159 (Biloela) was salt tolerant, productive and produced higher biomass than other genotypes at different water salinities (0, 10, 15, 20 dS m<sup>-1</sup>). Other commercial genotypes 158 (Gayndah) and 160 did not show the good performance in terms of fresh and dry biomass and were not stable at medium and high salinities (data not shown).

## DISCUSSION

Yield assessment conducted in field trials provides the most reliable information on salt tolerance; however, spatial and temporal variability in soil salinity (Richards, 1983) make it difficult to obtain reproducible information. The advantages of using pots buried in open field are that the saline conditions can be controlled and constantly monitored. Salt affects the growth of plants by limiting the absorption of water and essential nutrients through roots. Salt stress has an immediate effect on cell growth and enlargement, and high concentrations of salts can be extremely toxic (Munns and Tester, 2008; Hussain et al., 2016).

In the present study, there was wide variation among *C. ciliaris* genotypes grown under different salinity levels regarding their biomass yield. The fresh and dry weight of all genotypes was decreased with increase in salt stress. However, salt-sensitive genotypes showed more reduction in their biomass as compared to tolerant genotypes. Lanza Castelli et al. (2010) demonstrated a significant reduction (50–70%) in height and shoot fresh weight at 30 dS m<sup>-1</sup> in Texas and Biloela accessions. In the present study, the accessions 49, 127, 129, 30, 38, were the least affected by increasing salinity. The accession 129 produced good biomass at low salinity; however, it was not a stable genotype at higher salinity. Therefore, it is quite suitable for low and medium saline areas where water salinity falls in the range 5–15 dS m<sup>-1</sup>, respectively. Our results demonstrate that accessions from Africa were more stable, salt tolerant and attained high biomass yield as compared to accessions from other regions. Previously, it was reported that, *C. ciliaris* is a dominant grass along the sandy beaches of Eritrea on soils with salt content of 0.006–0.251% (Hemming, 1961). Skerman and Riveros (1990), demonstrated that it grows well in saline soils; although it is less salt tolerant than *Chloris gayana*, *Cynodon* spp., and *Panicum antidotale*.

The DW from all *C. ciliaris* accessions shows a high variability under the various salinity levels. The highest DW was produced by accession 49 and the lowest by accession 23. Similar studies were conducted on other grass species such as Bermuda grass (Rodríguez and Miller, 2000), *Catharanthus roseus* (Jaleel et al., 2008) and *C. ciliaris* (Lanza Castelli et al., 2010) and showed that the plant growth, tiller, number of leaves and dry



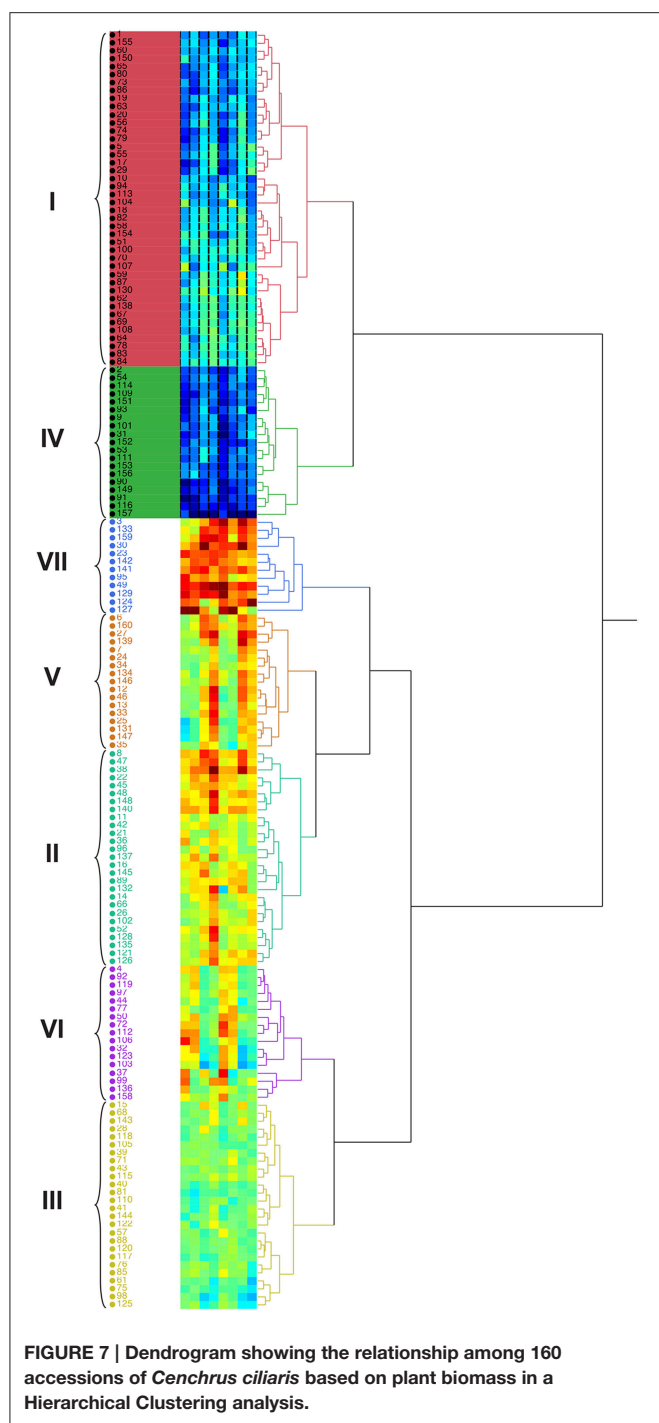
**FIGURE 6 | Principal component analysis of the variables: Fresh weight (FW) and dry weight (DW) at four water salinity levels (0, 10, 15, 20 dS m<sup>-1</sup>).** Two dimensional scatter plot of genotypic relationship among 160 *Cenchrus ciliaris* accessions as revealed by first two principal components.

weight decreased under salt stress. Ions of sodium (Na<sup>+</sup>) and chlorine (Cl<sup>-</sup>) (>40 mmol/L) can be toxic to plants because they can create imbalance in plant nutrition due to decreased nutrient uptake and translocation to new shoots (Munns et al., 2000; Tester and Davenport, 2003; Munns and Tester, 2008). Nawazish et al. (2006) found that shoot dry weight was severely affected in the *C. ciliaris* ecotype from Faisalabad, where it decreased from 29.83 to 8.02 g/plant. High salinity modifies plant metabolisms, which results in altered plant morphology; cultivar type, duration, and intensity of stress determine the extent of morphological modification. These results indicate that dry biomass yield can be used as a screening and selection criteria for evaluating the salt tolerance behavior among a large collection of plant accessions. Based on their relative performance (DW) there

is a chance of selecting the genotypes within each salinity level. Several accessions from the top 10 selection with high biomass yield potential are of particular interest as forage resource in salt-affected agro-ecosystems for livestock and are highly suitable for arid and semi-arid hot dry regions.

From present results, we found that, PCA grouped accessions together with more morphological similarities; the clusters did not necessarily include all the accessions from the same or nearby sites. Diversity of populations within a geographical origin and similarity of populations beyond geographical limits have also been reported in *C. ciliaris* genotypes (Jorge et al., 2008). Among the good performed accessions, genotypes 127, 129, 49, 23, 27, 16, and 42 attained higher DW. It was also found that variables (genotypes) close to an axis correlate more with that principal





component; one may consider that axis is a combination of its neighboring variables. Divergence studies of morphological and agronomical traits using principal component and cluster analysis have also been reported by other researchers (Warwick et al., 2006; Dawood et al., 2009) which were in support of present investigation that both cluster and PCA disclosed complex relationship among the accessions in a more understandable way. The variability among the accessions from diverse origin could be related primarily to their morphological differences

and secondly to agronomic use. Accessions evaluated in this work exhibited a reasonable level of diversity for some of the studied traits (FW, DW) of economic significance providing a resource for future crop improvement. For the improvement of *C. ciliaris*, it would be desirable to use diverse accessions with more variability for use in future breeding programs. For use in natural resource management and for soil stabilization, *C. ciliaris* accessions which are shorter, more prostrate and more rhizomatous should be selected, because they would provide better ground cover (IBPGR, 1984).

A cluster analysis was used in this study to facilitate the evaluation of salt tolerance among different genotypes. The major advantage of the utilization of a multivariate analysis in the evaluation for salt tolerance is the allowance of a simultaneous analysis on multiple parameters and increase in the accuracy in the rankings of genotypes. Another advantage is the convenience to rank genotypes when plants are evaluated at different salt levels, e.g., moderate and high salt levels. Cluster analysis proved their validity to establish genotypic diversity within accessions. These findings are supported by Iqbal et al. (2010) who concluded that quantitative characters revealed more reliability than biochemical markers that are specific for particular conditions. Germplasm evaluation for quantitative traits may help to work out the relative importance of various traits within each cluster. Accessions from different geographical regions were pooled in the same cluster. Our results coincide with the classification of 68 populations of *C. ciliaris* where most of the populations were grouped mainly due to their agronomical attributes and not related to their geographic origin (Jorge et al., 2008). Group VII consists of highly salt tolerant genotypes that acquired maximum biomass yield at all target salinities and were also among the top 10 accessions. Grouping was not associated with the geographic distribution; instead accessions were mainly grouped based on their morphological traits (FW, DW). Thus, it cannot be generalized that all the accessions having same origin would always have low diversity among them. For example, in Australia, where *C. ciliaris* is grown extensively, rainfall can occur during the cooler months and most cultivars of *C. ciliaris* respond poorly to these rains owing to the cold temperatures. Accessions originating from cool, dry environments could be promising for these areas through improved performance in spring, owing to a better response to rains during the cooler seasons (Hacker and Waite, 2001). Divergence studies of morphological and agronomical attributes using principal component and cluster analysis have been made by different researchers (Dawood et al., 2009).

A good starting point in selecting suitable accessions for dry areas is to test accessions with good agro-morphological attributes originated from areas with very low rainfall. Several accessions of potential interest were collected from arid areas of South Africa, Zimbabwe, Tanzania and hot arid environments of UAE and Kenya. Salinity tolerance is generally an important parameter when choosing accessions to evaluate, especially within a country, because of the strong correlation between genotype and environment ( $G \times E$ ). A total of 100 accessions, originating from South Africa are found in four of the five clusters, showing that there is not a strong link between

environment and morphology, confirming the findings in a study on dry matter and phenology of *C. ciliaris* in Tunisia (Mseddi et al., 2004). Some other studies also classified American and Gayndah in the same group (Pengelly et al., 1992).

The present study was largely consistent with the available descriptions of commercial cultivars. One main reason for discrepancies among experimental findings could be a strong influence of the environment. Within-cultivar high variation in *C. ciliaris* is a well-established fact (Jorge et al., 2008) that may be one of the reasons why commercial cultivars are not consistent in their performance. They further concluded that accessions belonging to identical agro-morphological groups were found from a wide range of environments in sub-Saharan Africa. Inherent genetic variability present in perennial grasses collected from different habitats of Cholistan desert were evaluated and high amount of genotypic variations were recorded among the accessions (Arshad et al., 2000). A genotypic characterization of the collection could provide further evidence to confirm the classification of this study.

## CONCLUSIONS

We found significant variation for salt tolerance among a large collection of buffelgrass genotypes representing global diversity, USDA collection, commercial varieties and local landraces. We identified several accessions that were stable and productive at high salinity and other accessions whom were stable but low yielder. Accession 3, 133, 159, 30, 23, 142, 141, 95, 49, 129, 124, and 127 were stable, salt tolerant and produced good dry biomass yield at target salinities during all years. These

accessions can be exploited to widen the genetic base of existing *C. ciliaris* accessions against salt tolerance. These accessions hold an appropriate salinity tolerance potential and can be grown to enhance farm productivity in arid and semi-arid areas of the globe. Furthermore; introduction of salt-tolerant perennial species is a promising alternatives to overcome salinity problems in the arid regions. The conservation of fresh water resources and its use for high value purposes, while using low quality saline water can provide both ecological and economic benefits, essential for sustainable agriculture in dry lands.

## AUTHOR CONTRIBUTIONS

MH conducted the experiment, collected data, and drafted the manuscript. AA designed the experiment, followed upon data collection, provided support through the PCA statistical analysis and edited the manuscript.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge the International Fund for Agricultural Development (IFAD), Arab Fund for Economic and Social Development (AFESD), and the Islamic Development Bank (IDB) for their financial support through several regional projects. We are highly grateful to Dr. Susan Robertson (ICBA, Dubai, UAE) for help in English language and grammar correction. We are thankful to Dr. Asad Qureshi (ICBA, Dubai, UAE) for help in critical revision of the article. We are extremely thankful to anonymous reviewers for the critical and valuable comments that have greatly improved the article.

## REFERENCES

- Ahmad, F. (2010). *Leptochloa Fusca* cultivation for utilization of salt-affected soil and water resources in Cholistan desert. *Rev. Soc. Nat.* 22, 141–149. doi: 10.1590/S1982-45132010000100010
- Angassa, A., and Baars, R. M. T. (2000). Ecological condition of encroached and non-encroached rangelands in Borana, Ethiopia. *Afr. J. Ecol.* 38, 321–328. doi: 10.1046/j.1365-2028.2000.00250.x
- Arshad, M., Rao, A., Akbar, G., and Akhtar, M. (2000). Eco-physiological adaptability of *Sporobolus iocladius*, from Cholistan, Pakistan. *Pak. J. Biol. Sci.* 2, 1604–1607. doi: 10.3923/pjbs.1999.1604.1607
- Ayerza, R. (1981). *Buffelgrass: Utilisation and Productivity of a Promising Grass*. Buenos Aires: Hemisferio Sur.
- Buldgen, A., and Francois, J. (1998). Physiological reactions to imposed water deficit by *Andropogon gayanus* cv. Bisquamulatus and *Cenchrus ciliaris* cv. Biloela in a mixed fodder crop. *J. Agric. Sci.* 131, 31–38. doi: 10.1017/S0021859698005462
- Chatfield, C., and Collin, A. J. (1980). *Introduction to Multivariate Analysis*. New York, NY: Chapman and Hall in Association with Methuen.
- Daniells, I. G., Holland, J. F., Young, R. R., Alston, C. L., and Bernardi, A. L. (2001). Relationship between yield of grain sorghum (*Sorghum bicolor*) and soil salinity under field conditions. *Aust. J. Exp. Agric.* 41, 211–217. doi: 10.1071/EA00084
- Dawood, A., Kikuchi, A., Jatoti, S. A., Kiuchi, F., and Watanabe, K. N. (2009). Genetic variation in the chloroplast DNA (cpDNA) of *Zingiberaceae* taxa from Myanmar envisaged by the PCR-RFLP analysis. *Ann. Appl. Biol.* 155, 91–101. doi: 10.1111/j.1744-7348.2009.00322.x
- Food and Agriculture Organization of the United Nations (2012). *FAO Statistical Yearbook 2012: World Food and Agriculture*. Available online at: <http://reliefweb.int/sites/reliefweb.int/files/resources/i2490e.pdf> (Accessed March 12, 2014).
- Fernando, A. L., Boléo, S., Barbosa, B., Costa, J., Sidella, S., Nocentini, A., et al. (2012). “Perennial grasses: environmental benefits and constraints of its cultivation in Europe,” in *20th European Biomass Conference and Exhibition, 18–22 June 2012* (Milano), 2092–2094.
- Graham, T. W., and Humphreys, L. R. (1970). Salinity responses of cultivars of Buffelgrass (*Cenchrus ciliaris* L.). *Aust. J. Exp. Agric. Anim. Husb.* 10, 725–728.
- Hacker, J. B., and Waite, R. B. (2001). Selecting buffelgrass (*Cenchrus ciliaris*) with improved spring yield in subtropical Australia. *Trop. Grassl.* 35, 205–210.
- Hemming, D. F. (1961). The ecology of the coastal area of northern Eritrea. *J. Ecol.* 49, 55–78. doi: 10.2307/2257423
- Hussain, M. I., Lyra, D. A., Farooq, M., Nikoloudakis, N., and Khalid, N. (2016). Salt and drought stresses in Safflower: a review. *Agron. Sustain. Dev.* 36, 4. doi: 10.1007/s13593-015-0344-8
- IBPGR (1984). *Forage and Browse Plants for Arid and Semiarid Africa*. Kew: IBPGR/Royal Botanic Gardens.
- Iqbal, M. S., Ghafoor, A., and Qureshi, A. S. (2010). Evaluation of *Nigella sativa* L. for genetic variation and *ex-situ* conservation. *Pak. J. Bot.* 42, 2489–2495.
- Jaleel, C. A., Beemaroo, S., Ramalingam, S., and Rajaram, P. (2008). Soil salinity alters growth, chlorophyll content and secondary metabolite accumulation in *Catharanthus roseus*. *Turk. J. Biol.* 32, 79–83.
- Jorge, M. A. B., Van DE Wouw, M., Hanson, J., and Mohammed, J. (2008). Characterization of a collection of buffelgrass (*Cenchrus ciliaris*). *Trop. Grassl.* 42, 27–39.

- Krishnamurthy, L., Serraj, R., Hash, C. T., Al-Dakheel, A. J., and Reddy, B. V. S. (2007). Screening sorghum genotypes for salinity tolerant biomass production. *Euphytica* 156, 15–24. doi: 10.1007/s10681-006-9343-9
- Ksiksi, T. S., and El-Shaigy, N. O. (2012). Growth responses of nutrient-stressed *Cenchrus ciliaris* under carbon dioxide enrichment. *J. Earth Sci. Clim. Change* 3:127. doi: 10.4172/2157-7617.1000127
- Lanza Castelli, S., Grunberg, K., Munoz, N., Griffa, S., Colomba, E. L., Ribotta, A., et al. (2010). Oxidative damage and antioxidants defenses as potential indicators of salt-tolerant *Cenchrus ciliaris* L. genotypes. *Flora* 205, 622–626. doi: 10.1016/j.flora.2010.04.004
- Mseddi, K., Mnif, L., Chaied, M., Neffati, M., and Roux, M. (2004). Aboveground phytomass productivity and morphological variability of Tunisian accessions of *Cenchrus ciliaris* L. *Afr. J. Range Forage Sci.* 21, 49–55. doi: 10.2989/10220110409485833
- Munns, R., Guo, J., Passioura, J. B., and Cramer, G. R. (2000). Leaf water status controls daytime but not daily rates of leaf expansion in salt-treated barley. *Aust. J. Plant Physiol.* 27, 949–957. doi: 10.1071/PP99193
- Munns, R., and Tester, M. (2008). Mechanisms of salinity tolerance. *Ann. Rev. Plant Biol.* 59, 651–681. doi: 10.1146/annurev.arplant.59.032607.092911
- Nawazish, S., Hameed, M., and Naurin, S. (2006). Leaf anatomical adaptations of *Cenchrus ciliaris* L. from the Salt Range, Pakistan against drought stress. *Pak. J. Bot.* 38, 1723–1730.
- Peng, Y. L., Gao, Z. W., Gao, Y., Liu, G. F., Sheng, L. X., and Wang, D. L. (2008). Ecophysiological characteristics of alfalfa seedlings in response to various mixed salt alkaline stresses. *J. Integr. Plant Biol.* 50, 29–39. doi: 10.1111/j.1744-7909.2007.00607.x
- Pengelly, B. C., Hacker, J. B., and Eagles, D. A. (1992). The classification of a collection of buffelgrasses and related species. *Trop. Grassl.* 26, 1–6.
- Ponsens, J., Hanson, J., Schellberg, J., and Moeseler, B. M. (2010). Characterization of phenotypic diversity, yield and response to drought stress in a collection of Rhodes grass (*Chloris gayana* Kunth) accessions. *Field Crops Res.* 118, 57–72. doi: 10.1016/j.fcr.2010.04.008
- Qadir, M., Quillerou, E., Nangia, V., Murtaza, G., Singh, M., Thomas, R. J., et al. (2014). Economics of salt-induced land degradation and restoration. *Nat. Res. Forum* 38, 282–295. doi: 10.1111/1477-8947.12054
- Richards, R. (1983). Should selection for yield in saline regions be made on saline or non-saline soils? *Euphytica* 32, 431–438. doi: 10.1007/BF00021452
- Rodriguez, I. R., and Miller, G. L. (2000). Using a chlorophyll meter to determine the chlorophyll concentration, nitrogen concentration and visual quality of St. Augustine grass. *Hortscience* 35, 751–754.
- Roy, S. J., Negrao, S., and Tester, M. (2014). Salt resistant crop plants. *Curr. Opin. Biotechnol.* 26, 115–124. doi: 10.1016/j.copbio.2013.12.004
- Shannon, M. C. (1997). Adaptation of plants to salinity. *Adv. Agron.* 60, 75–120. doi: 10.1016/S0065-2113(08)60601-X
- Skerman, P. J., and Riveros, F. (1990). *Tropical Grasses*. FAO Plant Production and Protection Series No. 23. Rome: FAO.
- Tester, M., and Davenport, R. J. (2003). Na<sup>+</sup> transport and Na<sup>+</sup> tolerance in higher plants. *Ann. Bot.* 91, 503–527. doi: 10.1093/aob/mcg058
- Wang, Q., Wu, C., Xie, B., Liu, Y., Cui, J., Chen, G., et al. (2012). Model analysing the antioxidant responses of leaves and roots of switchgrass to NaCl-salinity stress. *Plant Physiol. Biochem.* 58, 288–296. doi: 10.1016/j.plaphy.2012.06.021
- Warwick, S. I., Gugel, R. K., McDonald, T., and Falk, K. C. (2006). Genetic variation of Ethiopian mustard (*Brassica carinata*) germplasm in western Canada. *Genet. Res. Crop Evol.* 53, 297–312. doi: 10.1007/s10722-004-6108-y
- Zhang, Y., Li, Y., Jiang, L., Tian, C., Li, J., and Xiao, Z. (2011). Potential of perennial crop on environmental sustainability of agriculture. *Procedia Environ. Sci.* 10, 1141–1147. doi: 10.1016/j.proenv.2011.09.182

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Al-Dakheel and Hussain. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Comparative Analysis of *WRKY* Genes Potentially Involved in Salt Stress Responses in *Triticum turgidum* L. ssp. *durum*

Fatma-Ezzahra Yousfi<sup>1,2,3</sup>, Emna Makhoulfi<sup>1,2,3,4</sup>, William Marande<sup>2</sup>, Abdel W. Ghorbel<sup>1</sup>, Mondher Bouzayen<sup>3,4</sup> and Hélène Bergès<sup>2\*</sup>

<sup>1</sup> Laboratory of Plant Molecular Physiology, Center of Biotechnology of Borj Cedria, Borj Cedria Science and Technology Park, Hammam-If, Tunisia, <sup>2</sup> Centre National de Ressources Genomiques Vegetales, French Plant Genomic Center, INRA-CNRGV, Castanet-Tolosan, France, <sup>3</sup> INRA, UMR990 Genomique et Biotechnologie des Fruits, Castanet-Tolosan, France, <sup>4</sup> INPT, Laboratoire de Genomique et Biotechnologie des Fruits, University of Toulouse, Castanet-Tolosan, France

## OPEN ACCESS

### Edited by:

Teresa Donze,  
West Chester University, USA

### Reviewed by:

Hao Wang,  
University of Georgia, USA  
Dong-Ha Oh,  
Louisiana State University, USA  
Zoe Joly-Lopez,  
New York University, USA

### \*Correspondence:

Hélène Bergès  
helene.berges@inra.fr

### Specialty section:

This article was submitted to  
Crop Science and Horticulture,  
a section of the journal  
Frontiers in Plant Science

**Received:** 29 June 2016

**Accepted:** 20 December 2016

**Published:** 31 January 2017

### Citation:

Yousfi F-E, Makhoulfi E, Marande W,  
Ghorbel AW, Bouzayen M and  
Bergès H (2017) Comparative  
Analysis of *WRKY* Genes Potentially  
Involved in Salt Stress Responses in  
*Triticum turgidum* L. ssp. *durum*.  
Front. Plant Sci. 7:2034.  
doi: 10.3389/fpls.2016.02034

*WRKY* transcription factors are involved in multiple aspects of plant growth, development and responses to biotic stresses. Although they have been found to play roles in regulating plant responses to environmental stresses, these roles still need to be explored, especially those pertaining to crops. Durum wheat is the second most widely produced cereal in the world. Complex, large and unsequenced genomes, in addition to a lack of genomic resources, hinder the molecular characterization of tolerance mechanisms. This paper describes the isolation and characterization of five *TdWRKY* genes from durum wheat (*Triticum turgidum* L. ssp. *durum*). A PCR-based screening of a *T. turgidum* BAC genomic library using primers within the conserved region of *WRKY* genes resulted in the isolation of five BAC clones. Following sequencing fully the five BACs, fine annotation through Triannot pipeline revealed 74.6% of the entire sequences as transposable elements and a 3.2% gene content with genes organized as islands within oceans of TEs. Each BAC clone harbored a *TdWRKY* gene. The study showed a very extensive conservation of genomic structure between *TdWRKYs* and their orthologs from Brachypodium, barley, and *T. aestivum*. The structural features of *TdWRKY* proteins suggested that they are novel members of the *WRKY* family in durum wheat. *TdWRKY1/2/4*, *TdWRKY3*, and *TdWRKY5* belong to the group Ia, IIa, and IIc, respectively. Enrichment of *cis*-regulatory elements related to stress responses in the promoters of some *TdWRKY* genes indicated their potential roles in mediating plant responses to a wide variety of environmental stresses. *TdWRKY* genes displayed different expression patterns in response to salt stress that distinguishes two durum wheat genotypes with contrasting salt stress tolerance phenotypes. *TdWRKY* genes tended to react earlier with a down-regulation in sensitive genotype leaves and with an up-regulation in tolerant genotype leaves. The *TdWRKY* transcripts levels in roots increased in tolerant genotype compared to sensitive genotype. The present results indicate that these genes might play some functional role in the salt tolerance in durum wheat.

**Keywords:** BAC sequencing, *cis*-regulatory elements, *WRKY*, transposable elements, salt stress, *Triticum durum*, wheat

## INTRODUCTION

Durum wheat (*Triticum turgidum* ssp. *durum*) is a monocot of the *Poaceae* family, of the *Triticeae* tribe and the *Triticum* genus. Tetraploid *Triticum durum* ( $2n = 4 \times = 28$ , AABB), originated when two diploid wild grasses were crossed. The A genome originated from *T. urartu* (Dvorak, 1988). The origin of the B genome is still under discussion, but so far, *Ae. Speltoides* has been put forward as the closest to the donor of this genome (Fernandez-Calvin and Orellana, 1994). Durum wheat has one of the largest and most complex genomes; its size is estimated at 13,000 Mb.

The unavailability of *T. durum* genome sequences complicates and hinders the identification of the genetic factors and mechanisms behind responses to abiotic stresses such as drought and salinity. These are just some of the major constraints affecting cereal crops. In terms of commercial production and human food, this species is the second largest of its kind after bread wheat (*Triticum aestivum* L.). Durum wheat represents 8% of total wheat production, but 80% grow under Mediterranean climates (Monneveux et al., 2000). To date, durum wheat is the only tetraploid wheat species of commercial importance that is widely cultivated in these regions, where drought, heat and salinity limit yield considerably. Durum wheat is also more salt sensitive than bread wheat (Gorham et al., 1990) and saline soil has a negative effect on production (Maas and Grieve, 1990). Consequently, special efforts must be made to increase its tolerance.

Plants adapt to adverse environmental conditions through the induction of stress-responsive and stress-tolerant genes, a process that occurs mainly through transcription factors (TFs). TFs are known to mediate stress signal transduction pathways regulating downstream target gene expression and lead to stress tolerance (Shinozaki and Dennis, 2003; Chen and Zhu, 2004; Yamaguchi-Shinozaki and Shinozaki, 2005; Budak et al., 2013).

The WRKY transcription factor belongs to a very large family of transcription factors potentially involved in drought/salt stress response (Budak et al., 2013). This family originated in early eukaryotes and greatly expanded in plants (Zhang and Wang, 2005). It counts over 70 members in the *Arabidopsis* genus (*Arabidopsis thaliana*; Eulgem et al., 2000; Dong et al., 2003), 55 members among cucumber plants (*Cucumis sativus*; Wei et al., 2012), 119 members among maize plants (*Zea Mays*; Ling et al., 2011), 94 members among barley plants (*Hordeum vulgare*; Liu et al., 2014), 182 encoding genes in soybean (*Glycine max*; Bencke-Malato et al., 2014), 62 members among diploid woodland strawberry plants (*Fragaria vesca*; Wei et al., 2016) and 100 members among rice grasses (*Oryza sativa*; Xie et al., 2005).

This transcription factors family is characterized by a 60 amino acids domain containing the WRKY amino acid sequence at its amino-terminal end and a putative zinc finger motif at its carboxy-terminal end. It binds specifically to the (T)(T)TGAC(C/T) sequence motif, known as the W box, which

requires both the invariable WRKY amino-acid signature and the cysteine and histidine residues of the WRKY domain, which tetrahedrally coordinate a zinc atom (Rushton et al., 2010).

The existence of either one or two highly conserved WRKY domains is the most vital structural characteristic of the *TdWRKY* gene (Bi et al., 2016). Furthermore, the global structures of WRKY proteins are highly divergent and can be classified into different groups, which might reflect their distinct roles. WRKY proteins are classified into 3 main groups (I, II, III) based on the number of WRKY domains and the structure of the zinc finger-like-motif.

Group I proteins contain two WRKY domains followed by a C2H2 zinc finger motif. The other WRKY proteins from group II and III contain one WRKY domain followed by a C2H2 or C2HC accordingly. Group II can be divided into five subgroups (IIa, IIb, IIc, IId, and IIe) based on additional amino acid motifs (Yamasaki et al., 2005).

WRKY genes are known to participate in various developmental and physiological metabolisms, including disease resistance (Bhattarai et al., 2010), senescence (Besseau et al., 2012), growth and developmental processes (Guillaumie et al., 2010), as well as biotic and abiotic stress responses (Mingyu et al., 2012). Recently, transgenic *Arabidopsis* plants overexpressing *TaWRKY2* (EU665425) or *TaWRKY19* (EU665430) have shown improved tolerance to salt, drought and/or freezing stresses when compared with the wild-type plants (Niu et al., 2012). Marè et al. (2004) described *HvWRKY38* (AY541586), a barley gene coding for a WRKY protein, whose expression is involved in cold and drought stress response.

In durum wheat, very few WRKY EST sequences have been identified. Only partial cDNAs were found (Budak et al., 2013; Cifarelli et al., 2013). Literature does not include any report describing the characterization of a *TdWRKY* gene from tetraploid wheat species.

Available plant genome sequences from monocot plants are key resources that enable a better understanding of their gene content, structure and function. They are also indispensable for understanding transposable elements, intergenic space organization and composition. Genomic comparisons between the genome A diploid wheat donor (*Triticum urartu*), hexaploid wheat (*T. aestivum*, The International Wheat Genome Sequencing Consortium (IWGSC), 2014), barley (*H. vulgare*, The International Barley Sequencing Consortium, 2014) and Brachypodium (*Brachypodium distachyon*, The International Brachypodium Initiative, 2010), have not only confirmed the broad synteny between *Poaceae* gene content but have also helped to deduce their function within a phylogenetic context (Dubcovsky et al., 2001).

In this paper, we obtained genomic sequences of the region harboring WRKY genes from *T. turgidum* ssp. *durum* from Langdom#65 BAC clone library screening. First, we presented the WRKY gene characterization, including phylogenetic analysis and orthologous gene comparison. Then, we analyzed the genomic environment of WRKY genes, including intergenic space composition, and *cis*-acting elements that helped to associate a putative function with *TdWRKYs*. We finally performed *TdWRKYs* differential gene expression. *TdWRKY1*, 3,

**Abbreviations:** ABRE, ABA-responsive element; BAC, bacterial artificial chromosome; CRT, C-repeat element; DRE, dehydration-responsive element; ERF, ethylene response factor; EST, expressed sequence tag; LTRE, low-temperature-responsive element; NLS, nuclear localization signal; ORF, open reading frame; TE, transposable element; TF, transcription factor; UTR, untranslated region.

4, and 5 were induced by high-salt treatment in two durum wheat varieties, Grecale (GR) and Om Rabiaa (OR), shown to be salt-tolerant and -sensitive, respectively, suggesting that *TdWRKYs* may be involved in salt-stress responses. These data provide new leads toward improving durum wheat tolerance to abiotic stresses.

## MATERIALS AND METHODS

### Plant Material

Three genotypes of tetraploid *T. turgidum* L. ssp. *durum* ( $2n = 4 \times = 28$ ) were used in this study; Langdon LDN#65 for PCR bacterial artificial chromosome (BAC) screening, and OR and GR for functional analyses. The latter two were a local Tunisian variety and an Italian variety introduced in Tunisia, respectively.

### BAC Library Screening

The BAC library constructed from the tetraploid *T. turgidum* L. ssp. *durum* ( $2n = 4 \times = 28$ ) Langdon LDN#65 genotype was used in this study for PCR Bacterial Artificial Chromosome (BAC) screenings. The BAC library is represented by a total of 516,096 clones organized in 1,344,384-well plates. The average insert size of the BAC clones was estimated at around 131 kb resulting in 5.1 genome coverage (Cenci et al., 2003). The LDN#65 BAC library and related tools are available at the French Plant Genomic Resource Center upon request ([http://cnrgv.toulouse.inra.fr/en/library/genomic\\_resource/Ttu-B-LDN65](http://cnrgv.toulouse.inra.fr/en/library/genomic_resource/Ttu-B-LDN65)). The library was organized into a two-dimensional pool and BAC library screening was performed as described by Cenci et al. (2003, 2004). The pooling strategy used required 56 PCRs for the superpools (9216 BAC clones each) and 50 additional PCRs for each positive superpool. The screenings have been done as described in Makhloufi et al. (2014).

### PCR Primer Design and PCR Amplification for BAC Screening

Two primer design strategies have been used. *O. sativa*, *T. aestivum* and *H. vulgare* sequences harboring conserved WRKY domains were used as queries (NCBI: <http://www.ncbi.nlm.nih.gov/>; GrainGenes: <http://wheat.pw.usda.gov/GG2/index.shtml>; Gramene: <http://www.gramene.org/>; TIGR: <http://www.jcvi.org/>; HarvEST: <http://harvest.ucr.edu/>; Phytozome: <http://www.phytozome.net/>; PTDB: <http://plntfdb.bio.uni-potsdam.de/v3.0/>; and plant GDB: <http://www.plantgdb.org/>), in a BLAST search to find homologous *T. turgidum* ESTs (Budak et al., 2013; Cifarelli et al., 2013). Multiple alignments of the DNA sequences were performed by ClustalW software (Larkin et al., 2007). In order to avoid an amplification of the exon-intron junction, prediction of the exon boundaries within *Triticeae* expressed sequence tags (ESTs) were performed based on rice and Arabidopsis genomic sequences.

The second has been deduced due to the very strong homology between different species of *Triticeae*. We tried to identify non-specific *T. turgidum* primers from genes whose function was studied. For this, some WRKY genes that had been specifically studied in the context of abiotic stress, were chosen for primer

design: *HvWRKY38* (AY541586) (group IIa) (Marè et al., 2004) and *TaWRKY2* (EU665425) (group I); *TaWRKY19* (EU665430) (group I) (Niu et al., 2012).

PCR primers were then designed to cover exons of the entire selected sequence using the Perl primer tool v.1.1.2.1 (Marshall, 2004) (Table S1). Primers were tested on genomic DNA of LDN#65 before BAC library screening. Total DNA was extracted from wheat Langdon 65 variety using the Plant DNAzol® reagent. PCR conditions used were as follows: initial denaturation at 95°C for 5 min, followed by 45 cycles of 20 s at 95°C, 16 s at 60°C, and 20 s at 72°C, performing a melting curve with an increment of 0.5°C per cycle. PCR products for the selected BACs were separated by electrophoresis (2% Agarose).

### BAC Sequencing, Assembly, and Annotation

BAC DNA was extracted using a NucleoSpin® 96 Flash kit (Macherey-Nagel) and the insert size was estimated with NotI digestion (Fast Digest NotI; fermentas). Positive BAC clones were outsourced for 454 Life Sciences pyrosequencing technology using the GS Junior Roche system (Kit 454 Titanium; Roche). Sequence data assembly was performed with Newbler 2.8 software sold by 454 Life Sciences/Roche for 454 data (Veras et al., 2013). The assembly was performed on data previously processed by the software Pyrocleaner after clearing reads from contamination by the host *E. coli*. Sequenced BAC DNA was analyzed using TriAnnot Pipeline v.3.8 improved for wheat species (<http://wheat-urgi.versailles.inra.fr/Tools/TriAnnot-Pipeline>) enabling annotation, masking of transposable elements, and gene structure organization (Text S1) (Leroy et al., 2012).

### Alignment, Phylogenetic Tree, and Sequence Analysis of TdWRKY Genes

The TdWRKY protein sequences were submitted to the CDD (Conserved Domains Database) from NCBI, and to Motif Scan detection in MyHits (<http://myhits.isb-sib.ch/>) with Prosite databases selected. Homologous proteins from whole genome sequenced monocot plants (Wheat, Barley, Brachypodium, Maize, Sorghum and Rice) and Arabidopsis were selected for sequence alignments using DNAMAN software (<http://www.lynnon.com/>).

A neighbor joining phylogenetic tree was derived from a MUSCLE alignment 3.8 (Edgar, 2004) of TdWRKY proteins, their homologous WRKY proteins from cereals and their closest members from *A. thaliana*. The tree was then produced by MEGA 6 software (Tamura et al., 2013) using the Neighbor-Joining method with 1000 bootstrap replicates.

A 1.5 kb DNA fragment on the 5'-regulatory region upstream of the transcription start of the *TdWRKY* gene was subjected to *in silico* analysis using the PLACE signal scan and NSITE-PL (Recognition of PLANT Regulatory motifs with statistics) from the Softberry tool (<http://www.softberry.com/>) to search for putative *cis*-regulatory elements in the promoter region potentially involved in controlling *TdWRKY* gene expression and also in the promoter region of their orthologs in *T. aestivum* and



*H. vulgare* to see the average frequencies of *cis*-acting elements related to abiotic stress in these promoters.

## Salt Stress

Sterile seeds of two independent genotypes from *T. turgidum* subsp. durum, OR and GR, were first stored for 48 h at 4°C for initialization of germination. Seedlings were sown in recipient Magenta™ vessels containing 50 ml of 50% MS-based medium (Murashige and Skoog, 1962) and were left for 10 d in an *in vitro* growth chamber maintained at a controlled photoperiod of 14 h during the day at 25°C with 80% humidity and an intense luminosity of 250  $\mu\text{mol m}^{-2} \text{s}^{-1}$ , and for 10 h during the night at 20°C. They were then subjected to abiotic and hormonal stress treatments. For salinity treatment, seedlings were transferred into 50% MS medium containing 200 mM NaCl for 6 or 24 h. Leaves and roots were then harvested separately, dropped immediately into liquid nitrogen, and stored at –80°C for RNA extraction.

## Gene Expression Analysis

Total RNA from at least 30 salt-treated and untreated leaves and roots from OR and GR genotypes were extracted using a Pure Link Plant RNA Reagent kit (Invitrogen). Total RNA was DNase treated (Promega), and first-strand cDNA was reverse transcribed from 2  $\mu\text{g}$  of total RNA using an M-MLV Reverse Transcriptase kit (Promega) according to the manufacturer's instructions. First-strand cDNA generated from total RNA including salt-treated and untreated samples from either the OR or GR genotype was subjected to real time quantitative expression analysis. The latter was performed in a fluometric thermal cycler (DNA Engine Opticon 2; MJ Research, Waltham, MA, USA) using SYBR Green fluorescent dye following the manufacturer's instructions. Results were shown using SDS2.2 software on an Applied Biosystem 7900 HT Fast Real-Time PCR System. Comparisons of repeated samples were assessed using CT values among the three replications. Linear data were normalized to the mean CT of 26S rRNA as an internal reference gene and the relative expression ratio was calculated using the formula  $2^{-\Delta\Delta\text{CT}}$ . Log2-transformed signal ratio was carried out. The gene specific primers used for PCR are listed in Table S2.

## RESULTS

### Identification, Sequencing and General Features of the Sequence Composition of *T. durum* BAC Clones

516,000 clones from the Durum wheat bacterial artificial chromosome (BAC) library, were screened for individual clones harboring WRKY genes. Six *T. durum* BAC clones were selected and fully sequenced by 454 technologies.

BAC clone annotation revealed 12 non-TE genic features that were classified into two categories: 9 protein coding genes, one hypothetical and 2 gene fragments (Table S4). The gene assignments were all supported by at least one full-length cDNA, an EST and/or homolog in another monocot plant such as *T. aestivum*, *O. sativa*, *H. vulgare*, *Z. mays*, and *B. distachyon*. One gene per insert was predicted for TD14H23 (GenBank accession no. KY091673), TD473J01 (GenBank accession no.

KY091677), TD493B21 (GenBank accession no. KY091678) and TD315C07 (GenBank accession no. KY091676). The remaining two BAC clones TD16L16 (GenBank accession no. KY091674) and TD789O23 (GenBank accession no. KY091679), contain four genes (Table 1). Two genes (of known or unknown function), three genes with putative function, four genes with domain-containing proteins, one hypothetical gene and one truncated (fragmented gene), were identified (Table 1, Table S4).

An average GC content of 47.5% was found for all BACs. In contrast with the constant GC content, gene and TE composition was highly variable between the different BAC sequences. The proportion of TEs ranged from 64 to 91%, while gene content ranged from 1 to 4 genes per BAC. The coding fraction of the 960,769 kb total samples represents 3.2% of the sequences, while the TE content is 74.6%, distributed as follows: 63.6% for class I, 7.9% for class II and 3.1% for unclassified TEs (Table S3). While class I retrotransposons constitute the highest TE proportion of the 6 sequenced regions, BAC clone TD473J01 shows the highest proportion of CACTA class II (17.0%). Class I TE DNA sequences were distributed as follows: 27.2% *Gypsy*- (150 TEs) and 15.9% *Copia*- (71 TEs) like “long terminal repeats (LTR)”-retrotransposons. New class I transposable elements were identified for the first time in this study. They account for 17.9% of length and were identified as *de novo* LTR-retrotransposons (Table S5). Class II TEs (DNA transposons) represent 7.9% of the cumulative sequence length. The CACTA TEs represent the majority (32.4%) of class II DNA sequences. Novel class II TE families were identified, sharing weak homologies with known CACTA, *Mutator* and *Mariner*. The 3.1% novel unclassified elements, share a stretch of weak homology with other Triticeae unclassified transposable elements (Table S3, Text S2).

### TdWRKY and Monocot Gene Structure

A comparison of durum wheat, *T. aestivum*, *H. vulgare*, and *Brachypodium distachyon* WRKY gene sequences helped to establish their structure.

Gene length varies from 1.2 to 4.2 Kb. It is conserved amongst the orthologous monocots (*TdWRKY1* and *TdWRKY5* length was highly conserved among orthologous members). The number of introns varies from 2 to 4. The intron size is also variable; it ranges from 102 to 925 bp (Figure 1, Table 1). All exon-intron junction sites obey the GT/AG rule as identified in other eukaryotic genes. To date, the relative organization of the exons and introns is the same for the other WRKY genes characterized in cereal, i.e., the number of exons and introns remains the same and individual introns occur at relatively the same sites for barley, *Brachypodium* and wheat genes.

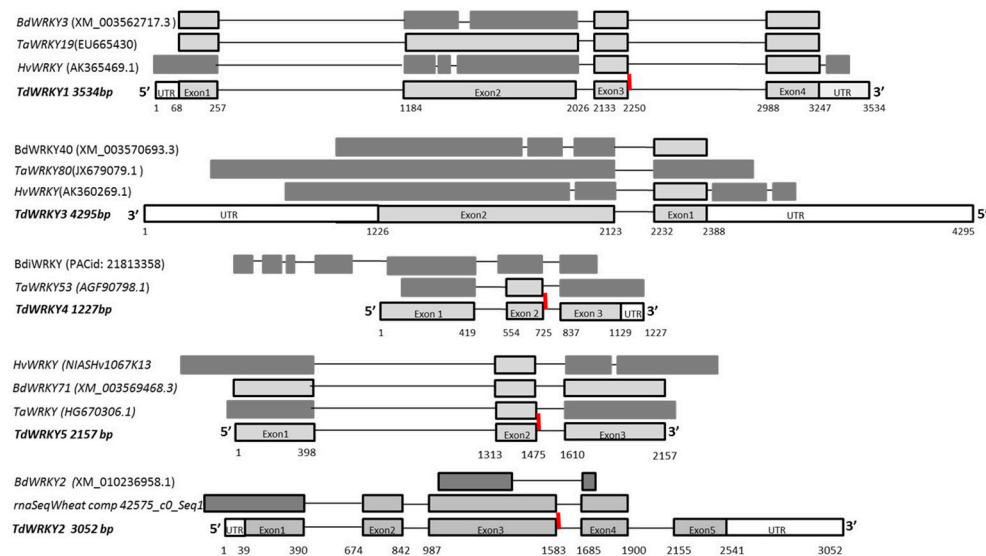
The size of coding sequences varies from 882 to 1716 bp (Figure 1). The size of exonic regions between the orthologous genes was similar although the overall region structure was slightly different between *TdWRKYs* and their orthologs (exon 2 from *TdWRKY1* and *TdWRKY2* were almost the same size, but with 1 to 3 more intron phases in corresponding exons. The same was observed with exon 3 from *TdWRKY5*). Small stretches of exonic region sequences or more exons (*BdiWRKY* has 4 more exons, in the 5' moiety, than *TdWRKY4*) do not contradict the



TABLE 1 | TdWRKY gene models and comparisons with homologous sequences in grass species using Blastp and Blastn.

BAC clone ID	T.turgidum predicted gene	BLASTP				BLASTn			
		Protein Acc.No	Identity (%)	Score	Expect	cDNA Acc.No	Identity (%)	Score	E-value
TD14H23	Known_function, WRKY DNA binding domain	<i>T. aestivum</i> ACD80362.1	98	953	0.0	<i>T. aestivum</i> EU665430.1	99	2521	0.0
	3534bp	<i>T. urartu</i> EMS57536.1	94	811	0.0	-	-	-	-
	1407bp	<i>H. vulgare</i> BAJ98672.1	78	755	0.0	<i>H. vulgare</i> AK365469.1	93	1625	0.0
	469aa	<i>B. distachyon</i> XP_003562765.1	70	635	0.0	<i>B. distachyon</i> XM_003562717.3	86	1137	0.0
	CAT01	<i>A. thaliana</i> NP_178433.1	45	228	1e-67	-	-	-	-
TD16L16	Putative_function, Putative WRKY transcription WRKY	<i>T. urartu</i> EMS51071.1	93	1124	0.0	<i>B. distachyon</i> XM_010236958.1	86	522	5e-69
	3052bp	<i>A. taushii</i> EMT16491.1	88	1010	0.0	-	-	-	-
	1716bp	<i>B. distachyon</i> XP_010235260.1	53	536	2e-180	-	-	-	-
	572aa	<i>A. thaliana</i> NP_200438.1	30	199	9e-55	-	-	-	-
	CAT04	-	-	-	-	-	-	-	-
TD493B21	Putative_function, WRKY DNA binding domain	<i>T. aestivum</i> ABN43185.1	99	605	0.0	<i>T. aestivum</i> EF368364.1	99	1591	0.0
	1227bp	<i>T. aestivum</i> AGF90798.1	98	602	0.0	<i>S. italica</i> XM_004962359.3	79	503	3e-138
	882bp	<i>A. tauschii</i> EMT33135.1	87	508	2e-176	<i>T. aestivum</i> EF368357.1	92	1151	0.0
	294aa	<i>A. thaliana</i> AAM34736.1	54	278	5e-89	<i>T. aestivum</i> EU665443.1	91	1138	0.0
	CAT01	-	-	-	-	-	-	-	-
TD473J01	Putative_function, WRKY transcription factor	<i>T. aestivum</i> AFW98256.1	100	724	0.0	<i>T. aestivum</i> JX679079.1	99	1940	0.0
	4295bp	<i>H. vulgare</i> CAD60651.1	91	626	0.0	<i>H. vulgare</i> AK360269.1	92	1487	0.0
	1053bp	<i>S. purpurea</i> AJF34891.1	81	554	0.0	<i>H. vulgare</i> AV541586.1	94	1367	0.0
	351aa	<i>B. distachyon</i> XP_003570741.1	70	437	3e-149	<i>B. distachyon</i> XM_003570693.3	80	734	0.0
	CAT04	<i>A. thaliana</i> NP_178199.1	36	198	9e-60	-	-	-	-
TD315C07	Domain_containing_protein, WRKY DNA binding domain	<i>T. aestivum</i> CDM83596.1	100	751	0.0	<i>H. vulgare</i> AK358052.1	92	1507	0.0
	2157bp	<i>H. vulgare</i> BAJ98040.1	89	511	1e-177	<i>T. aestivum</i> HG670306.1	99	1011	0.0
	1107bp	<i>B. distachyon</i> XP_003569516.1	70	356	1e-116	<i>B. distachyon</i> XM_003569468.3	82	861	0.0
	369aa	<i>A. thaliana</i> NP_174279.1	46	173	3e-50	-	-	-	-
	CAT02	-	-	-	-	-	-	-	-
TD789O23	Gene fragment, WRKY DNA binding domain	<i>T. aestivum</i> CDM83596.1	98	285	3e-85	<i>T. aestivum</i> HG670306.1	93	990	0.0
	683bp	<i>H. vulgare</i> BAJ98040.1	91	177	4e-51	<i>H. vulgare</i> AK358052.1	92	785	9e-162
	402bp	<i>B. distachyon</i> XP_003569516.1	74	100	4e-22	<i>B. distachyon</i> XM_003569468.3	78	246	4e-61
	134aa	<i>T. aestivum</i> CDM83596.1	98	375	1e-125	<i>H. vulgare</i> AK358052.1	93	1016	0.0
	CAT04	<i>H. vulgare</i> BAJ98040.1	91	352	8e-117	<i>T. aestivum</i> HG670306.1	94	1689	0.0
	Gene fragment, WRKY DNA binding domain	<i>B. distachyon</i> XP_003569516.1	74	273	2e-85	<i>B. distachyon</i> XM_003569468.3	84	584	9e-163
	1567bp	<i>A. thaliana</i> NP_193551.1	75	174	5e-51	-	-	-	-
	861bp	-	-	-	-	-	-	-	-
	287aa	-	-	-	-	-	-	-	-
	CAT03	-	-	-	-	-	-	-	-

The genes models were defined on the table using the genomic sequence length (bp), the CDS length (bp), the protein length (aa) and the categories (CAT) as a quality index for the annotator (CAT01, CAT02, and CAT03 correspond to biological evidence found on *Triticum* and *Aegilops* Full-length cDNAs, Poaceae Full-length cDNAs and CDS from *O. sativa* (IRGSP) and *B. distachyon* genomes annotation, respectively, the genes models predicted by EuGene belong to CAT04.



**FIGURE 1 | Isolation of 5 *TdWRKYs* by screening of a durum wheat BAC library and Comparative Structure of *TdWRKYs* and orthologous from *B.distachyon*, *T. aestivum* and *H. vulgare*.** Genomic structure of the durum wheat *TdWRKYs* gene; white portion are UTR regions, dark gray portions are orthologous exons from different sizes, and light gray boxes are exons from same size between orthologous. The line represents the intron. The red motives are special WRKY family conserved intron. *TdWRKYs* untranslated regions, introns and exons boundaries are highlighted on the bottom of *TdWRKYs*.

general pattern of overall high WRKY gene conservation between the *TdWRKYs* and their homologous genes (Text S3).

Comparison of the *TdWRKYs* cDNA with sequences of other species showed an identity ranging from 99% for *T. aestivum*; 86% for Brachypodium and 92% for barley (Table 1, Text S3).

Interestingly, we also identified a cluster of non syntenic collinear genes, probably originating from genomic rearrangements of gene blocks such as *gene 1* and *gene 4* from TD789O23 (the two gene fragments from Table 1). They were annotated as being similar to the WRKY domain protein. *Genes 1* and *4* (order on the BAC clone), share 93 and 94% of similarity, respectively, with *TdWRKY5*. An alignment of *TdWRKY5* from TD315C07 and the two fragments (*gene 1* and *gene 4*) from TD789O23, revealed that *gene 4* might be a truncated second allelic form of *TdWRKY5* (Figure 1). In fact, the *gene 4* genomic sequence is 989 bp shorter (from position 1 to 990, Figure S2) than *TdWRKY5*. It results in the formation of a new coding protein. The predicted protein from *gene 4* belongs to group II as well as *TdWRKY5*. The WRKY domain on exon2, the PR intron special feature, and the zinc finger-like motif on exon3, are perfectly conserved. The alignment of *TdWRKY5*, *gene 4* and *gene 1*, showed that *gene 1* and *gene 4* happened to be fragments of *TdWRKY5* (Figure S2). This disruption was the result of a deletion and an insertion. These are two of the most important genomic rearrangement events. They play significant roles in genome evolution. This reshuffling created two new, non-syntenic genes with orthologous species (Table 1). This might, or might not, be a loss-of-function mutation of the second *TdWRKY5* allelic form. Unfortunately, we were only able to find approximately 450 bp from the 898 bp of newly inserted fragment (*gene1*) since it is located at the extreme 5' start of

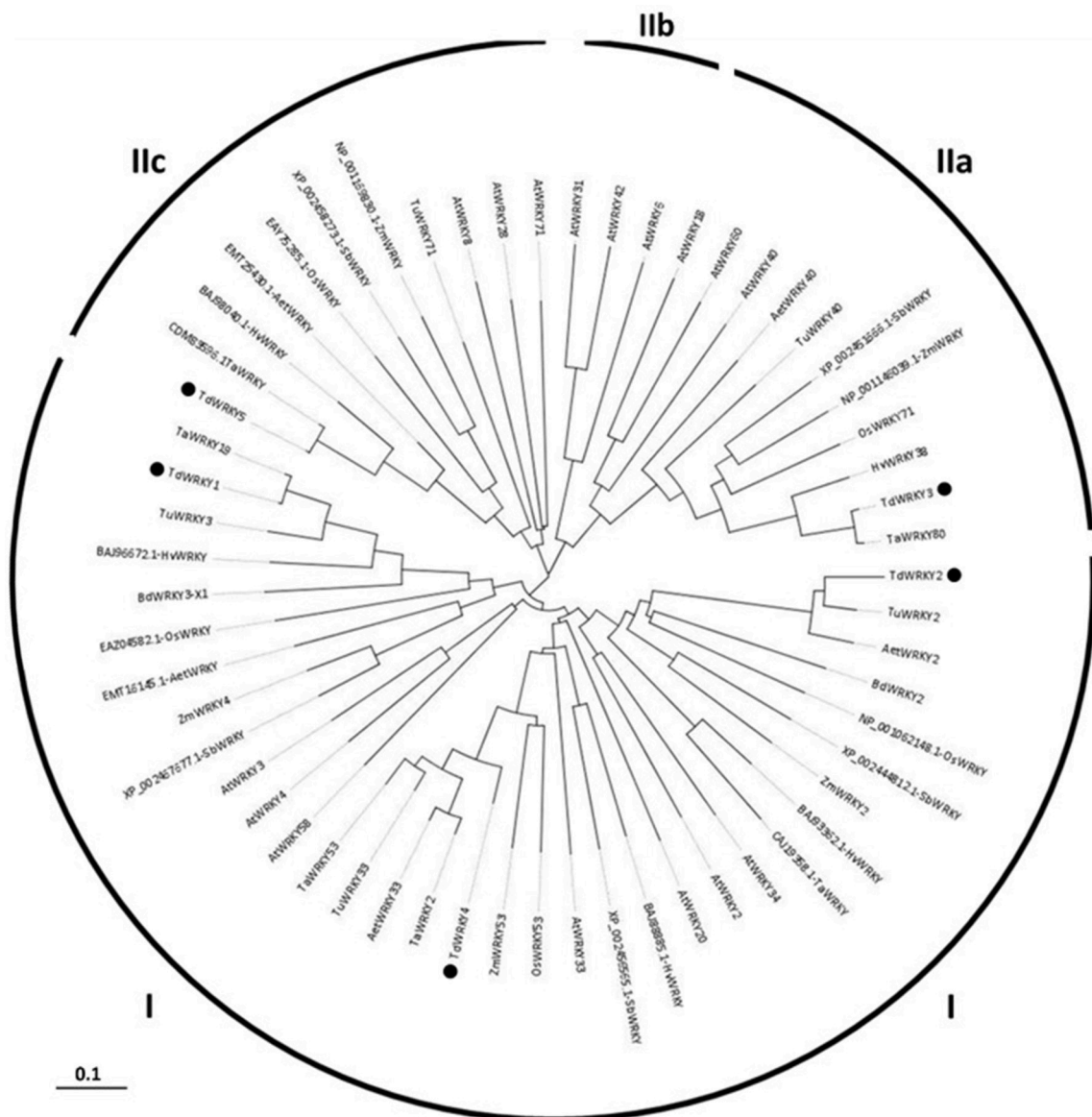
TD789O23 DNA insert. There are many ways in which exon shuffling may occur. Shuffling involves transposable elements such as *LINE-1* retroelements and *Helitron* transposons, as well as *CACTA* elements and *LTR* retroelements, a crossover during sexual recombination or alternative splicing. These hypothesized mechanisms should be thoroughly explored.

## TdWRKYs Classification and Phylogeny

WRKY proteins are classified into 3 groups. We used phylogeny to assign a group to our WRKY sequences (Figure 2). WRKY1, 2, and 4 belong to group I. They all have 2 WRKY domains with a C-X (4,5)-C-X (22,23) H-X-H zinc-finger-like type motif on each domain. *TdWRKY2* predicted protein contains two WRKY domains. The N-terminal domain has an altered WRKY motif WRKYGKK. An alignment of *TdWRKY2*, with its orthologous proteins from Aegilops, Brachypodium and *T.urartu* on the UniProtKB database (<http://www.uniprot.org/align/>), indicates that the protein is biased at its C-terminus WRKY conserved domain. 29 aa are missing from the zinc finger motif (Figure S1).

This might be the cause of alteration in Wbox recognition. The role of conserved residues in this domain has been studied by Maeo et al. (2001) using mutation experiments. Any mutation occurring either in the WRKYGQK or the zinc finger motif of WRKY domains, affecting cysteine or histidine, cancels DNA binding activity (Knoth et al., 2007). The WRKY domain from the C-terminal region is responsible for binding to DNA, whereas the role of the N-terminal is to promote protein-protein interactions (Maeo et al., 2001).

*TdWRKY3* and 5 were assigned to groups IIa and IIc respectively, due to the presence of only one WRKY domain with a specific zinc finger motif, C-X (4,5)-C-X (22,25)-H-X-H.



**FIGURE 2 |** Phylogenetic analysis of the TdWRKYs proteins with orthologous members from cereals (*T. aestivum*, *T. urartu*, *Ae. taushii*, *H. vulgare*, *Z. mays*, *S. bicolor*, and *O. sativa*) and *A. thaliana*. The Arabidopsis WRKY sequences were downloaded from GenBank (we used gene names for the alignment). TdWRKYs are spotted with black motives. Alignment of the 59 amino acids sequences was performed by MUSCLE 3.8 (<http://www.ebi.ac.uk/Tools/msa/muscle/>). The evolutionary history was inferred using the Neighbor-Joining method. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed. The evolutionary distances were computed using p-distance method and are in the units of the number of amino acid differences per site. Partial deletion parameters were used, all positions with less than 50% site coverage were eliminated, meaning that, fewer than 50% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 459 positions in the final dataset. Evolutionary analyses were conducted in MEGA6. All the branches were supported with bootstrap values  $\geq 50\%$ .

We investigated the Prosite and Pfam databases using the TdWRKYs sequences as queries to identify conserved domains, motifs and active phosphorylation sites (Figure 3, Text S4). The analysis predicted 11 specific sequence features for all TdWRKYs identified, including the WRKY domain, a Plant Zinc Clust domain, and a Gly\_Rich region and a His\_Rich region (Figure 3). Seven active sites were detected, including 32 CK2\_Phospho sites, 22 PKC\_Phospho sites, 12

Asn\_Glycosylation sites, 29 Myristyle sites, 2 TYR\_pospho sites, one amidition\_site and 3 CAMP\_phospho\_sites. These features are related to subcellular localization, signal transduction, transcriptional regulation and protein interaction, and build a basis for TdWRKY function. TdWRKY1, 4, and 5 might act as activators. TdWRKY2 and 3 contain the active repressor motif (LXLXLX) (Xie et al., 2005) (Figure 3).

For additional phylogeny support, the detection of the position and the phase of the intron in each region encoding a WRKY domain is essential. Four of the eight WRKY domains, found in *TdWRKY* genes, contain an intron in a conserved position (**Figure 1**). This phase 2 intron is localized at the 11th codon downstream of the WRKY motif, interrupting the codon encoding Arg (on 100% of genes). In Group Ia (*TdWRKY1*, 2, and 4) (the C-terminal WRKY domain), Group IIc (*TdWRKY5*), IId, IIe, and III genes, this intron comes after the codons for the invariant amino acid sequence PR and separates the WRKY sequence from the zinc finger motif. In Group IIa (*TdWRKY3*) and IIb genes the intron occurs at a nucleotide position that corresponds to five amino acids after the C-X5-C and separates this from the rest of the zinc finger structure (Tripathi et al., 2012). On *TdWRKY3*, the WRKY domain was separated from the zinc finger motif by the C-X5-C core but there was not an intron at this position (**Figure 1**).

### Hypothetical Function Deduced By Protein Domain and Orthologous Gene Function Sequence

The phylogenetic analyses revealed distinct clusters of TdWRKYs. Within the WRKY clade, distinct clusters corresponding to WRKY groups and subgroups comprising of further sub-clusters emerged. TdWRKYs with WRKYs from wheat, barley, *T. urartu* and *Aegilops* formed one sub-cluster whereas rice, maize and sorghum were grouped as a separate sub-cluster, and Arabidopsis were clustered together. Within sub-clusters, durum wheat, and *T. aestivum*, WRKYs showed less divergence (as indicated by a shorter branch length) than Brachypodium and barley WRKYs. Similarly, maize and sorghum WRKYs showed less divergence compared to rice. Monocot WRKYs showed less divergence than the closest Arabidopsis WRKYs (**Figure 2**).

TdWRKY1 was clustered to TaWRKY19, involved in salt stress response. TdWRKY2 was clustered to BdWRKY2 and AtWRKY34, involved in cold stress response (Zou et al., 2010). TdWRKY4 belongs to the TaWRKY53, TaWRKY2 and AtWRKY33 cluster, involved in salt stress (Jiang and Deyholos, 2009). TdWRKY3 is clustered to TaWRKY80 and HvWRKY38, involved in dehydration response and to AtWRKY18, 60, and 40 (Liu et al., 2012). They were described as regulating defense response. TdWRKY5 and its orthologous TaWRKY71, BdWRKY71, and AtWRKY8, involved in salt stress response, were grouped together (**Figure 2**).

Multiple sequence alignment of the conserved sequence from group Ia TdWRKYs (the two WRKY domains were aligned separately) and group IIa and IIc TdWRKY members with their closest Arabidopsis and monocot WRKY proteins (**Figure 4**) revealed the very conserved structure of the WRKY motifs and the amino acid residues potentially interacting with zinc ligands. TdWRKYs consist of a four-stranded  $\beta$ -sheet (1, 2, 3, and 4), with a zinc binding pocket formed by the conserved Cys/His residues located at one end of the  $\beta$ -sheet, and the WRKYGQK cores, corresponding to the most N-terminal  $\beta$ -strand (strand

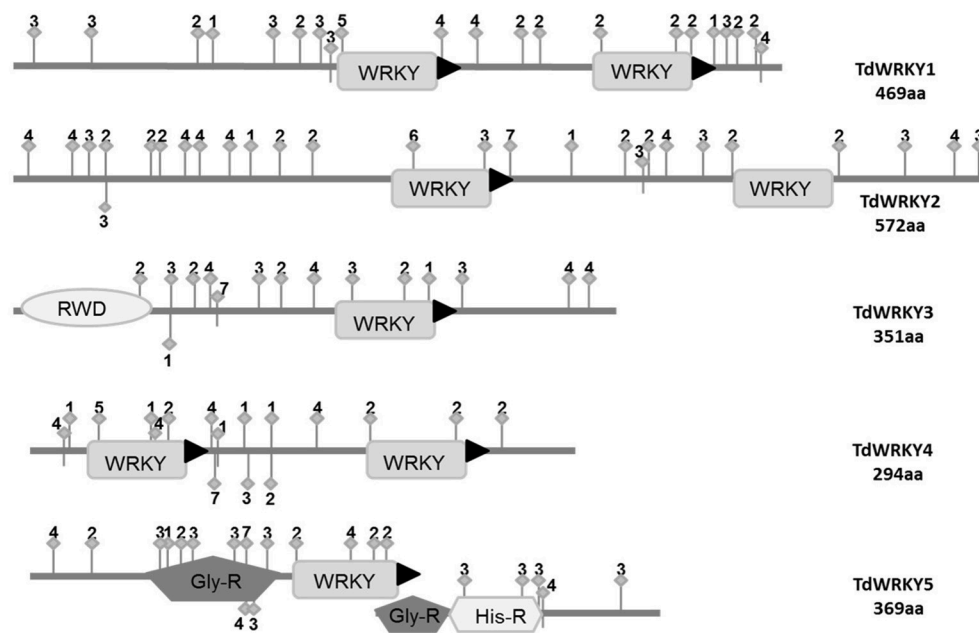
$\beta$ -1), kinked in the middle of the sequence by the Glyresidue (Yamasaki et al., 2005).

### Putative cis-Acting Elements Identified in the *TdWRKY* Promoter Region

BAC sequences generated more data about gene environment. It affords greater reliable and precise information for further functional studies. A 1.5 kb 5' regulatory region upstream of the transcription start of the *TdWRKY* genes was subjected to *in silico* analysis using a plant cis-acting regulatory DNA elements (PLACE) signal scan to search for putative *cis*-regulatory elements potentially involved in the control of *TdWRKY* gene expression. The data indicated the presence of a large number of conserved *cis*-regulatory elements that are putative targets for TFs reported to mediate responses to environmental stresses or to stress-related hormones (**Table 2**). To regulate gene expression, WRKY factors show a binding preference to a DNA sequence called W box: (C/T) TGAC (C/T) (Ciolkowski et al., 2008). This DNA core was over-represented on *TdWRKY3*, 4 and 5. 6 and 5 boxes were found on *TdWRKY1* and 2, respectively.

It has also been shown that some WRKY factor types can bind to other types of *cis*-elements. ABRE-like motifs (ACGTG) and ABRE-related motifs (ACGTGKC and TACGTGTC) were also found in the promoter region of *TdWRKY1* 3, 4, and 5. The MYB-core element (TAAGT) and a number of MYB-related motifs (YAACKG, CNGTTR, and GGATG), as well as a MYC (CANNTG) motif and the MYC-related motifs (CATGTG and CACATG), were present on the entire promoters. MYC and MYB had the biggest number among others. DRE (TACCGACAT), CRT (RCCGAC), and low-temperature responsive elements (LTREs) (CCGAC), all containing the CCGAC motif that forms the core of the DRE sequence, were well represented in the promoter of *TdWRKY1* and *TdWRKY3*. DRE-like elements such as CBFHv (RYCGAC) were also identified mostly on *TdWRKY1* and *TdWRKY3*. Two GCC-box motifs (AGCCGCC), target sequences for ERF proteins, were found in the *TdWRKY1* and *TdWRKY4* promoters. RAV1-A motifs (CAACA), to which RAV1 proteins can bind through their AP2 and B3-like domains, were also present in a large number on *TdWRKY* promoters (**Table 2**). The average *cis*-elements frequencies on the WRKY promoter gene were deduced from *TdWRKYs* and from orthologous *TaWRKYs* and *HvWRKYs* scanned also for regulatory elements. The results showed that the promoters of *TdWRKY* genes are the richest in terms of putative regulatory elements compared to *HvWRKYs* and *TaWRKYs*. Moreover, the frequencies of GCC boxes, MYB and Wboxes in *A. thaliana* promoter regions are significantly lower than in hexaploid wheat, durum wheat and barley, which might be related to the fact that the average frequency was calculated within the 74 WRKY members (Dong et al., 2003). Promoter analysis in *Populus* revealed that various *cis*-acting elements (LTRE, ABRE, ABA, and Wboxes) involved in abiotic stress and phytohormone responses were highly present in the promoter region of *PtrWRKY* genes (Jiang et al., 2014). The data presented by Makhloufi et al. (2014) on durum wheat *TdERF* indicated the presence of a large number of conserved *cis*-regulatory elements that are putative targets for





**FIGURE 3 | Sequence features of the TdWRKYs protein.** Domains or motifs: WRKY (WRKY domain), RWD, Pla\_Zn (Plant Zinc Clust domain), PRO\_R (Pro\_Rich region), MET\_R (Met\_Rich region), NLS (Nuclear Localization Signal), LXLXLX (repressor motif), and Fila (Filamin domain; Active sites: 1, Asn\_Glycosylation; 2, CK2\_Phospho\_Site; 3, Microbodies\_Cter site; 4, Myristyle site; 5, PKC\_Phospho\_Site; 6, Amidation site; 7, CAMP phosphosite).

TFs reported to mediate responses to environmental stresses or to stress-related hormones.

## The Expression Patterns of *TdWRKYs* under Salt-Stress Conditions in Two Contrasting Genotypes

The expression pattern of *TdWRKYs* in response to short term salt stress, in both leaves and roots, was analyzed in the OR and GR genotypes of durum wheat shown to be sensitive and tolerant to high salt, respectively (Makhloufi et al., 2014). Specific primers (Table S2) were designed and used in a quantitative real-time PCR. Upon high-salt treatment (200 mM NaCl), the expression levels of *TdWRKY1*, 3, 4, and 5 were altered (up to 19-fold in *TdWRKY4*) in the leaves of sensitive genotype after 6 h under salt stress (Figure 5). Meanwhile, the transcripts level at the same time in tolerant genotype remain almost unchanged for *TdWRKY3*, 4, and 5 and increased almost 5 times than control (Figure 5A). Thereafter, the expression of the *TdWRKY* genes in leaf tissues displayed a dramatic increase at 24 h in both genotypes, even though the upregulation was substantially higher in GR (27-fold, 17-fold, 11-fold, and 15-fold for *TdWRKY1*, 3, 4, and 5, respectively) than in OR (1.5-fold altered, 13-fold, 3-fold, and 12-fold for *TdWRKY1*, 3, 4, and 5, respectively) (Figure 5A). In treated OR roots, transcripts levels of the four *TdWRKYs* remain unchanged at 6h in both sensitive and tolerant genotypes. Application of salt induced a decrease in transcript levels of *TdWRKY3* (12-fold) and *TdWRKY4* (14-fold) after 24 h in sensitive genotype. The expression levels of all *TdWRKYs* after 24 h

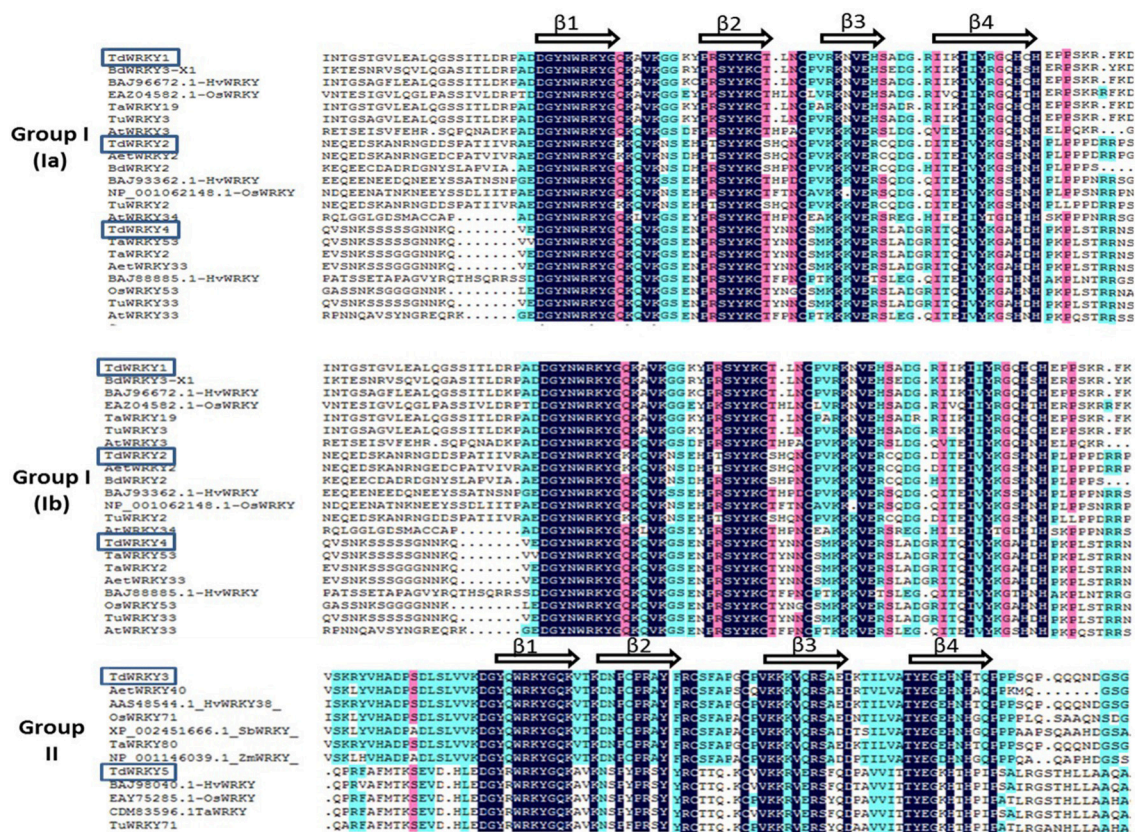
of stress treatment, in tolerant genotype, remain constitutive (Figure 5B).

*TdWRKY2* carries a deletion within the fourth and last exon, just after the position encoding the second Cys of the zinc finger motif (Figure S1). By qRT-PCR analysis, we verified disruption of the gene. 2 primer pairs positioned 3' around the deletion site did not amplify any product (Table S2). As a disruption of the zinc finger motif has been shown to completely abolish the W-box-specific DNA binding activity of WRKY transcription factors (Maeo et al., 2001), it is very likely that *TdWRKY2* is a true loss-of-function gene, as was suggested for *TdWRKY2* protein.

## DISCUSSION

### TE Expansion Responsible for Durum Wheat Genome Organization

Despite the accumulation of complete plant genome sequences, the most comprehensive studies on the organization of gene space throughout the sequence were carried from individual BAC clones or broader regions composed of BAC straddling, also called contigs (Brenner et al., 1993; Vitte and Bennetzen, 2006; Liu et al., 2007). We obtained, from the representative 6 BACs, a cumulative sequence length of almost 1 Mb. The TE proliferation was pronounced (representing 74.6%). LTR (Long Terminal Repeats, class I) appears to be the most represented in durum wheat sequences (63.6%) and among all grasses (Devos, 2010). Class II DNA transposons are generally less invasive than the retrotransposons in plant genomes. Durum wheat class II representation is slightly lower than bread wheat (16%) with



**FIGURE 4 | Multiple alignment of the WRKY conserved domain of TdWRKYs with closely related WRKY proteins using DNAMAN software (<http://en.bio-soft.net/format/DNAMAN.html>). Blue, clear blue and pink represent 100, 75, and 50% similarity, respectively. Four  $\beta$ -sheets indicated over the corresponding sequences.**

7.9%. Within class II members, MITE elements represent only 0.5% on durum wheat sequences; the same percentage is observed in bread wheat (Choulet et al., 2010). The composition of durum wheat is closely comparable to bread wheat composition and consequently to maize genome one (Schnable et al., 2009).

Although they are representative of the abundant wheat TEs available in the TREP database (Wicker et al., 2002, 2007; <http://boterv2.uzh.ch/kelldata/trep-db/TEClassification.html>), the class I and class II TEs observed in the genomic sequences of the wheat genomes may not cover all wheat TEs. It is expected that more wheat TEs will be identified, as more wheat genomic sequences become available and as more *de novo* TE annotation tools are developed (Choulet et al., 2010; Flutre et al., 2011). This is particularly supported by the identification in this study of different novel TE families, most of which are retrotransposons (17.9% of cumulative sequence).

## Durum Wheat Genes Clustered into Small Islands

Our data show that durum wheat genes are clustered mainly into several very small islands (from one to four genes) per BAC, separated by large blocks of repetitive elements. Overall durum wheat gene density was estimated at one gene every 80 Kb.

Gene islands reflect a proliferation of the genome. They are common features of large and repetitive genomes, such as wheat and maize genomes and are not found in small genomes such as rice, *Arabidopsis thaliana* (The Arabidopsis Genome Initiative, 2000), and Brachypodium (Huo et al., 2009). In fact, two BAC contigs (961 and 594 kb), in maize, were analyzed and gene blocks varied between one and four genes per block (Kronmiller and Wise, 2008). This suggests that within large plant genomes, gene islands may originate from a specific selection against the separation of genes by TE insertions that would be deleterious for gene expression or regulation and second a homogeneous expansion combined with preferential deletions in gene-rich regions (Choulet et al., 2010).

Genes from islands would share common functional characteristics (Hurst et al., 2004). Genes are maintained close to one another because this configuration would provide a selection advantage and functional significance (Batada et al., 2007; Janga et al., 2008). This is confirmed by the identification of co-expressed genes or relatives sharing the same functions that are conserved between the Arabidopsis, rice and poplar genomes (Liu and Han, 2009). *TdWRKY2* and *gene2* coding for a chloride channel protein (CLC) (Table 1) (from TD16L16), involved in vacuolar compartmentation during salt stress (Hechenberger

TABLE 2 | Putative cis-regulatory elements in the TdWRKY promoter regions.

Cis-elements	Sequences	Conditions	References	Cis-regulatory elements numbers					Average frequency of cis-elements in WRKY genes promoters			
				TdWRKY1	TdWRKY2	TdWRKY3	TdWRKY4	TdWRKY5	AtWRKY	TaWRKY	HvWRKY	
ABRE	YACGTGKC	ABA, dehydration, salt	Yamaguchi-Shinozaki and Shinozaki, 2005	5	0	6	6	5	0	2.2		3.4
ARF	TGTCTC	Auxin	Hagen and Guilfoyle, 2002	0	2	0	0	2	0	0		0
CBH-HV	RYCGAC	Dehydration, Cold	Svensson et al., 2006	4	0	4	0	0	0	0.6		1.4
GCC-box	GCGGCC	Ethylene, dehydration	Wu et al., 2007	5	0	0	6	0	0.73	3.8		3
LTRE	COGAC	Cold	Baker et al., 1994	6	1	5	2	3	0	1.8		1.4
MYB	YAACKG	ABA, dehydration salt	Abe et al., 1997	12	12	11	12	12	1.47	5		6.2
MYC	CANNTG	ABA, dehydrationsalt	Abe et al., 1997	11	13	5	5	7	0	3.6		0
RAV	CAACA	Ethylene	Kagaya et al., 1999	8	9	2	4	3	0	2.8		0.4
Wbox	TGAC	Biotic and abiotic stress	Eulgem et al., 1999, 2000	6	5	14	12	15	4.4	4.2		4.6
DRE/CRT	R(A/G)CCGAC	Dehydration, Cold	Dubouzet et al., 2003; Skinner et al., 2005	4	0	3	0	1	0	1.2		2.6

The 1.5 Kb DNA fragments corresponding to the 5' -regulatory regions upstream of the ORF (Open Reading Frame) transcription start sites of TdWRKY genes were subjected to an in silico analysis using a PLACE signal scan and NSITE-PL to search for putative cis-regulatory motifs potentially involved in the control of TdWRKY genes expression. R, A/G; Y, C/T; K, G/T; N, any nucleotide. Frequencies of potential binding sites in AtWRKY promoters were derived from the study of Dong et al., 2003.

et al., 1996), might share the same function. This assumption is more defensible, because of their very close genomic distance; they are separated only by 500 bp. The same gene structure is conserved on *B. distachyon* orthologous region. Gene ontology and expression profiles along the chromosome in *T. aestivum* revealed that these islands are enriched in genes sharing the same function or expression profiles suggesting the existence of long-distance regulation mechanisms in wheat (Choulet et al., 2010).

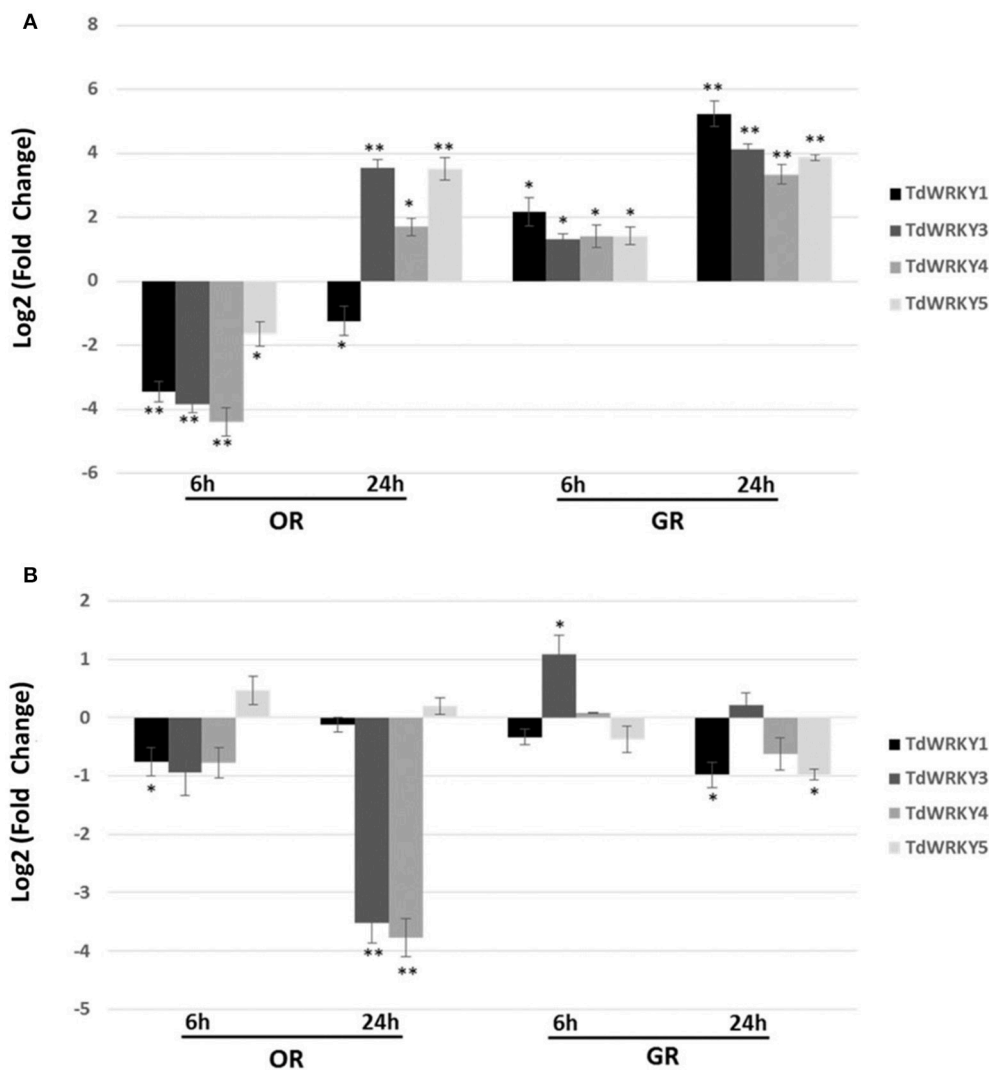
### Correlation between Sequence Similarity and Functional Similarity

Protein sequences contain important information for protein function. We found that TdWRKYs have sequence features including domains/motifs: WRKY domains, which bind WRKY proteins to the W-box motif in the promoter of target genes for transcription regulation; Plant Zinc finger motifs which function in association with WRKY domains; NLS (nuclear Localization Signal) peptides, responsible for leading proteins to the cell nucleus. And active sites such as protein kinases, which play important roles in signal transduction pathways. Special features such as Myristyl sites could function during TdWRKY5-mediated gene responses to stresses, as myristoylation sites play a vital role in membrane targeting and signal transduction in plant responses to environmental stresses (Podell and Gribskov, 2004).

Homologous genes with similar sequences are likely to have equivalent functions and to play the same functional role in equivalent biological processes. It is thus very important to identify homologous genes, especially those which are supported by experimental data. TdWRKYs from different groups (Ia, IIa, and IIc) have a close relationship with orthologous genes from barley, Brachypodium, bread wheat, maize, sorghum, aegilops, *T. urartu*, rice and Arabidopsis. Genes from *T. aestivum*, *Ae. tauschii*, *T. urartu*, and *H. vulgare* shared the closest similarity, while rice, sorghum and maize were out-grouped from other monocot plants.

As summarized in Table 2, the promoter regions of TdWRKYs were highly rich in cis-acting elements, and most of them were related to stress-induced gene expression, suggesting the putative role of essentially TdWRKY1, 3, 4, and 5 genes in wheat responses to a variety of environmental stresses. A TBLASTN on NCBI nucleotide collection database of the *TdWRKYs* 1.5 Kb cis-elements showed that *TdWRKY1*-UTR region shares 92% similarity from base 1 to 419 bp with *T. aestivum* 3B genome scaffold HG670306.1 from 28534545576 to 285345992. This 419 bp contains 4 W Box, 4 MYB, 5MYC, 4RAV, and 1 ABRE elements. 1.5 Kb *TdWRKY2*-UTR (632-949) region also shares 84% similarity with HG670306.1 (74734822-74734532). Elements which might be common to the orthologous regions are 3 Wbox, 4MYB, and 4MYC. 100% of homology was found between the 1.5 Kb from *TdWRKY5* putative cis-elements and its homolog HG670306 from coordinate 436276014 to coordinate 436275955. They share all the cis-elements on Table 2. Finally, 99% was shared between all UTR regions from *KC174859.1* (*TaWRKY53*) and 1.5 Kb from 736 to 1501. This region contains 4 Wbox, 6 MYB, 1 MYC, 4 GCC cores, 2 RAV, 2 LTREs, and 3 ABREs.





**FIGURE 5 | Expression pattern of *TdWRKY1*, *TdWRKY3*, *TdWRKY4*, and *TdWRKY5* in response to salt stress.** Expression profile of *TdWRKY1*, 3, 4, and 5 in leaves (A) and roots (B) from GR-salt tolerant and OR-salt sensitive durum wheat genotypes following 200 mM NaCl treatment. The levels of *TdWRKY1*, 3, 4, and 5 transcripts were assessed by real-time quantitative PCR. mRNA accumulation was monitored in 10-d-old roots and leaves, after 6 and 24 h of NaCl treatment (200 mM NaCl). For each sample, relative fold changes were determined by normalizing the CT value of the *TdWRKY1*, 3, 4, and 5 genes in different tissues in both sensitive and tolerant varieties to the CT value of *Td26S* (internal control) and calculating the ratios of transcripts abundance compared to samples from untreated tissues, using the formula  $2^{-\Delta\Delta CT}$ .  $\Delta\Delta CT$  refers to fold differences in *TdWRKY1*, 3, 4, and 5 expression relative to untreated tissues. Log2-transformed signal ratio was carried out. Values are means  $\pm$  SD ( $n \geq 30$  plants were pooled for RNA extraction) of three replicates. Asterisks \* are used to represent  $P$  values: \* $0.01 < P < 0.05$  (Statistically significant); \* $0.01 < P < 0.05$ ; \*\* $P < 0.01$  (Statistically highly significant) (Student's  $t$ -test).

Two hypotheses might be formulated. Firstly, genes that share the same acting elements on their promoters may function the same way and, secondly, homologous *cis*-acting elements of *TdWRKY1*, *TdWRKY2*, and *TdWRKY4* are assigned to the 3B chromosome in bread wheat, suggesting that *TdWRKY1*, 2, and 4 might be also accommodated within chromosome 3B or 3A from *T. turgidum*.

Promoter region analyses, gene structure within orthologous members, and phylogeny studies all indicated that *TdWRKYs* are novel members of the WRKY family in durum wheat and, given their high sequence homology with orthologous monocots and

*Arabidopsis*' known function, it can be postulated that they play similar roles in mediating responses to biotic and abiotic stresses.

### Differential Expression of *TdWRKYs*

In a context where no functional characterization has yet been carried out for WRKYs in durum wheat, our data showed that the *TdWRKY* genes were inducible by high-salt treatment. Moreover, *TdWRKY* expression in response to salt stress displayed distinctive patterns in two durum wheat genotypes with contrasting behavior regarding tolerance to abiotic stresses. In the tolerant GR variety, *TdWRKYs* were strongly induced by



salt stress within a few hours (6 h), while it was downregulated in the sensitive OR variety. Notably, differences between tolerant and sensitive genotypes were detected, mainly in the expression levels in tolerant genotype leaves at 24 h stress treatment. Peng et al. (2014) showed that *WRKY* members' unigenes were mostly up-regulated under salt stress in cotton (*Gossypium hirsutum* L.). They noted that some *WRKY* genes were expressed in the salt-tolerant genotype Earlistaple 7, but were repressed, weakly induced, or not induced at all in salt-sensitive Nan Dan Ba Di Da Hua, within 24 h. *HvWRKY2* was induced by salt stress in TR1 (Tolerant variety) but not in TS1 (Sensitive variety) (Li et al., 2014). Similarly, expression of wheat *TaWRKY2* and *TaWRKY19* was induced by salt, and both *TaWRKY2* and *TaWRKY19* enhanced salt tolerance in transgenic Arabidopsis plants compared with wild type (Niu et al., 2012). The induction pattern showed that the highest gene expression occurs at 3–6 h after salt stress initiation, for *TaWRKY2* and *TaWRKY19* and at 6 and 24 h for *HvWRKY2*, which is likely to be the same for *TdWRKY* members.

## CONCLUDING REMARKS

The durum wheat complexity and large genome size (13,000 Mb) have largely prevented the development of genomic resources. Meanwhile, efforts have focused on sequencing of target regions selected as covering one or more genes of interest, called locus. Several BAC clones covering the corresponding region in one or more wheat genomes were sequenced as well.

In this study, we targeted durum wheat BAC clones harboring *TdWRKY* genes potentially involved in response to salt stress. We validated six BAC clones relative to the genomic library TtuLDN65. The size of these clone inserts after sequencing varies between 120 and 190 Kb. The added value of such an approach is that we obtain the coding sequence with introns and promoter regions that are essential for expression or functional study. Furthermore, it enabled us to access the entire genomic

environment of the coding sequence that might provide new information about the structure, conservation, position, order and genomic dynamics. A structural study of the environment of a gene encoding a resistance or tolerance protein can bring a multitude of information that can enrich functional study.

In this study, we identified 5 *WRKY* genes potentially involved in the salt stress tolerance. This article reports the identification of 5 novel *TdWRKY* genes. Sequence comparison with orthologs from barley, wheat and Brachypodium provide valuable information for determining gene structure. *TdWRKY1*, *TdWRKY3*, *TdWRKY4*, and *TdWRKY5* gene sequences were highly conserved as well as exon-intron boundaries, even with Arabidopsis. These important structural similarities, between orthologous *WRKY* genes, are indicative of a potential functional conservation.

## AUTHOR CONTRIBUTIONS

HB, MB, AG initiated the project. FY, EM, WM performed experiments. FY performed analysis and interpretation of data for the work. FY wrote the paper. HB, WM revised the paper critically for important intellectual content.

## ACKNOWLEDGMENTS

This research was realized at the French Plant Genomic Resource Center (CNRGV) (INRA, Toulouse) and Laboratoire de Genomique et Biotechnologie des Fruits (GBF) (INRA, Toulouse). The authors would like to thank Philippe Leroy from (INRA, Clermont-Ferrand) for the Triannot pipelines training.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.02034/full#supplementary-material>

## REFERENCES

- Abe, H., Yamaguchi-Shinozaki, K., Urao, T., Iwasaki, T., Hosokawa, D., and Shinozaki, K. (1997). Role of Arabidopsis MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. *Plant Cell* 9, 1859–1868.
- Baker, S. S., Wilhelm, K. S., and Thomashow, M. F. (1994). The 5'-region of *Arabidopsis thaliana* cor15a has cis-acting elements that confer cold-, drought- and ABA-regulated gene expression. *Plant Mol. Biol.* 24, 701–713. doi: 10.1007/BF00029852
- Batada, N. N., Urrutia, A. O., and Hurst, L. D. (2007). Chromatin remodelling is a major source of coexpression of linked genes in yeast. *Trends Genet.* 23, 480–484. doi: 10.1016/j.tig.2007.08.003
- Bencke-Malato, M., Cabreira, C., Wiebke-Strohm, B., Bucker-Neto, L., Mancini, E., Osorio, M. B., et al. (2014). Genome-wide annotation of the soybean *WRKY* family and functional characterization of genes involved in response to *Phakopsora pachyrhizi* infection. *BMC Plant Biol.* 14:236. doi: 10.1186/s12870-014-0236-0
- Besseau, S., Li, J., and Palva, E. T. (2012). *WRKY54* and *WRKY70* co-operate as negative regulators of leaf senescence in *Arabidopsis thaliana*. *J. Exp. Bot.* 63, 2667–2679. doi: 10.1093/jxb/err450
- Bhattarai, K. K., Atamian, H. S., Kaloshian, I., and Eulgem, T. (2010). *WRKY72*-type transcription factors contribute to basal immunity in tomato and Arabidopsis as well as gene-for-gene resistance mediated by the tomato R gene Mi-1. *Plant J.* 63, 229–240. doi: 10.1111/j.1365-3113.2010.04232.x
- Bi, C., Xu, Y., Ye, Q., Yin, T., and Ye, N. (2016). Genome-wide identification and characterization of *WRKY* gene family in *Salix suchowensis*. *PeerJ* 4:2437. doi: 10.7717/peerj.2437
- Brenner, S., Elgar, G., Sandford, R., Macrae, A., Venkatesh, B., and Aparicio, S. (1993). Characterization of the pufferfish (Fugu) genome as a compact model vertebrate genome. *Nature* 366, 265–268. doi: 10.1038/366265a0
- Budak, H., Kantar, M., and Kurtoglu, K. Y. (2013). Drought tolerance in modern and wild wheat. *Sci. World J.* 13:16. doi: 10.1155/2013/548246
- Cenci, A., Chantret, N., Kong, X., Gu, Y., Anderson, O. D., Fahima, T., et al. (2003). Construction and characterization of a half million clone BAC library of durum wheat (*Triticum turgidum* ssp. *durum*). *Theor. Appl. Genet.* 107, 931–939. doi: 10.1007/s00122-003-1331-z
- Cenci, A., Somma, S., Chantret, N., Dubcovsky, J., and Blanco, A. (2004). PCR identification of durum wheat BAC clones containing genes coding for carotenoid biosynthesis enzymes and their chromosome localization. *Genome* 47, 911–917. doi: 10.1139/g04-033

- Chen, W. J., and Zhu, T. (2004). Networks of transcription factors with roles in environmental stress response. *Trends Plant Sci.* 9, 591–596. doi: 10.1016/j.tplants.2004.10.007
- Choulet, F., Wicker, T., Rustenholz, C., Paux, E., Salse, J., Leroy, P., et al. (2010). Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. *Plant Cell* 22, 1686–1701. doi: 10.1105/tpc.110.074187
- Cifarelli, R. A., D'Onofrio, O., Grillo, R., Mango, T., Cellini, F., Piarulli, L., et al. (2013). Development of a new wheat microarray from a durum wheat totipotent cDNA library used for a powdery mildew resistance study. *Cell. Mol. Biol. Lett.* 18, 231–248. doi: 10.2478/s11658-013-0086-z
- Ciolkowski, I., Wanke, D., Birkenbihl, R., and Somssich, I. E. (2008). Studies on DNA binding selectivity of WRKY transcription factors lend structural clues into WRKY domain function. *Plant Mol. Biol.* 68, 81–92. doi: 10.1007/s11103-008-9353-1
- Devos, K. M. (2010). Grass genome organization and evolution. *Curr. Opin. Plant Biol.* 13, 139–145. doi: 10.1016/j.pbi.2009.12.005
- Dong, J., Chen, C., and Chen, Z. (2003). Expression profiles of the Arabidopsis WRKY gene superfamily during plant defense response. *Plant Mol. Biol.* 51, 21–37. doi: 10.1023/A:1020780022549
- Dubcovsky, J., Ramakrishna, W., SanMiguel, P. J., Busso, C. S., Yan, L., Shiloff, B. A., et al. (2001). Comparative sequence analysis of colinear barley and rice bacterial artificial chromosomes1. *Plant Physiol.* 125, 1342–1353. doi: 10.1104/pp.125.3.1342
- Dubouzet, J. G., Sakuma, Y., Ito, Y., Kasuga, M., Dubouzet, E. G., Miura, S., et al. (2003). OsDREB genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. *Plant J.* 33, 751–763. doi: 10.1046/j.1365-3113X.2003.01661.x
- Dvorak, J. (1988). "Cytogenetical and molecular inferences about the evolution of wheat," in *7th International Wheat Genetics Symposium*, eds T. E. Miller and R. M. D. Koebner (Cambridge: Institute of Plant Science Research), 187–192.
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Eulgem, T., Rushton, P. J., Robatzek, S., and Somssich, I. E. (2000). The WRKY superfamily of plant transcription factors. *Trends Plant Sci.* 5, 199–206. doi: 10.1016/S1360-1385(00)01600-9
- Eulgem, T., Rushton, P. J., Schmelzer, E., Hahlbrock, K., and Somssich, I. E. (1999). Early nuclear events in plant defence signalling: rapid gene activation by WRKY transcription factors. *EMBO J.* 18, 4689–4699. doi: 10.1093/emboj/18.17.4689
- Fernandez-Calvin, B., and Orellana, J. (1994). Metaphase I-bound arms frequency and genome analysis in wheat-Aegilops hybrids. Similar relationships between the B genome of wheat and S and Sl genomes of *Ae. speltoideus*, *Ae. longissima* and *Ae. sharonensis*. *Theor. Appl. Genet.* 88, 1043–1049. doi: 10.1007/BF00220814
- Flutre, T., Duprat, E., Feuillet, C., and Quesneville, H. (2011). Considering transposable element diversification in de novo annotation approaches. *PLoS ONE* 6:e16526. doi: 10.1371/journal.pone.0016526
- Gorham, J., Jones, R. G., and Bristol, A. (1990). Partial characterization of the trait for enhanced K<sup>+</sup>-Na<sup>+</sup> discrimination in the D genome of wheat. *Planta* 180, 590–597. doi: 10.1007/BF02411458
- Guillaumie, S., Mzid, R., Méchin, V., Léon, C., Hichri, I., Destrac-Irvine, A., et al. (2010). The grapevine transcription factor WRKY2 influences the lignin pathway and xylem development in tobacco. *Plant Mol. Biol.* 72, 215–234. doi: 10.1007/s11103-009-9563-1
- Hagen, G., and Guilfoyle, T. (2002). Auxin-responsive gene expression: genes, promoters and regulatory factors. *Plant Mol. Biol.* 49, 373–385. doi: 10.1023/A:1015207114117
- Hechenberger, M., Schwappach, B., Fischer, W. N., Frommer, W. B., Jentsch, T. J., and Steinmeyer, K. (1996). A family of putative chloride channels from *Arabidopsis* and functional complementation of a yeast strain with a *CLC* gene disruption. *J. Biol. Chem.* 271, 33632–33638. doi: 10.1074/jbc.271.52.33632
- Huo, N., Vogel, J. P., Lazo, G. R., You, F. M., Ma, Y., McMahon, S., et al. (2009). Structural characterization of Brachypodium genome and its syntenic relationship with rice and wheat. *Plant Mol. Biol.* 70, 47–61. doi: 10.1007/s11103-009-9456-3
- Hurst, L. D., Pál, C., and Lercher, M. J. (2004). The evolutionary dynamics of eukaryotic gene order. *Nat. Rev. Genet.* 5, 299–310. doi: 10.1038/nrg1319
- Janga, S. C., Collado-Vides, J., and Babu, M. M. (2008). Transcriptional regulation constrains the organization of genes on eukaryotic chromosomes. *PNAS* 105, 15761–15766. doi: 10.1073/pnas.0806317105
- Jiang, Y., and Deyholos, M. K. (2009). Functional characterization of Arabidopsis NaCl-inducible WRKY25 and WRKY33 transcription factors in abiotic stresses. *Plant Mol. Biol.* 69, 91–105. doi: 10.1007/s11103-008-9408-3
- Jiang, Y., Duan, H., Yin, J., Ye, S., Zhu, J., and Zhang, F. (2014). Genome-wide identification and characterization of the Populus WRKY transcription factor family and analysis of their expression in response to biotic and abiotic stresses. *J. Exp. Bot.* 65, 6629–6644. doi: 10.1093/jxb/eru381
- Kagaya, Y., Ohmiya, K., and Hattori, T. (1999). RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA binding domains uniquely found in higher plants. *Nucleic Acids Res.* 27, 470–478. doi: 10.1093/nar/27.2.470
- Knob, C., Ringler, J., Dangel, J. L., and Eulgem, T. (2007). Arabidopsis WRKY70 is required for full RPP4-mediated disease resistance and basal defense against *Hyaloperonospora parasitica*. *Mol. Plant Microbe Interact.* 20, 120–128. doi: 10.1094/MPMI-20-2-0120
- Kronmiller, B. A., and Wise, R. P. (2008). TE nest: automated chronological annotation and visualization of nested plant transposable elements. *Plant Physiol.* 146, 45–59. doi: 10.1104/pp.107.110353
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404
- Leroy, P., Guilhot, N., Sakai, H., Bernard, A., Choulet, F., Theil, S., et al. (2012). TriAnnot: a versatile and high performance pipeline for the automated annotation of plant genomes. *Front. Plant Sci.* 3:5. doi: 10.3389/fpls.2012.00005
- Li, H., Guo, Q., Lan, X., Zhou, Q., and Wei, N. (2014). Comparative expression analysis of five WRKY genes from Tibetan hulless barley under various abiotic stresses between drought-resistant and sensitive genotype. *Acta Physiol. Plant.* 36, 963–973. doi: 10.1007/s11738-013-1475-7
- Ling, J., Jiang, W., Zhang, Y., Yu, H., Mao, Z., Gu, X., Huang, S., et al. (2011). Genome-wide analysis of WRKY gene family in *Cucumis sativus*. *BMC Genomics* 12:471. doi: 10.1186/1471-2164-12-471
- Liu, D., Leib, K., Zhao, P., Kogel, K. H., and Langen, G. (2014). Phylogenetic analysis of barley WRKY proteins and characterization of HvWRKY1 and -2 as repressors of the pathogen-inducible gene HvGERAc. *Mol. Genet. Genomics* 289, 1331–1345. doi: 10.1007/s00438-014-0893-6
- Liu, R., Vitte, C., Ma, J., Mahama, A. A., Dhlwayo, T., Lee, M., et al. (2007). A GeneTrek analysis of the maize genome. *Proc. Natl. Acad. Sci. U.S.A.* 104, 11844–11849. doi: 10.1073/pnas.0704258104
- Liu, X., and Han, B. (2009). Evolutionary conservation of neighbouring gene pairs in plants. *Gene* 437, 71–79. doi: 10.1016/j.gene.2009.02.012
- Liu, Z. Q., Lu, Y., Wu, Z., Mei, C., Lu, K., and Yu, T. O. (2012). Cooperation of three WRKY-domain transcription factors WRKY18, WRKY40, and WRKY60 in repressing two ABA-responsive genes ABI4 and ABI5 in Arabidopsis. *J. Exp. Bot.* 63, 6371–6392. doi: 10.1093/jxb/ers293
- Maas, E. V., and Grieve, C. M. (1990). Spike and leaf development in salt-stressed wheat. *Crop Sci.* 30, 1309–1313. doi: 10.2135/cropsci1990.0011183X003000060031x
- Maeo, K., Hayashi, S., Kojima-Suzuki, H., Morikami, A., and Nakamura, D. (2001). Role of conserved residues of the WRKY domain in the DNA-binding of tobacco WRKY family proteins. *Biosci. Biotechnol. Biochem.* 65, 2428–2436. doi: 10.1271/bbb.65.2428
- Makhlofi, E., Yousfi, F. E., Marande, W., Mila, I., Hanana, M., Bergès, H., et al. (2014). Isolation and molecular characterization of ERF1, an ethylene response factor gene from durum wheat (*Triticum turgidum* L. subsp. durum), potentially involved in salt-stress responses. *J. Exp. Bot.* 65, 6359–6371. doi: 10.1093/jxb/eru352
- Marè, C., Mazzucotelli, E., Crosatti, C., Francia, E., Stanca, A. M., and Cattivelli, L. (2004). Hv-WRKY38: a new transcription factor involved in cold- and drought-response in barley. *Plant Mol. Biol.* 55, 399–416. doi: 10.1007/s11103-004-0906-7
- Marshall, O. J. (2004). PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR. *Bioinformatics* 20, 2471–2472. doi: 10.1093/bioinformatics/bth254
- Mingyu, Z., Zhengbin, Z., Shouyi, C., Jinsong, Z., and Hongbo, S. (2012). WRKY transcription factor superfamily: structure, origin

- and functions. *Afr. J. Biotechnol.* 11, 8051–8059. doi: 10.5897/AJB11.549
- Monneveux, P., Zaharieva, M., and Rekika, D. (2000). “The utilisation of *Triticum* and *Aegilops* species for the improvement of durum wheat,” in *Presented at Durum Wheat Improvement in the Mediterranean Region: New Challenges* Zaragoza (2000-04-12 – 2000-04-14). Available online at: <http://prodnra.inra.fr/record/158629>
- Murashige, T., and Skoog, F. (1962). A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol. Plant.* 15, 473–497. doi: 10.1111/j.1399-3054.1962.tb08052.x
- Nieva, C., Busk, P. K., Domínguez-Puigianer, V., Lumbreras, P. S., Testillano, M. C., and Risueño, M. (2005). Isolation and functional characterisation of two new bZIP maize regulators of the ABA responsive gene *rab28*. *Plant Mol. Biol.* 58, 899–914. doi: 10.1007/s11103-005-8407-x
- Niu, C. F., Wei, W., Zhou, Q. Y., Tian, A. G., Hao, Y. J., Zhang, W. K., et al. (2012). Wheat WRKY genes TaWRKY2 and TaWRKY19 regulate abiotic stress tolerance in transgenic *Arabidopsis* plants. *Plant Cell. Environ.* 35, 1156–1170. doi: 10.1111/j.1365-3040.2012.02480.x
- Peng, Z., He, S., Gong, W., Sun, J., Pan, Z., and Xu, F. (2014). Comprehensive analysis of differentially expressed genes and transcriptional regulation induced by salt stress in two contrasting cotton genotypes. *BMC Genomics* 15:760. doi: 10.1186/1471-2164-15-760
- Podell, S., and Gribskov, M. (2004). Predicting N-terminal myristoylation sites in plant proteins. *BMC Genomics* 5:e37. doi: 10.1186/1471-2164-5-37
- Quesneville, H., Bergman, C. M., Andrieu, O., Autard, D., Nouaud, D., Ashburner, M., et al. (2005). Combined evidence annotation of transposable elements in genome sequences. *PLoS Comput. Biol.* 1:e22. doi: 10.1371/journal.pcbi.0010022
- Rushton, P. J., Somssich, I. E., Ringler, P., and Shen, Q. J. (2010). WRKY transcription factors. *Trends Plant Sci.* 15, 247–258. doi: 10.1016/j.tplants.2010.02.006
- Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science* 326, 1112–1115. doi: 10.1126/science.1178534
- Shinozaki, K., and Dennis, E. S. (2003). Cell signalling and gene regulation: global analyses of signal transduction and gene expression profiles. *Curr. Opin. Plant Biol.* 6, 405–409. doi: 10.1016/S1369-5266(03)00093-1
- Skinner, J. S., vonZitzewitz, J., Szucs, P., Marquez-Cedillo, L., Filichkin, T., Amundsen, K., et al. (2005). Structural, functional, and phylogenetic characterization of a large CBF gene family in barley. *Plant Mol. Biol.* 59, 533–551. doi: 10.1007/s11103-005-2498-2
- Svensson, J. T., Crosatti, C., Campoli, C., Bassi, R., Stanca, A. M., Close, T. J., et al. (2006). Transcriptome analysis of cold acclimation in barley albina and xantha mutants. *Plant Physiol.* 141, 257–270. doi: 10.1104/pp.105.072645
- Tamura, K., Stecher, G., Peterson, D., Filipi, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- The Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815. doi: 10.1038/35048692
- The International Brachypodium Initiative (2010). The International Brachypodium Initiative. *Nature* 463, 763–768.
- The International Wheat Genome Sequencing Consortium (IWGSC) (2014). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* 345:1251788. doi: 10.1126/science.1251788
- Towler, D. A., Gordon, J. I., Adams, S. P., and et Glaser, L. (1988). The biology and enzymology of eukaryotic protein acylation. *Annu. Rev. Biochem.* 57, 69–99. doi: 10.1146/annurev.bi.57.070188.000441
- Tripathi, P., Rabara, R. C., Langum, T. J., Boken, A. K., Rushton, D. L., Boomsma, D. D., et al. (2012). The WRKY transcription factor family in *Brachypodium distachyon*. *BMC Genomics* 13:270. doi: 10.1186/1471-2164-13-270
- Ulm, R., Ichimura, K., Mizoguchi, T., Peck, S. C., Zhu, T., Wang, X., et al. (2002). Distinct regulation of salinity and genotoxic stress responses by *Arabidopsis* MAP kinase phosphatase 1. *EMBO J.* 21, 6483–6493. doi: 10.1093/emboj/cdf646
- Veras, A. A., de Sá, P. H., Azevedo, V., Silva, A., and Ramos, R. T. J. (2013). AutoAssemblyD: a graphical user interface system for several genome assemblers. *Bioinformatics* 9, 840–841. doi: 10.6026/97320630009840
- Vitte, C., and Bennetzen, J. L. (2006). Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. *Proc. Natl. Acad. Sci.* 103, 17638–17643. doi: 10.1073/pnas.0605618103
- Wei, K., Chen, J., Chen, Y. F., Wu, L. J., and Xie, D. O. (2012). Molecular phylogenetic and expression analysis of the complete WRKY transcription factor family in maize. *DNA Res.* 19, 153–164. doi: 10.1093/dnares/dsr048
- Wei, W., Hu, Y., Han, Y. T., Zhang, K., Zhao, F. L., and Feng, J. Y. (2016). The WRKY transcription factors in the diploid woodland strawberry *Fragaria vesca*: identification and expression analysis under biotic and abiotic stresses. *Plant Physiol. Biochem.* 105, 129–144. doi: 10.1016/j.plaphy.2016.04.014
- Wicker, T., Matthews, D. E., and Keller, B. (2002). TREP: a database for Triticeae repetitive elements. *Trends Plant Sci.* 7, 561–562. doi: 10.1016/S1360-1385(02)02372-5
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., et al. (2007). A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8, 973–982. doi: 10.1038/nrg2165
- Wicker, T., Zimmermann, W., Perovic, D., Paterson, A. H., Ganai, M., Graner, A., et al. (2005). A detailed look at 7 million years of genome evolution in a 439 kb contiguous sequence at the barley Hv-eIF4E locus: recombination, rearrangements and repeats. *Plant J.* 41, 184–194. doi: 10.1111/j.1365-313X.2004.02285.x
- Wu, L., Chen, X., Ren, H., Zhang, Z., Zhang, H., Wang, J., et al. (2007). ERF protein JERF1 that transcriptionally modulates the expression of abscisic acid biosynthesis-related gene enhances the tolerance under salinity and cold in tobacco. *Planta* 226, 815–825. doi: 10.1007/s00425-007-0528-9
- Xie, Z., Zhang, Z. L., Zou, X., Huang, J., Ruas, P., Thompson, D., et al. (2005). Annotations and functional analyses of the rice WRKY gene superfamily reveal positive and negative regulators of abscisic acid signaling in aleurone cells. *Plant Physiol.* 137, 176–189. doi: 10.1104/pp.104.054312
- Yamaguchi-Shinozaki, K., and Shinozaki, K. (2005). Organization of cis-acting regulatory elements in osmotic- and cold-stress-responsive promoters. *Trends Plant Sci.* 10, 88–94. doi: 10.1016/j.tplants.2004.12.012
- Yamamoto, S., Nakano, T., Suzuki, K., and et Shinshi, H. (2004). Elicitor-induced activation of transcription via W box-related cis-acting elements from a basic chitinase gene by WRKY transcription factors in tobacco. *Biochim. Biophys. Acta* 1679, 279–287. doi: 10.1016/j.bbaexp.2004.07.005
- Yamasaki, K., Kigawa, T., Inoue, M., Tatenno, M., Yamasaki, T., Yabuki, T., et al. (2005). Solution structure of an Arabidopsis WRKY DNA binding domain. *Plant Cell* 17, 944–956. doi: 10.1105/tpc.104.026435
- Yang, P., Chen, C., Wang, Z., Fan, B., and et Chen, Z. (1999). Pathogen- and salicylic acid induced WRKY DNA-binding activity recognizes the elicitor response element of the tobacco class I chitinase gene promoter. *Plant J.* 18, 141–149. doi: 10.1046/j.1365-313X.1999.00437.x
- Zhang, Y., and Wang, L. (2005). The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants. *BMC Evol. Biol.* 5:1. doi: 10.1186/1471-2148-5-1
- Zou, C., Jiang, W., and Yu, D. (2010). Male gametophyte-specific WRKY34 transcription factor mediates cold sensitivity of mature pollen in Arabidopsis. *J. Exp. Bot.* 61, 3901–3914. doi: 10.1093/jxb/erq204

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Yousfi, Makhoulfi, Marande, Ghorbel, Bouzayen and Bergès. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# De novo Assembly and Characterization of the Transcriptome of Broomcorn Millet (*Panicum miliaceum* L.) for Gene Discovery and Marker Development

Hong Yue<sup>††</sup>, Le Wang<sup>††</sup>, Hui Liu<sup>1</sup>, Wenjie Yue<sup>1</sup>, Xianghong Du<sup>1</sup>, Weining Song<sup>1,2,3\*</sup> and Xiaojun Nie<sup>1,2\*</sup>

## OPEN ACCESS

### Edited by:

Gautam Sarath,  
United States Department of  
Agriculture – Agricultural Research  
Service, USA

### Reviewed by:

Teresa Donze,  
University of Nebraska Lincoln, USA  
Erin D. Scully,  
United States Department of  
Agriculture – Agricultural Research  
Service, USA

### \*Correspondence:

Weining Song  
sweining2002@yahoo.com  
Xiaojun Nie  
small@nwsuaf.edu.cn

<sup>††</sup> These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Crop Science and Horticulture,  
a section of the journal  
Frontiers in Plant Science

**Received:** 31 March 2016

**Accepted:** 08 July 2016

**Published:** 21 July 2016

### Citation:

Yue H, Wang L, Liu H, Yue W, Du X,  
Song W and Nie X (2016) De novo  
Assembly and Characterization of the  
Transcriptome of Broomcorn Millet  
(*Panicum miliaceum* L.) for Gene  
Discovery and Marker Development.  
Front. Plant Sci. 7:1083.  
doi: 10.3389/fpls.2016.01083

<sup>1</sup> College of Agronomy, Northwest A&F University, Yangling, China, <sup>2</sup> State Key Laboratory of Crop Stress Biology in Arid Areas, Northwest A&F University, Yangling, China, <sup>3</sup> Australia-China Joint Research Centre for Abiotic and Biotic Stress Management in Agriculture, Horticulture and Forestry, Northwest A&F University, Yangling, China

Broomcorn millet (*Panicum miliaceum* L.) is one of the world's oldest cultivated cereals, which is well-adapted to extreme environments such as drought, heat, and salinity with an efficient C4 carbon fixation. Discovery and identification of genes involved in these processes will provide valuable information to improve the crop for meeting the challenge of global climate change. However, the lack of genetic resources and genomic information make gene discovery and molecular mechanism studies very difficult. Here, we sequenced and assembled the transcriptome of broomcorn millet using Illumina sequencing technology. After sequencing, a total of 45,406,730 and 51,160,820 clean paired-end reads were obtained for two genotypes Yumi No. 2 and Yumi No. 3. These reads were mixed and then assembled into 113,643 unigenes, with the length ranging from 351 to 15,691 bp, of which 62,543 contigs could be assigned to 315 gene ontology (GO) categories. Cluster of orthologous groups and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses assigned could map 15,514 unigenes into 202 KEGG pathways and 51,020 unigenes to 25 COG categories, respectively. Furthermore, 35,216 simple sequence repeats (SSRs) were identified in 27,055 unigene sequences, of which trinucleotides were the most abundant repeat unit, accounting for 66.72% of SSRs. In addition, 292 differentially expressed genes were identified between the two genotypes, which were significantly enriched in 88 GO terms and 12 KEGG pathways. Finally, the expression patterns of four selected transcripts were validated through quantitative reverse transcription polymerase chain reaction analysis. Our study for the first time sequenced and assembled the transcriptome of broomcorn millet, which not only provided a rich sequence resource for gene discovery and marker development in this important crop, but will also facilitate the further investigation of the molecular mechanism of its favored agronomic traits and beyond.

**Keywords:** abiotic stress, broomcorn millet, transcriptome, qRT-PCR, SSR



## INTRODUCTION

Broomcorn millet (*Panicum miliaceum* L.) is one of the earliest domesticated cereals worldwide with historical, evolutionary and agricultural significance (Yang et al., 2012; Stephens et al., 2014). It has been revealed that Broomcorn millet was domesticated as early as 10,000 years ago in the semiarid regions of China (Lu et al., 2009), and it was an indispensable food staple in many semiarid regions of East Asia before the cultivation of rice and wheat (Hu et al., 2008). Thus, Broomcorn millet played a vital impact on human civilization (Crawford, 2006; Bellwood et al., 2007). Genetically, Broomcorn millet is a tetraploid species with a chromosome number of 36 ( $2N = 4X = 36$ ; Hunt et al., 2010). As a short-day C4 crop, it has many favored agronomic traits, such as high productivity yield, short growing season (60–90 days), a lower water requirement and is well-adapted to extreme conditions (Hunt et al., 2014). Previous studies revealed that Broomcorn millet showed high drought and salinity tolerance (Dai et al., 2011). Thus, discovery and identification of the stress-responsive genes from broomcorn millet will not only provide useful information for better understanding the molecular mechanism of stress tolerance, but also provide indispensable gene resources for tolerance improvement in this species as well as other crops (Jiang and Deyholos, 2006).

With the development of next generation sequence (NGS) technology, RNA sequencing (RNA-seq) has gradually become a powerful and high-efficiency method to obtain a large number of transcripts and identify differentially expressed genes (DEGs) at the transcriptome level, which has been widely used in various plants (Clark et al., 2015; Sartelet et al., 2015; Xu et al., 2015; Zhan et al., 2015). Prior to this study, transcriptome analysis has not been performed in broomcorn millet, which has limited further gene identification and molecular mechanism studies (Miller et al., 2002; Salgado et al., 2014; Yates et al., 2014). Here, *de novo* assembly and analysis of the broomcorn millet transcriptome was performed using High-throughput Illumina paired-end RNA sequencing technology, with the purpose to enrich the genetic information and sequence resources for facilitating gene discovery and marker development studies. This is the first study to report the transcriptome characteristics of broomcorn millet, which will provide the useful information for molecular studies, and also shed light on the molecular mechanism of stress tolerance in Broomcorn millet.

## MATERIALS AND METHODS

### Plant Materials and RNA Extraction

Two Broomcorn millet genotypes, namely Yumi No. 2 and Yumi No. 3, which were kindly provided by Dr. Bai-Li Feng, College of Agriculture, Northwest A&F University, were used as materials in this study. Yumi No. 2 is a waxy, drought-sensitive cultivar, while Yumi No. 3 is a non-waxy, drought and salt tolerant cultivar. Seeds of both cultivars were germinated and grown in pots containing peat mixed sand at a 1:1 ratio. Plants were normally watered and grown under glasshouse conditions (22°C, 16 h photoperiod/20°C, 8 h dark period). For RNA-Seq library

construction, tissue samples including leaves, stems, spikes, and roots were collected from five plant individuals grown at different time points. Leaves and roots were harvested at the seedling, jointing, booting, and filling stages while stems were harvested at the jointing, booting, filling, and mature stages. Spikes were harvested at the mature stages. Three independent tissue samples collection were performed as biological replicates. All the samples were immediately frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  for RNA isolation.

Total RNA from collected samples was separately isolated using TRIzol reagent (Invitrogen) according to the manufacturer's instructions. Then, the quality of RNA was checked by agarose gel electrophoresis and Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and the quantity was checked by NanoDropND-1000 Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). Finally, the equal amounts of RNA isolated from Yumi No. 2 and Yumi No. 3 tissues were pooled together into a single RNA sample respectively, and then the pooled RNA sample was used for RNA sequencing.

### Library Construction and Illumina Sequencing

The RNA-seq library construction and sequencing were performed using the Illumina's standard pipeline (Illumina, San Diego, CA, USA). In brief, magnetic oligospheres were used to remove rRNA or tRNA and enrich mRNA. mRNA was further sheared into short fragments of 180 bp in size were recovered and column purified, then subjected to first strand cDNA synthesis using random hexamer-primed synthesis of first-strand cDNA, followed by second-strand cDNA synthesis using DNA polymerase I and RNase H (Invitrogen). The cDNA fragments were end-repaired and A-tailed, and then index adapters were ligated. After the purified cDNA libraries were amplified by polymerase chain reaction (PCR) for 15 cycles and PCR products were separated by Certified Low Range Ultra Agarose, the suitable fragments were selected for deep sequencing. Sequencing was performed on the IlluminaHiSeq2500 platform with the PE100 approach by Shanghai Majorbio Bio-pharm Technology Corporation. All clean Illumina sequencing data have been deposited in SRA database with the accession number of SAMN05255231.

### Sequence Data Analysis and *De novo* Assembly

The quality of the raw reads of two libraries was checked using SeqPrep and Sickle software. Adapter contamination, low quality bases ( $\leq 20$ ), reads containing more than 10% ambiguous bases, and reads containing less than 20 bases were removed. The obtained clean reads were mixed and then *de novo* assembled using the Trinity program with the default parameters<sup>1</sup>. K was set to 25 for inchworm analysis and the transcript isoform with the highest relative abundance after butterfly analysis was selected as the representative transcript for each gene.

<sup>1</sup><http://trinityrnaseq.sourceforge.net>

Clustering analysis with CD-HIT software, with 95% identity, was performed to reduce redundancy from transcripts derived from homeologous genes or different alleles of the same genes, which are common artifacts of pooling tissues from multiple individuals for transcriptome sequencing.

## Functional Annotation

To determine the predicted function, all unigenes were used as blastx queries against the following databases with an  $e$ -value  $< 1e-5$ , including NCBI Non-redundant Protein (NR) database (2016/1/12), Swiss-Prot protein database (UniProtKB/Swiss-Prot protein knowledgebase release 2015\_12 statistics) and Cluster of Orthologous Groups database (COG) as well as Kyoto Encyclopedia of Genes and Genomes pathway database (KEGG). To evaluate sequences similarity and predicted function according to homologous genes, three close relative species of broomcorn millet, including *Panicum halli*, *Panicum virgatum* and *Setaria italica*, were used to analyze. The genome annotation sequences of *P. halli*<sup>2</sup> and *P. virgatum*<sup>3</sup> were obtained from JGI phytozome. The genome annotation sequences of *S. italica* were obtained from the Foxtail millet Database<sup>4</sup>. The hits with the highest sequence similarity were retrieved for analysis. Based on NR annotation, 10 top-hit species were identified and gene ontology (GO) classifications were annotated by the Blast2GO program. KEGG produced annotation of metabolic pathways. Goatools<sup>5</sup> and KOBAS (Xie et al., 2011) was used to identify enriched GO and KEGG in the DEGs between Yumi No. 2 and Yumi No. 3. DEGs were identified using Fisher's exact test and  $P$ -values were corrected for multiple hypothesis testing.

## SSRs and SNPs Markers Screening

Potential simple sequence repeats (SSRs) were detected using MISA software<sup>6</sup>. In this study, repeats of one to six nucleotides in length were considered. The minimum iterations units were 10 repeat units for mononucleotides, six for dinucleotides, and four for tri-, tetra-, penta-, and hexa-nucleotides. The maximal distance was 100 nucleotides interrupting two SSRs in a compound microsatellite. Assembled contigs were scanned for single nucleotide polymorphisms (SNPs) with SNP detection software SOAPsnp (Li et al., 2009) using the method as described by Chen et al. (2014), which has been used for SNP calling in *Stevia rebaudiana* with low quality mapping scores ( $\leq 20$ ) and less than five nucleotides apart were discarded.

## Analysis and Annotation Differentially Expressed Genes (DEGs)

Bowtie was used to perform the read mapping. The number of reads per kilobase of exon region per million mapped reads (RPKM) was used to normalize the expression values of reads.

Raw read counts were utilized with edgeR<sup>7</sup> to identify DEGs. In multiple hypothesis testing, false discovery rate (FDR) was used to select the threshold  $P$ -value (Benjamini et al., 2001). An  $FDR \leq 0.05$  and fold change (FC) ratio larger than 2 ( $|\log_2 FC| \geq 1$ ) were chosen in our study to determine the DEGs. Scatter and volcano plots were drawn by geWorkbench platform<sup>8</sup> using the value of logRPKM and log<sub>2</sub>FC. All DEGs were subjected to GO and KEGG annotation and cluster analysis of DEG patterns was performed using hCluster software<sup>9</sup> following the method previously described by Liu G. et al. (2012).

## qRT-PCR Analysis

First, Broomcorn millet plants, including Yumi No. 2 and Yumi No. 3, were adequately watered and grown under glasshouse conditions (22°C, 16 h photoperiod/20°C, 8 h dark period). After grown for 1 month, the plants were transferred into 4°C, 38°C or 200 mM NaCl, which represented low temperature, heat and salt treatment, respectively (Li et al., 2014). Whole plants were separately collected at 0, 3, 6, 12, and 24 h under above treatment as well as under no stress treatment as control. The experiment was conducted three independent times with five biological replicate seedlings per treatment. Tissue from the five replicate samples was pooled for RNA isolation using the RNAiso reagent (TaKaRa, Japan) and the three experiments were treated as biological replicates for the RNA extraction, and then equivalent amounts of RNA from three biological replicates of each sample was pooled into a single RNA. cDNA was prepared by using the PrimeScript™ RT reagent Kit (TaKaRa, Japan) following the manufacturer's instructions. Quantitative reverse transcription polymerase chain reaction (qRT-PCR) was analyzed using SYBR Green master mix (TaKaRa, Japan) and ABI 7300 real-time PCR system (Applied Biosystems, USA). The thermal cycling conditions were as follows: 3 min at 94°C, followed by 40 cycles each consisting of 95°C for 15 s, 60°C for 30 s, 72°C for 1 min. *Actin* was used as an internal control (Supplementary Table S1). Each reaction was performed in triplicate and the  $2^{-\Delta\Delta Ct}$  method was used to calculate the expression levels. Student's  $t$ -test was used to statistics analyze.

## RESULTS AND DISCUSSION

### Transcriptome Sequencing and *De novo* Assembly

To comprehensively generate a broomcorn millet transcriptome, the RNA isolated from leaves, stem, root, shoots, flower, and spike of two cultivars were equally pooled and sequenced separately using the Illumina HiSeq2000 platform. After sequencing and quality filtering, a total of 45,406,730 high-quality reads for Yumi No. 2 and 51,160,820 reads for Yumi No. 3 were obtained, accounting for approximately 4.3 and 4.9 Gb, respectively (Supplementary Table S2). Then, the generated reads were mixed and *de novo* assembled using Trinity software. A total of 113,643

<sup>2</sup>[https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org\\_Phalli](https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Phalli)

<sup>3</sup>[https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org\\_Pvirgatum](https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Pvirgatum)

<sup>4</sup><http://foxtailmillet.genomics.org.cn/page/species/download.jsp>

<sup>5</sup><https://github.com/tanghaibao/Goatools>

<sup>6</sup><http://pgrc.ipk-gatersleben.de/misa/>

<sup>7</sup><http://www.bioconductor.org/packages/release/bioc/html/edgeR.html>

<sup>8</sup><https://cabig.nci.nih.gov/community/tools/geWorkbench>

<sup>9</sup><https://pypi.python.org/pypi/hcluster/>

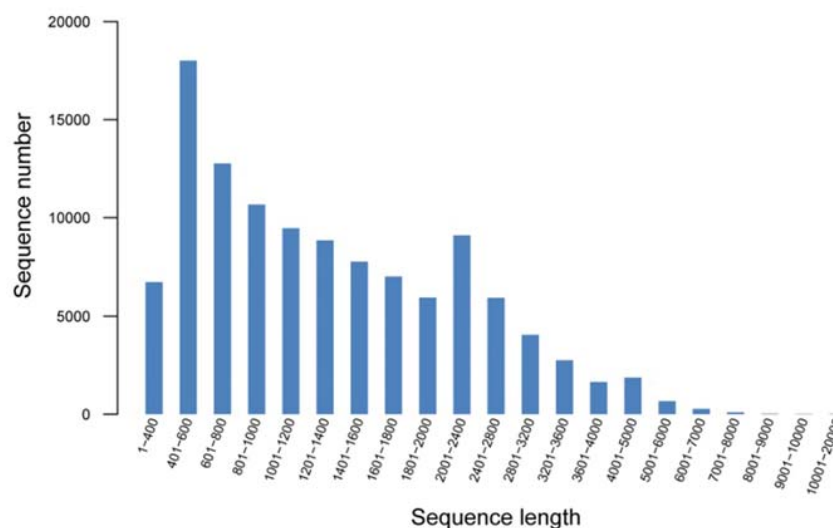
contigs with a total assembly length of 164,535,293 bp were obtained. To investigate the quality and coverage of assembly, we then mapped the paired reads of the two cultivars back to these contigs. Results showed that 132,599,016 (93.80%) and 113,424,962 (93.29%) reads belonging to Yumi No. 2 and Yumi No. 3 respectively could be mapped back to the contigs, supporting the accuracy of the assembly. Further analysis found that the size distribution of these contigs ranged from 351 to 15,691 bp with an average size of 1,448 bp, of which 24,750 (21.78%) ranged from 100 to 600 bp, and 23,423 (20.61%) ranged from 600 to 1000 bp; 65,470 unigenes with lengths longer than 1 kb were identified (**Figure 1**). To the best of our knowledge, the sequences of broomcorn millet are deficient at present (Hunt et al., 2011). Therefore, the unigene dataset reported here significantly enriches the sequence resources and genetic information of broomcorn millet, which provides the foundation for further study of gene expression, gene function and gene regulation pathways in broomcorn millet.

## Annotation and Functional Characterization of the Broomcorn Millet Transcriptome

In order to assess and annotate the assembled unigenes, 113,643 assembled unigenes generated by Trinity software were subjected to blastx similarity searches against NCBI's NR database, Swiss-Prot, COG, KEGG of proteins with an e-value cut off of  $10^{-5}$ . As a result, 51,629 unigenes had matches to proteins in the NR database, and 40,407 unigenes were similar to proteins in the Swiss-Prot database. Totally, 60,352 unigenes were profitably annotated in Nr, Swiss-Prot, GO, COG, and KEGG. The size distribution of these open reading frames was shown in Supplementary Figure S1. Furthermore, we examined the homologs of broomcorn millet unigenes with other monocot species, including *Sorghum bicolor*, *Oryza sativa*, *Brachypodium*

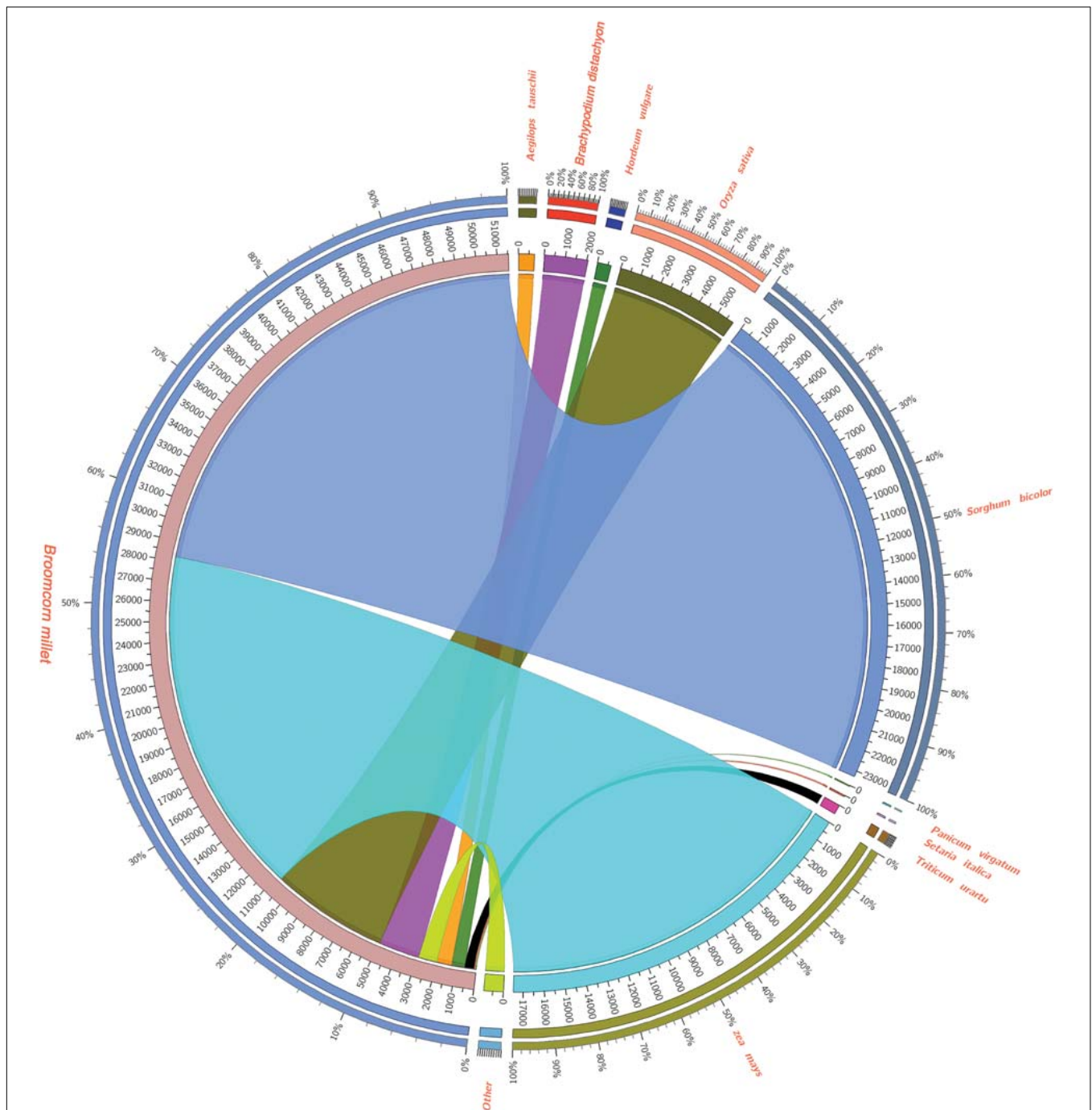
*distachyon*, *Hordeum vulgare*, *Aegilops tauschii*, *Triticum urartu*, *P. virgatum*, *S. italica* by protein similarity search in NCBI NR database. Results showed that 35,888 (43.68%) unigenes of broomcorn millet have homologous matches to *S. bicolor* transcripts, 28,537 (34.74%) have hits to *Zea mays*, 9,194 (11.19%) have hits to *O. sativa*, 2,995 (3.65%) have hits to *B. distachyon*, 1,247 (1.52%) have hits to *H. vulgare*, 1,190 (1.45%) have hits to *A. tauschii*, 742 (0.90%) have hits to *Triticum urartu*, 73 (0.14%) have hits to *S. italica*, and 50 (0.09%) have hits to *P. virgatum* as well as 921 (1.78%) have hits to other species, respectively (**Figure 2**). It is unexpected that only a few Broomcorn millet unigenes showed homologous significantly matches to *S. italica* and *P. virgatum*. Further analysis found this might be as the result of the best hits with better e-values and similarity distributions in *S. bicolor*, *Zea mays*, and *O. sativa* compared to *P. halli*, *P. virgatum*, and *S. italica*. To further evaluate accurately sequences similarity with *P. halli*, *P. virgatum* and *S. italica*, the predicted protein sequences of these three species were used to identify the homologs in broomcorn millet unigenes by blastx search with an e-value cutoff of  $10^{-5}$ . Results showed that 88,538 (77.91%) unigenes of broomcorn millet have homologs hits to *P. halli* genome annotations, 90,229 (79.40%) had hits to *P. virgatum* and 84,024 (73.94%) had hits to *S. italica*, respectively. At the same time, e-value and similarity distribution through against Nr databases, *P. halli*, *P. virgatum*, and *S. italica* genome annotations was shown in Supplementary Figure S2.

To preliminarily understand the function of the broomcorn millet unigenes, those with homologs to previously annotated sequences in the NR database were further annotated with GO terms using the Blast2GO tool (Conesa et al., 2005). A total of 62,543 unigene sequences were annotated to three major GO classes. The largest class was cellular component, accounting for 42.47% of the total annotated unigenes, followed by biological process (38.93%) and molecular function (18.60%).



**FIGURE 1 |** Size distribution of all assembled unigenes using the combined reads of common millet cultivars Yumi No. 2 and Yumi No. 3.



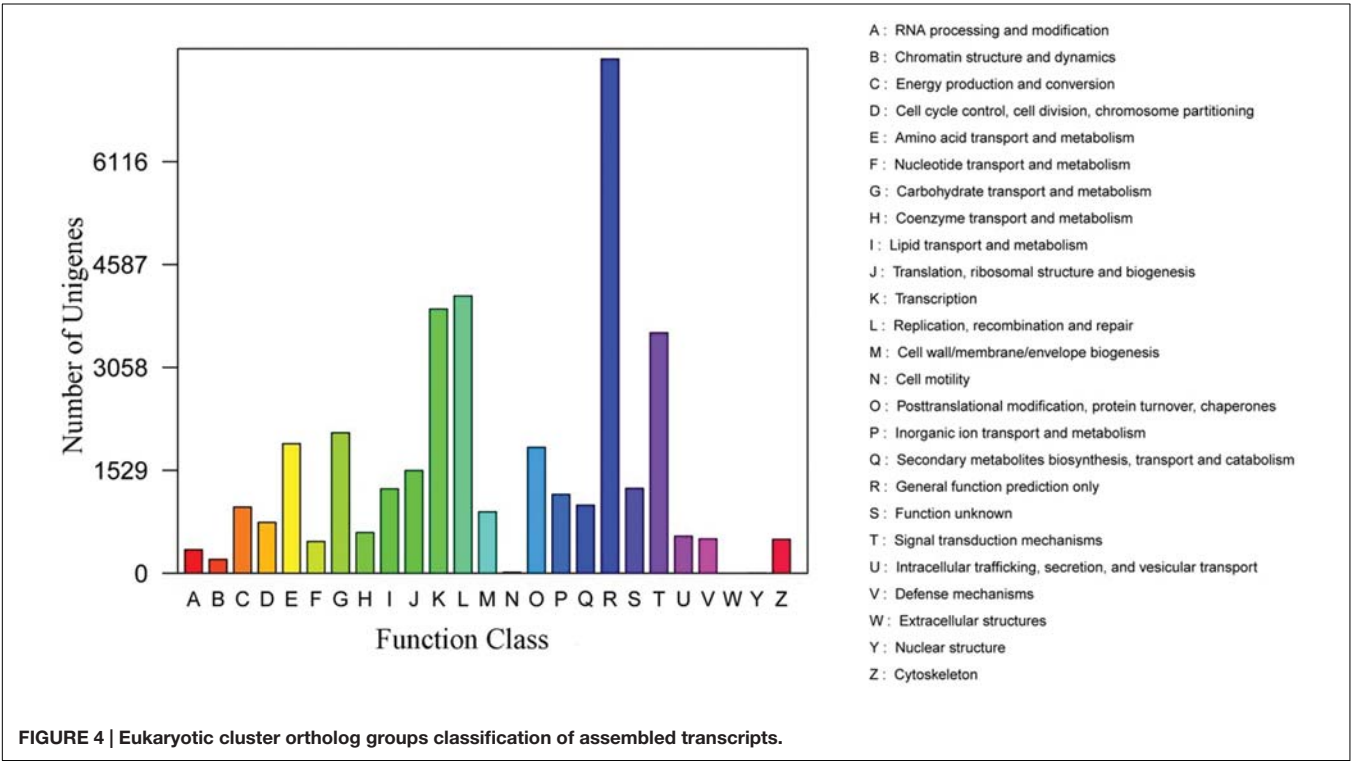
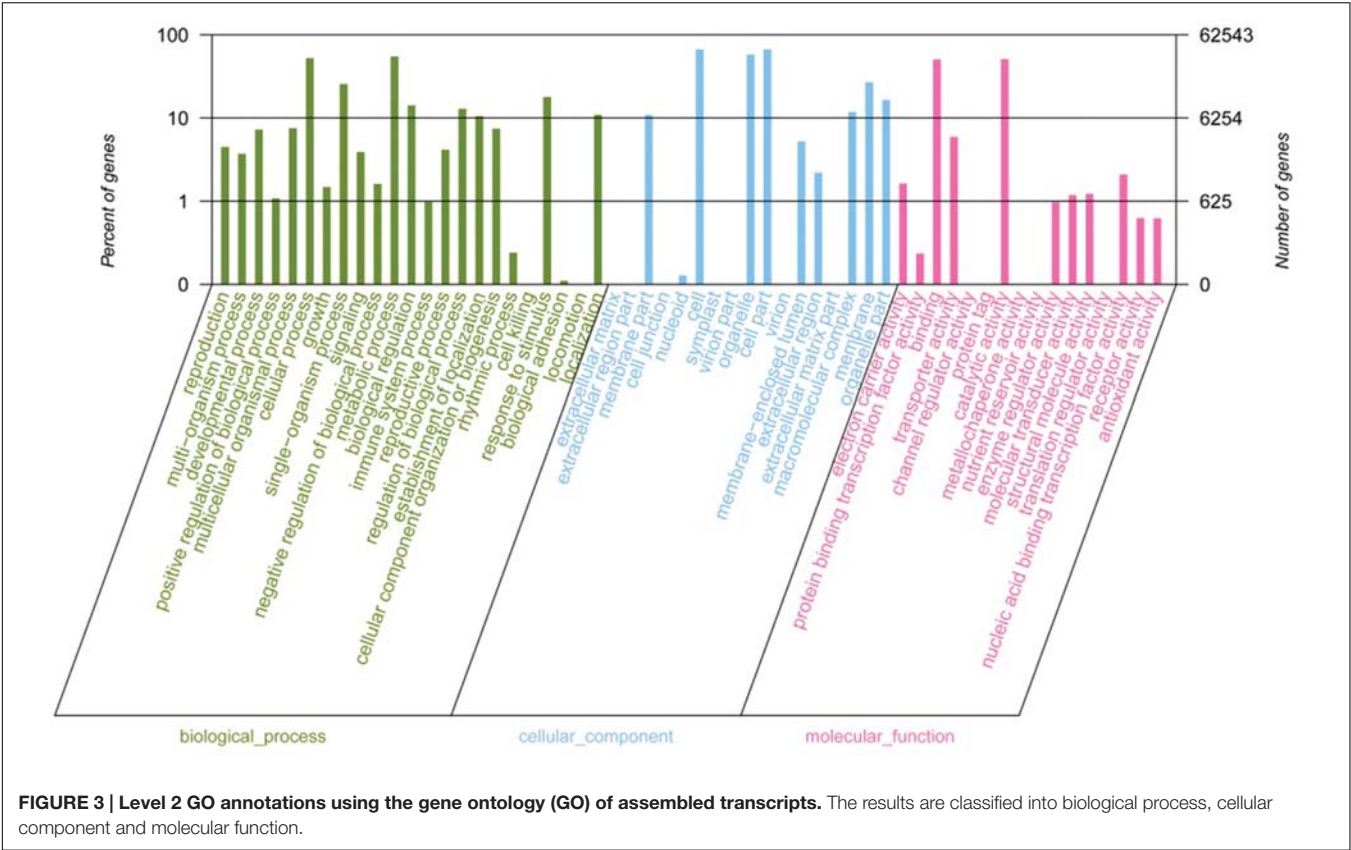


**FIGURE 2 | Summary of Broomcorn millet unigenes similarity search compared to other monocots using Circos (<http://circos.ca/>).** 43.68% unigenes of broomcorn millet have homologous matches to *Sorghum bicolor* transcripts, 34.74% have hits to *Zea mays*, 11.19% have hits to *Oryza sativa*, 3.65% have hits to *Brachypodium distachyon*, 1.52% have hits to *Hordeum vulgare*, 1.45% have hits to *Aegilops tauschii*, 0.90% have hits to *Triticum urartu*, 0.14% have hits to *Setaria italica*, and 0.09% have hits to *Panicum virgatum* as well as 1.78% have hits to other species, respectively.

Within the cellular component, 'cell,' 'cell part,' and 'organelle' were the most abundant among the 56 categories, which together accounted for 72.14% of the genes belonging to this class. In biological process, 'metabolic' and 'cellular processes' were the largest and second largest categories,

accounting for 44.09% of the sequences assigned to this term. In 16 different molecular function categories, the two most abundant categories were 'catalytic activity' and 'binding,' which accounted for 87.44% for the genes of this class (Figure 3).





**TABLE 1 | Simple sequence repeats (SSRs) identification in broomcorn millet.**

Type	Number of SSRs	Repeat in each unit	Number	Percent (%)
Mononucleotide motifs	6,433	A/T	6,167	95.88
		C/G	266	4.12
Dinucleotide motifs	3,635	AG/CT	2,274	62.56
		AC/GT	576	15.84
		AT/AT	412	11.34
		CG/CG	373	10.26
Trinucleotide motifs	23,499	CCG/CGG	10,970	46.67
		AGC/GCT	3,541	15.07
		AGG/CCT	3,182	13.54
		AAG/CTT	1,560	6.64
		ACC/GGT	1,351	5.75
		ACG/CGT	1,180	5.02
Tetranucleotide motifs	1,107	ATTT/AAAT	93	8.40
		AAAG/CTTT	92	8.31
		CCGG/CCGG	60	5.42
Pentanucleotide motifs	436	CCCTT/AAGGG	62	14.22
		GAGGT/ACTCT	58	13.30
Hexanucleotide motifs	106	CCGGCGC/GCCGG	20	18.87
		ACGCCC/GGGCGT	13	12.26
Total number of sequences examined			113,643	
Total size of examined sequences (bp)			164,535,293	
Total number of identified SSRs			35,216	
Number of SSR containing sequences			27,055	
Number of sequences containing more than one SSR			2,536	
Number of SSRs present in compound formation			786	

Furthermore, the assembled unigenes were aligned to the KOG database to classify their putative function. Result showed that a total of 33,671 unigenes could be matched to genes in the KOG database. These unigenes were classified into 25 different functional classes. Among transcripts with matches to the KOG database, 'the function prediction only class' (7,071, 21.00%) represented the largest group, followed by 'signal transduction mechanisms' (4,986; 14.81%), 'post-translational modification, protein turnover and chaperones' (3,835; 11.39%) 'transcription' (2,570; 7.63%), while only a few unigenes were assigned to 'extracellular structures' (70; 0.21%) and 'cell motility' (11; 0.03%; **Figure 4**).

Finally, KEGG analysis was used to assigned unigenes to metabolic pathways. 15,514 unigenes were mapped to 202 KEGG pathways. Among them, the most represented pathways were metabolic pathways (ko01100; 25.65%), followed by biosynthesis of secondary metabolites (ko01110; 10.71%), biosynthesis of amino acids (ko01230; 3.57%), pyrimidine metabolism (ko00240; 2.75%), purine metabolism (ko00230; 2.70%), peroxisome (ko04146; 2.55%), and spliceosome (ko03040; 2.50%) as well as plant-pathogen interaction (ko05169; 2.37%; Supplementary Figure S3 and Supplementary Table S3).

## Identification of SSRs and SNPs Loci in Broomcorn Millet

Simple sequence repeats are one of the most informative and versatile molecular markers, which are widely used in genetic diversity, genetic structure and genetic mapping studies (Varshney et al., 2005a,b; Arora et al., 2014; Ting et al., 2014). To provide useful information for marker development in broomcorn millet, we firstly investigated the EST-SSR loci using the assembled transcripts. A total of 35,216 SSR loci were identified in 27,055 sequences out of 113,643 unigenes. Among them, 2,536 sequences contained more than one SSR (**Table 1**). The repeat number of the SSRs ranged from 4 times to 24 times, and a repeat number of four was the most abundant repeats number accounting for 49.99% of SSRs, followed by repeat number of five accounting for 14.45% of SSRs, repeat number of 10 accounted for 13.38% of SSRs, repeat number six accounted for 8.79% of SSRs, while repeat number of more than 20 were rare, only accounted for 0.05% of SSRs. Within the different types of SSRs, trinucleotide was the most abundant repeat unit accounted for 66.72% of SSRs, which was consistent with cabbage, red clover, sweetpotato and cucumber (Guo et al., 2010; Wang et al., 2010; Izzah et al., 2014; Yates et al., 2014), followed by mononucleotide, which accounted for 18.26% of SSRs, dinucleotide, which accounted for 10.33% of SSRs, tetranucleotide, which accounted for 3.15% of SSRs, pentanucleotide, which accounted for 1.24% of SSRs and hexanucleotide accounted for 0.30% of SSRs (**Figure 5**). The most abundant repeat type was A/T, which accounted for 95.88% of the mononucleotide SSRs. The least repeat type was C/G, which accounted for 4.12% of the mononucleotide SSRs. Within dinucleotide repeats, the most abundant motif was AG/CT, which accounted for 62.56% of the dinucleotide SSRs, which was similar with the previous results in sweet potato, coffee, peanut, and Arachis (Poncet et al., 2006; Liang et al., 2009; Wang et al., 2010). In contrast, AC/GT were the most abundant dinucleotide repeats in soybean, maize, rice, wheat, and barley where AC/GT were the most frequent repeats. Within trinucleotide SSRs, CCG/CGG were the most abundant motifs accounting for 46.67% of the trinucleotide SSRs, which were the most abundant trinucleotide repeats in soybean (Xin et al., 2012), maize and barley (Kantety et al., 2002). Previous studies have reported that the CCG/CGG motif was very rare in dicotyledonous plants while abundant in monocots (Wang et al., 2011). The result of Broomcorn millet was consistent with this conclusion (**Table 1**). Finally, the length of SSRs loci were found to range from 10 to 25 bp, of which 12 bp were the most frequent accounting for 51.39% of the SSRs loci, followed by 15 bp accounting for 13.78% of the SSRs loci and 10 bp accounting for 11.91% of the SSRs loci (Supplementary Figure S4).

Single nucleotide polymorphisms have recently become a more popular marker for high density genetic mapping, association mapping and population genetic structure studies (Liu et al., 2011). SNPs occurring in coding regions (cSNP) may cause the loss or change of protein function, and thus, they could be used directly to assess the impact of mutation on important economic traits (Ellegren, 2008). To understand

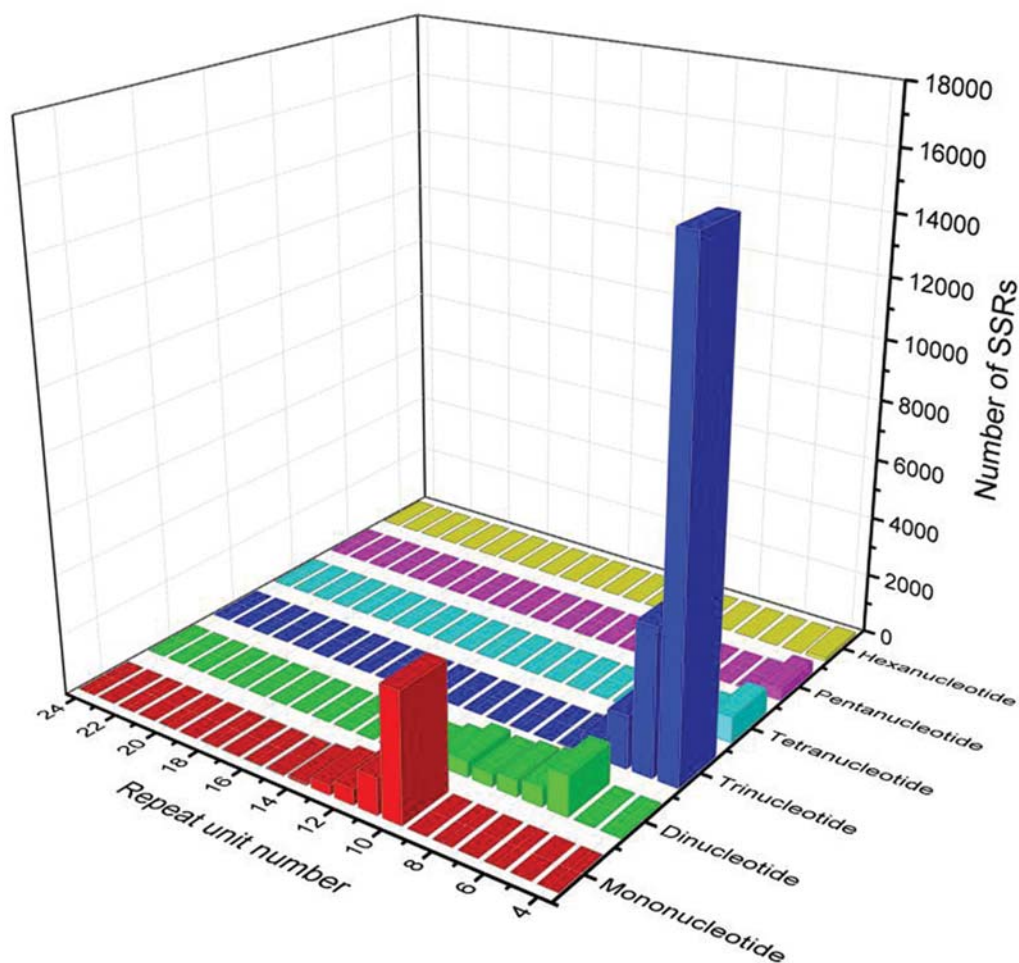


FIGURE 5 | Number of different SSRs repeat types between Yumi No. 2 and Yumi No. 3.

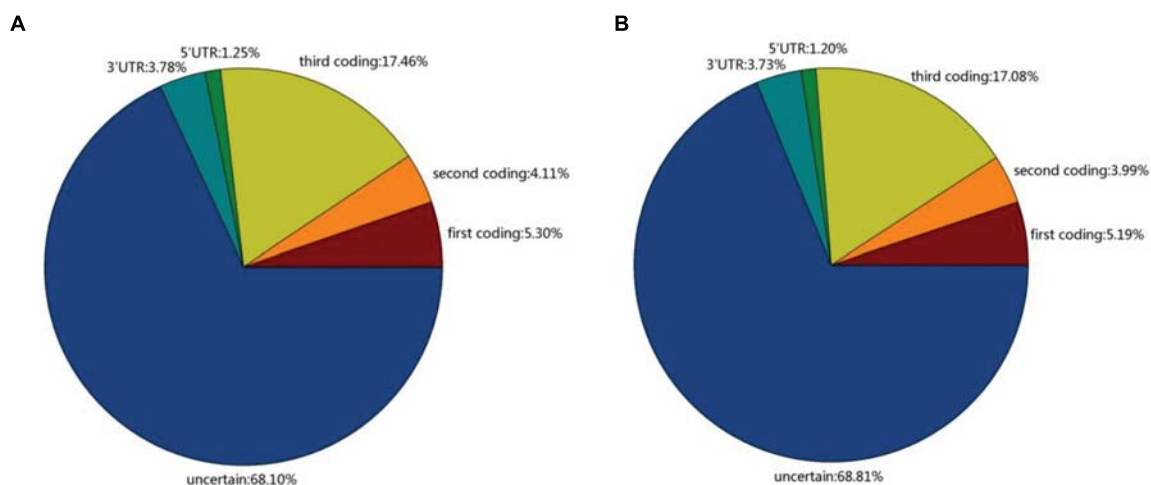
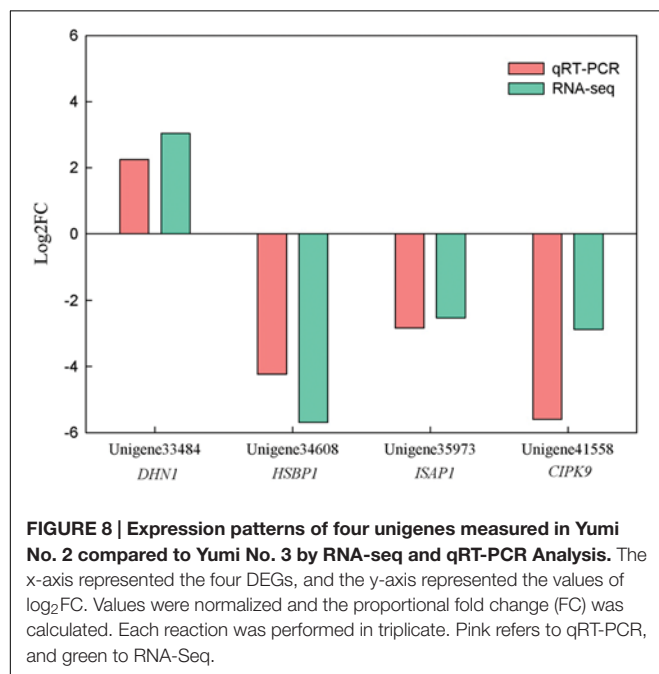
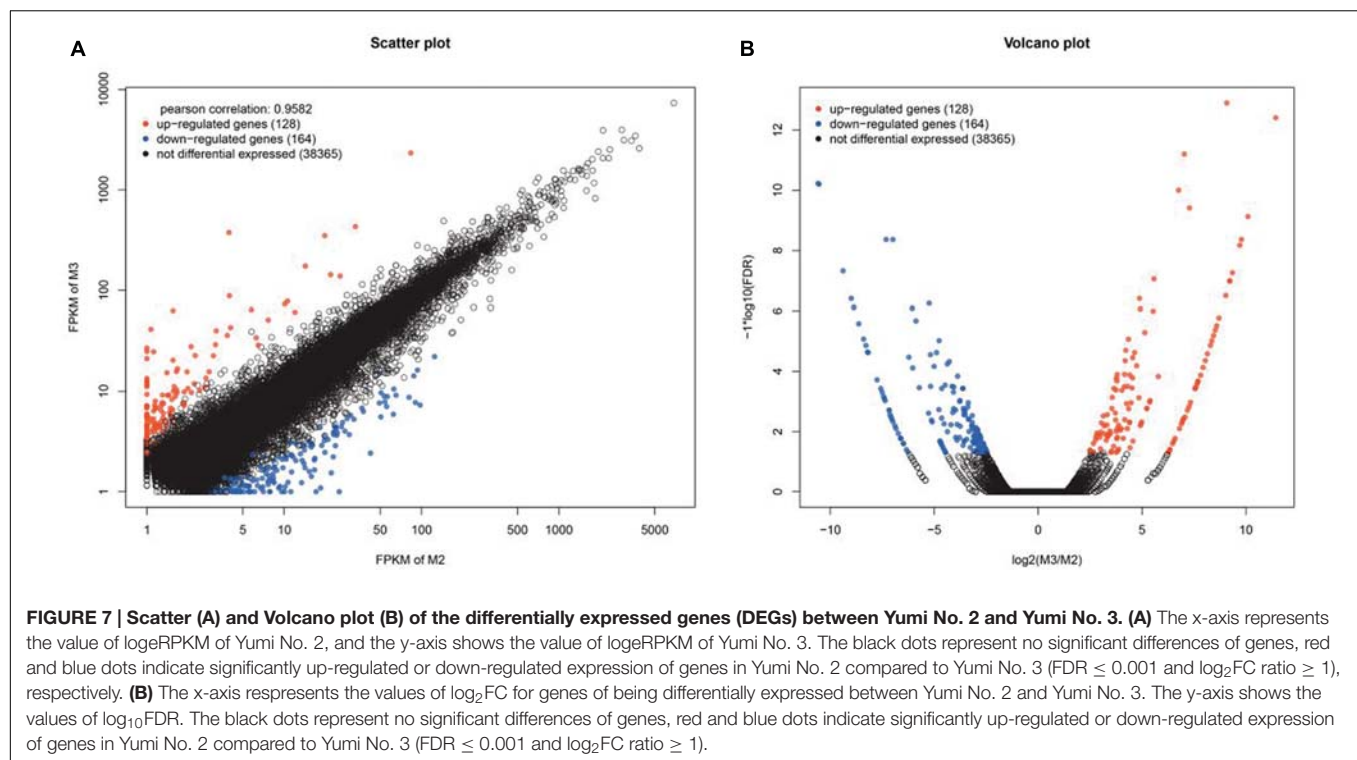


FIGURE 6 | The distribution of SNPs in Yumi No. 2 (A) and Yumi No. 3 (B). The transcript with no homologous in NR, Swiss-Prot protein database and also no predicted protein-coding potential were identified as non-coding regions.



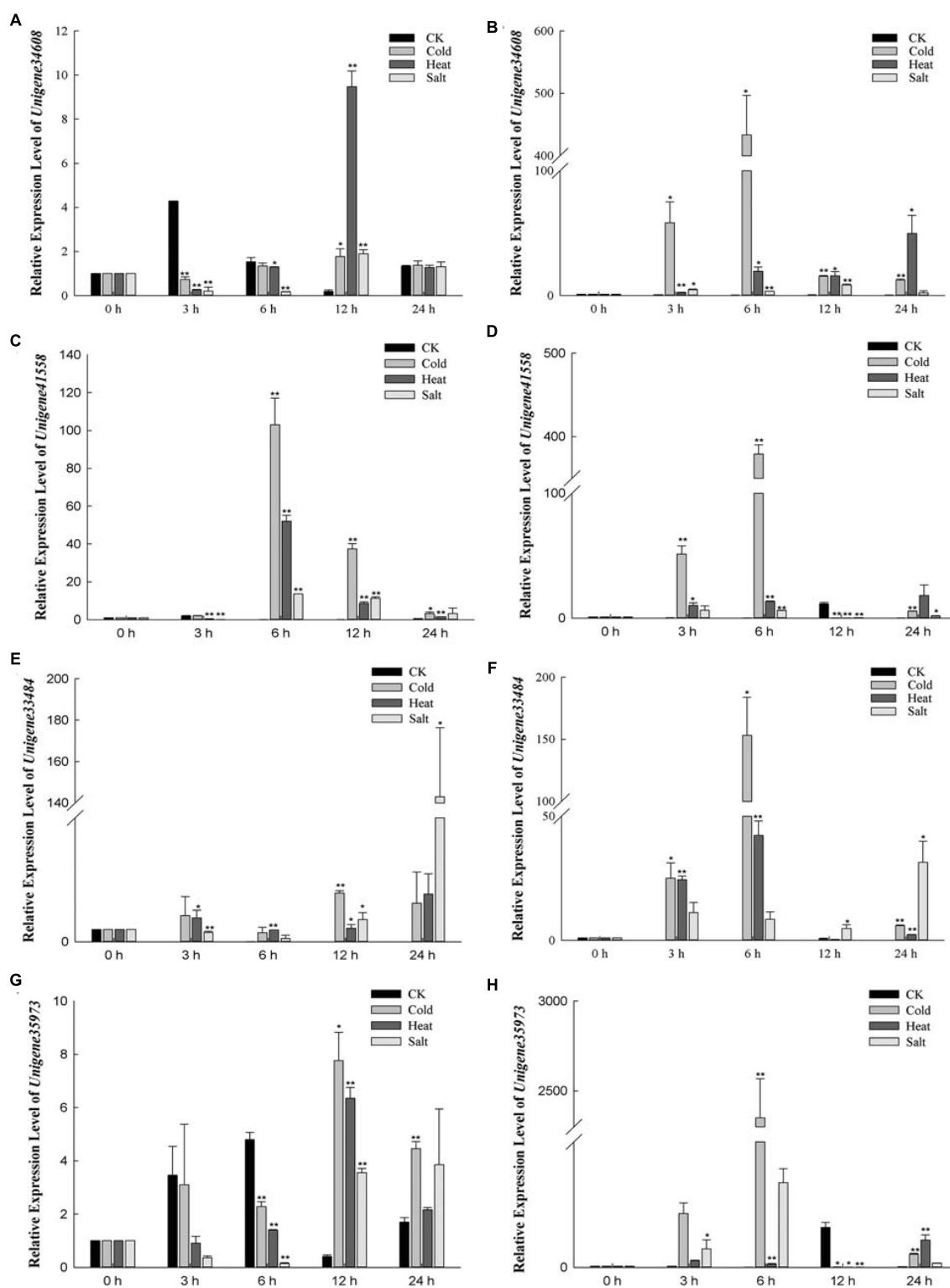
the cSNP in Broomcorn millet, we identified cSNPs between these two cultivars. A total of 406,062 high-quality SNPs were identified in Yumi No. 2 and 409,850 high-quality SNPs were identified in Yumi No. 3. For Yumi No. 2, the putative SNPs included 270,068 transitions (A/G, C/T) and 135,994 transversions (G/T, C/G, A/T, A/C), while 273,593 transitions

and 136,257 transversions were observed in Yumi No. 3 (Supplementary Figure S5). The average number was 2.46 and 2.49 SNPs per kb for Yumi No. 2 and Yumi No. 3, respectively. Further analysis found that in Yumi No. 2, 109,078 (26.87%) SNPs were distributed into coding sequences, 20,446 (5.03%) SNPs distributed into untranslated regions (UTRs) and 276,539 (68.10%) distributed in non-coding regions, respectively. Among the SNP of Yumi No. 2, the percentage distributed in the coding sequences (CDSs), UTRs and non-coding regions was 26.87, 5.03, and 68.10%, respectively. However, among the SNP of Yumi No. 3, the percentage distributed in the CDSs, UTRs and non-coding regions was 26.26, 4.93, and 68.81%, respectively (Figure 6).

## Detection of Differentially Expressed Genes (DEGs)

To detect DEGs, the expression level of these unigenes in the two cultivars was investigated. Pairwise comparison of RPKM and FC between Yumi No. 2 and Yumi No. 3 in the RNA-seq data sets was first conducted. The scatter plots showed the expression differences of each gene, and volcano plots showed the expression differences of gene among these two cultivars. The results indicated that most genes were expressed at similar levels (black dots) between the cultivars and only a small portion of genes were significantly up-regulated (red dots) and down-regulated (blue dots) expressed. A total of 292 DEGs were obtained, of which 128 genes (red dots) were up-regulated and 164 genes (blue dots) were down-regulated in Yumi No. 2 compared to Yumi No. 3, all of which had statistically significant differences in expression levels ( $P < 0.05$ ; Figure 7). The list of





**FIGURE 9 | Relative expression levels of four DEGs which may be involved in abiotic-stress response in Yumi No. 2 and YumiNo. 3 by qRT-PCR analysis.** The  $2^{-\Delta\Delta Ct}$  values were used to analyze differential expression levels after cold, heat, salt treatment. Error bars indicate SD (n = 3). \*P-value < 0.05, \*\*P-value < 0.01 (Student's t-test). **(A)** The expression level of *Unigene34608* in Yumi No. 2. **(B)** The expression level of *Unigene34608* in Yumi No. 3. **(C)** The expression level of *Unigene41558* in Yumi No. 2. **(D)** The expression level of *Unigene41558* in Yumi No. 3. **(E)** The expression level of *Unigene33484* in Yumi No. 2. **(F)** The expression level of *Unigene33484* in Yumi No. 3. **(G)** The expression level of *Unigene35973* in Yumi No. 2. **(H)** The expression level of *Unigene35973* in Yumi No. 3.

DEGs has shown in Supplementary Table S4 and Supplementary Figure S6.

To determine the potential function of these DEGs, GO, and KEGG analyses were performed. In total, 88 GO terms were significantly enriched in the DEGs at the stringent cut off level of  $P < 0.05$ . In biological process, the GO terms metabolic process (GO:0008152) and cellular process (GO:0009987) were enriched in the DEGs while in cellular component, the GO terms cytoplasmic part (GO:0044444) and intracellular organelle (GO:0043229) were enriched. Finally, heterocyclic compound binding (GO:1901363) and organic cyclic compound binding (GO:0097159) were enriched in the molecular function category (Supplementary Table S5). KEGG pathway enrichment analysis for DEGs included 12 enriched pathways (Supplementary Table S6). RNA degradation (ko03018), nucleotide excision repair (ko03420), ubiquinone and other terpenoid-quinone biosynthesis (ko00130) and phosphatidylinositol signaling system (ko04070) were the top four enriched pathways. The differential expressed genes clustered into different functional categories, which provided the important resource to discover and identify the important functional genes.

### Validation of the DEGs Using qRT-PCR

To validate the DEGs from the RNA-seq data, four DEGs which may be involved in abiotic stress response were selected to qRT-PCR analysis. Three DEGs, including *unigene34608*, *unigene35973* and *unigene41558*, were down-regulated in Yumi No. 2 compared to Yumi No. 3. *Unigene33484* was up-regulated in Yumi No. 2 compared to Yumi No. 3. Results showed that the expression patterns of three transcripts (*unigene33484*, *unigene34608*, and *unigene35973*) were consistent with that of RNA-seq (Figure 8). Although, the expression level of the remaining transcript *unigene41558* obtained from qRT-PCR was significantly higher than that of RNA-seq data, they showed a similar down-regulated expression trend to RNA-seq (Figure 8). Consequently, the RNA-seq could provide the useful information for gene expression and the identified DEGs provided a valuable resource for gene discovery and functional analysis in Broomcorn millet.

To identify some stress-related genes from Broomcorn millet, the expression profiles of these unigenes which may be involved in abiotic-stress response were further detected under different stresses (Figure 9). *Unigene34608* is predicted to encode heat shock factor-binding protein 1 (HSBP1). HSBP1 can affect HSF1 DNA binding activity and is a negative regulator in response to heat stress (Satyal et al., 1998). In Yumi No. 2, the results showed that the transcript levels of *Unigene34608* had small expression level changes under cold stress, while its expression level was reduced by 0.26-fold under heat stress for 3 h and 0.16-fold after 6 h under salt stress compared to control plants. In Yumi No. 3, the expression level of *Unigene34608* was temporarily elevated under cold and heat stress, especially expression level was increased more than 400-fold compared to control plants under cold treatment for 6 h. And it quickly declined to low levels under salt stress for 24 h. *Unigene41558* putative encodes a CBL-interacting protein kinase 9 (CIPK9),

which interacts with calcium sensor and plays important roles in low- $K^+$  stress (Liu L.L. et al., 2012; Hung et al., 2014). High expression levels of *Unigene41558* were observed under several stress treatments. For example, in Yumi No. 2, expression levels of *Unigene41558* increased 103.02-fold compared to untreated controls under cold treatment for 6 h, 52.02-fold under heat treatment for 6 h and 11.39-fold under salt treatment for 12 h. Similar trends were also observed in Yumi No. 3 where highest expression level of *unigene41558* were observed under cold stress for 6 h, heat stress for 24 h, and salt stress for 3 h. *Unigene33484* is homologous to an acidic Y2Kn dehydrin DHN1 (Allagulova et al., 2003). A previous study revealed over-expression of *DHN1* gene positively affected plant growth under abiotic stress (Beck et al., 2007). The expression levels of *Unigene33484* show slight increases in expression levels under cold and heat stress in Yumi No. 2 with values ranging from 1- to 4-fold higher than untreated controls, while under salt stress, expression levels initially declined to 0.26-fold at 6 h and gradually increased to 1.80-fold at 12 h and reached the highest expression levels for 24 h with expression levels 100-fold higher than untreated controls. In Yumi No. 3, the expression patterns of *Unigene33484* under stress treatment were different from Yumi No. 2, which increased 153.22-, 42.27-, 31.38-fold under cold treatment for 6 h, heat stress for 6 h and salt stress for 24 h, respectively. It is indicated that *Unigene33484* likely plays a role in osmoregulation in Broomcorn millet. *Unigene35973* is predicted to encode a zinc-finger protein gene *ISAP1*, which involved in regulating cold, dehydration, and salt tolerance in transgenic tobacco (Mukhopadhyay et al., 2004). In this study, expression levels of *Unigene35973* in Yumi No 2 were 7.72-, 6.34-, 3.75-fold higher under cold, heat, and salt stress compared to untreated controls, respectively. In Yumi No. 3, expression levels of this unigene were over 100-fold higher than untreated controls under both cold and salt stress. The expression profiles of these four unigenes suggested they may play an important role in response to abiotic stress in Broomcorn millet, which provided the foundation for further study of the molecular mechanism of stress tolerance of this important crop.

### CONCLUSION

This is the first large scale *de novo* assembly and analysis of the transcriptome in broomcorn millet. A total of 113,643 unigenes were obtained, of which 62,543 were functionally annotated. Furthermore, more than 35,000 SSRs and 406,000 SNP loci were identified, which provide an important resource for marker development in this species. This study provided the first insight into the transcriptome of Broomcorn millet, which not only provided an invaluable sequence resource and genomic information for molecular studies in this important crop, but also shed light on discovering the vital functional genes involving in the metabolism regulation network of involved in the adaptation to extreme climatic conditions as well as facilitating further studies on molecular mechanisms of stress tolerance in Broomcorn millet.

## AUTHOR CONTRIBUTIONS

HY, WS, and XN conceived and designed the experiments. HY and LW performed the experiments. HY, WY, and HL analyzed the data. XD contributed reagents, materials, and analytical tools. HY, WS, and XN wrote the paper. All authors read and approved the final manuscript.

## FUNDING

This work was mainly funded by the National Natural Science Foundation of China (Grant No. 31401373) and partially supported by the Open Project Program of State Key Laboratory

of Crop Stress Biology in Arid Areas, China (CSBAA2014002).

## ACKNOWLEDGMENT

We would like to thank Prof. Baili Feng for kindly providing two varieties of broomcorn millet.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.01083>

## REFERENCES

- Allagulova, C. R., Gimalov, F., Shakirova, F., and Vakhitov, V. (2003). The plant dehydrins: structure and putative functions. *Biochemistry (Mosc.)* 68, 945–951. doi: 10.1023/A:1026077825584
- Arora, A., Kundu, S., Dilbaghi, N., Sharma, I., and Tiwari, R. (2014). Population structure and genetic diversity among Indian wheat varieties using microsatellite (SSR) markers. *Aust. J. Crop Sci.* 8, 1281–1289.
- Beck, E. H., Fetting, S., Knake, C., Hartig, K., and Bhattarai, T. (2007). Specific and unspecific responses of plants to cold and drought stress. *J. Biosci.* 32, 501–510. doi: 10.1007/s12038-007-0049-5
- Bellwood, P., Gamble, C., Le Blanc, S. A., Pluciennik, M., Richards, M., and Terrell, J. E. (2007). First farmers: the origins of agricultural societies, by Peter Bellwood. Malden (MA): Blackwell, 2005; ISBN 0-631-20565-9 hardback£ 60; ISBN 0-631-20566-7 paperback£ 17.99, xix+ 360 pp., 59 figs., 3 tables. *Camb. Archaeol. J.* 17, 87–109. doi: 10.1017/S0959774307000078
- Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., and Golani, I. (2001). Controlling the false discovery rate in behavior genetics research. *Behav. Brain Res.* 125, 279–284. doi: 10.1016/S0166-4328(01)00297-2
- Chen, J., Hou, K., Qin, P., Liu, H., Yi, B., Yang, W., et al. (2014). RNA-Seq for gene identification and transcript profiling of three *Stevia rebaudiana* genotypes. *BMC Genomics* 15:571. doi: 10.1186/1471-2164-15-571
- Clark, M. B., Mercer, T. R., Bussotti, G., Leonardi, T., Haynes, K. R., Crawford, J., et al. (2015). Quantitative gene profiling of long noncoding RNAs with targeted RNA sequencing. *Nat. Methods* 12, 339–342. doi: 10.1038/nmeth.3321
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi: 10.1093/bioinformatics/bti610
- Crawford, G. W. (2006). “East Asian plant domestication,” in *Archaeology of Asia*, ed. M. T. Stark (Oxford: Blackwell Publishing Ltd.), 77–95.
- Dai, H., Jia, G., Lu, C., Wei, A., Feng, B., and Zhang, S. (2011). Studies of synergism between root system and leaves senescence in Broomcorn millet (*Panicum miliaceum* L.). *J. Food Agric. Environ.* 9, 177–180.
- Ellegren, H. (2008). Comparative genomics and the study of evolution by natural selection. *Mol. Ecol.* 17, 4586–4596. doi: 10.1111/j.1365-294X.2008.03954.x
- Guo, S., Zheng, Y., Joung, J.-G., Liu, S., Zhang, Z., Crasta, O. R., et al. (2010). Transcriptome sequencing and comparative analysis of cucumber flowers with different sex types. *BMC Genomics* 11:384. doi: 10.1186/1471-2164-11-384
- Hu, Y., Zhu, J., Liu, F., Zhang, Z., Chai, Y., and Weining, S. (2008). Genetic diversity among Chinese landraces and cultivars of broomcorn millet (*Panicum miliaceum*) revealed by the polymerase chain reaction. *Ann. Appl. Biol.* 153, 357–364. doi: 10.1111/j.1744-7348.2008.00263.x
- Hung, C.-Y., Aspesi, P. Jr., Hunter, M. R., Lomax, A. W., and Perera, I. Y. (2014). Phosphoinositide-signaling is one component of a robust plant defense response. *Front. Plant Sci.* 5:267. doi: 10.3389/fpls.2014.00267
- Hunt, H. V., Badakshi, F., Romanova, O., Howe, C. J., Jones, M. K., and Heslop-Harrison, J. P. (2014). Reticulate evolution in *Panicum* (Poaceae): the origin of tetraploid broomcorn millet, *P. miliaceum*. *J. Exp. Bot.* 65, 3165–3175. doi: 10.1093/jxb/eru161
- Hunt, H. V., Campana, M. G., Lawes, M. C., Park, Y. J., Bower, M. A., Howe, C. J., et al. (2011). Genetic diversity and phylogeography of broomcorn millet (*Panicum miliaceum* L.) across Eurasia. *Mol. Ecol.* 20, 4756–4771. doi: 10.1111/j.1365-294X.2011.05318.x
- Hunt, H. V., Denyer, K., Packman, L. C., Jones, M. K., and Howe, C. J. (2010). Molecular basis of the waxy endosperm starch phenotype in broomcorn millet (*Panicum miliaceum* L.). *Mol. Biol. Evol.* 27, 1478–1494. doi: 10.1093/molbev/msq040
- Izzah, N. K., Lee, J., Jayakodi, M., Perumal, S., Jin, M., Park, B.-S., et al. (2014). Transcriptome sequencing of two parental lines of cabbage (*Brassica oleracea* L. var. capitata L.) and construction of an EST-based genetic map. *BMC genomics* 15:149. doi: 10.1186/1471-2164-15-149
- Jiang, Y., and Deyholos, M. K. (2006). Comprehensive transcriptional profiling of NaCl-stressed *Arabidopsis* roots reveals novel classes of responsive genes. *BMC Plant Biol.* 6:25. doi: 10.1186/1471-2229-6-25
- Kantety, R. V., La Rota, M., Matthews, D. E., and Sorrells, M. E. (2002). Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol. Biol.* 48, 501–510. doi: 10.1023/A:1014875206165
- Li, M.-Y., Wang, F., Jiang, Q., Ma, J., and Xiong, A.-S. (2014). Identification of SSRs and differentially expressed genes in two cultivars of celery (*Apium graveolens* L.) by deep transcriptome sequencing. *Hortic. Res.* 1, 10. doi: 10.1038/hortres.2014.10
- Li, R., Yu, C., Li, Y., Lam, T.-W., Yiu, S.-M., Kristiansen, K., et al. (2009). SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 25, 1966–1967. doi: 10.1093/bioinformatics/btp336
- Liang, X., Chen, X., Hong, Y., Liu, H., Zhou, G., Li, S., et al. (2009). Utility of EST-derived SSR in cultivated peanut (*Arachis hypogaea* L.) and *Arachis* wild species. *BMC Plant Biol.* 9:35. doi: 10.1186/1471-2229-9-35
- Liu, G., Li, W., Zheng, P., Xu, T., Chen, L., Liu, D., et al. (2012). Transcriptomic analysis of ‘Suli’pear (*Pyrus pyrifolia* white pear group) buds during the dormancy by RNA-Seq. *BMC Genomics* 13:700. doi: 10.1186/1471-2164-13-700
- Liu, L. L., Ren, H. M., Chen, L. Q., Wang, Y., and Wu, W. H. (2012). A protein kinase CIPK9 interacts with calcium sensor CBL3 and regulates K<sup>+</sup> homeostasis under low-K<sup>+</sup> stress in *Arabidopsis*. *Plant Physiol.* 161, 266–277. doi: 10.1104/pp.112.206896
- Liu, S., Zhou, Z., Lu, J., Sun, F., Wang, S., Liu, H., et al. (2011). Generation of genome-scale gene-associated SNPs in catfish for the construction of a high-density SNP array. *BMC Genomics* 12:53. doi: 10.1186/1471-2164-12-53
- Lu, H., Zhang, J., Liu, K.-B., Wu, N., Li, Y., Zhou, K., et al. (2009). Earliest domestication of common millet (*Panicum miliaceum*) in East Asia extended to 10,000 years ago. *Proc. Natl. Acad. Sci. U.S.A.* 106, 7367–7372. doi: 10.1073/pnas.0900158106
- Miller, P. R., McConkey, B. G., Clayton, G. W., Brandt, S. A., Staricka, J. A., Johnston, A. M., et al. (2002). Pulse crop adaptation in the northern Great Plains. *Agron. J.* 94, 261–272. doi: 10.2134/agronj2002.0261

- Mukhopadhyay, A., Vij, S., and Tyagi, A. K. (2004). Overexpression of a zinc-finger protein gene from rice confers tolerance to cold, dehydration, and salt stress in transgenic tobacco. *Proc. Natl. Acad. Sci. U.S.A.* 101, 6309–6314. doi: 10.1073/pnas.0401572101
- Poncet, V., Rondeau, M., Tranchant, C., Cayrel, A., Hamon, S., de Kochko, A., et al. (2006). SSR mining in coffee tree EST databases: potential use of EST-SSRs as markers for the *Coffea* genus. *Mol. Genet. Genomics* 276, 436–449. doi: 10.1007/s00438-006-0153-5
- Salgado, L. R., Koop, D. M., Pinheiro, D. G., Rivallan, R., Le Guen, V., Nicolás, M. F., et al. (2014). De novo transcriptome analysis of *Hevea brasiliensis* tissues by RNA-seq and screening for molecular markers. *BMC Genomics* 15:236. doi: 10.1186/1471-2164-15-236
- Sartelet, A., Li, W., Pailhoux, E., Richard, C., Tamma, N., Karim, L., et al. (2015). Genome-wide next-generation DNA and RNA sequencing reveals a mutation that perturbs splicing of the phosphatidylinositol glycan anchor biosynthesis class H gene (PIGH) and causes arthrogryposis in Belgian Blue cattle. *BMC Genomics* 16:316. doi: 10.1186/s12864-015-1528-y
- Satyal, S. H., Chen, D., Fox, S. G., Kramer, J. M., and Morimoto, R. I. (1998). Negative regulation of the heat shock transcriptional response by HSBP1. *Genes Dev.* 12, 1962–1974. doi: 10.1101/gad.12.13.1962
- Stephens, A., Lombardi, M., Cogan, N. O., Forster, J. W., Hobson, K., Materne, M., et al. (2014). Genetic marker discovery, intraspecific linkage map construction and quantitative trait locus analysis of ascochyta blight resistance in chickpea (*Cicer arietinum* L.). *Mol. Breed.* 33, 297–313. doi: 10.1007/s11032-013-9950-9
- Ting, N.-C., Jansen, J., Mayes, S., Massawe, F., Sambanthamurthi, R., Ooi, L. C.-L., et al. (2014). High density SNP and SSR-based genetic maps of two independent oil palm hybrids. *BMC Genomics* 15:309. doi: 10.1186/1471-2164-15-309
- Varshney, R. K., Graner, A., and Sorrells, M. E. (2005a). Genic microsatellite markers in plants: features and applications. *Trends Biotechnol.* 23, 48–55. doi: 10.1016/j.tibtech.2004.11.005
- Varshney, R. K., Sigmund, R., Börner, A., Korzun, V., Stein, N., Sorrells, M. E., et al. (2005b). Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. *Plant Sci.* 168, 195–202. doi: 10.1016/j.plantsci.2004.08.001
- Wang, Z., Fang, B., Chen, J., Zhang, X., Luo, Z., Huang, L., et al. (2010). De novo assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweetpotato (*Ipomoea batatas*). *BMC Genomics* 11:726. doi: 10.1186/1471-2164-11-726
- Wang, Z., Li, J., Luo, Z., Huang, L., Chen, X., Fang, B., et al. (2011). Characterization and development of EST-derived SSR markers in cultivated sweetpotato (*Ipomoea batatas*). *BMC Plant Biol.* 11:139. doi: 10.1186/1471-2229-11-139
- Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., et al. (2011). KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 39(Suppl. 2), W316–W322. doi: 10.1093/nar/gkr483
- Xin, D., Sun, J., Wang, J., Jiang, H., Hu, G., Liu, C., et al. (2012). Identification and characterization of SSRs from soybean (*Glycine max*) ESTs. *Mol. Biol. Rep.* 39, 9047–9057. doi: 10.1007/s11033-012-1776-8
- Xu, K., Sun, F., Chai, G., Wang, Y., Shi, L., Liu, S., et al. (2015). De novo assembly and transcriptome analysis of two contrary tillering mutants to learn the mechanisms of tillers outgrowth in switchgrass (*Panicum virgatum* L.). *Front. Plant Sci.* 6:749. doi: 10.3389/fpls.2015.00749
- Yang, X., Wan, Z., Perry, L., Lu, H., Wang, Q., Zhao, C., et al. (2012). Early millet use in northern China. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3726–3730. doi: 10.1073/pnas.1115430109
- Yates, S. A., Swain, M. T., Hegarty, M. J., Chernukin, I., Lowe, M., Allison, G. G., et al. (2014). De novo assembly of red clover transcriptome based on RNA-Seq data provides insight into drought response, gene discovery and marker identification. *BMC Genomics* 15:453. doi: 10.1186/1471-2164-15-453
- Zhan, J., Thakare, D., Ma, C., Lloyd, A., Nixon, N. M., Arakaki, A. M., et al. (2015). RNA sequencing of laser-capture microdissected compartments of the maize kernel identifies regulatory modules associated with endosperm cell differentiation. *Plant Cell* 27, 513–531. doi: 10.1105/tpc.114.135657

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer ES and handling Editor declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.

Copyright © 2016 Yue, Wang, Liu, Yue, Du, Song and Nie. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Water Deficit Affects Primary Metabolism Differently in Two *Lolium multiflorum*/*Festuca arundinacea* Introgression Forms with a Distinct Capacity for Photosynthesis and Membrane Regeneration

Dawid Perlikowski<sup>1</sup>, Mariusz Czyżniewski<sup>1</sup>, Łukasz Marczak<sup>2</sup>, Adam Augustyniak<sup>1</sup> and Arkadiusz Kosmala<sup>1\*</sup>

<sup>1</sup> Institute of Plant Genetics, Polish Academy of Science, Poznań, Poland, <sup>2</sup> Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznań, Poland

## OPEN ACCESS

### Edited by:

Teresa Donze,  
University of Nebraska–Lincoln, USA

### Reviewed by:

Prateek Tripathi,  
Scripps Research Institute, USA

Hao Peng,  
Washington State University, USA

### \*Correspondence:

Arkadiusz Kosmala  
akos@igr.poznan.pl

### Specialty section:

This article was submitted to  
Crop Science and Horticulture,  
a section of the journal  
Frontiers in Plant Science

**Received:** 22 April 2016

**Accepted:** 06 July 2016

**Published:** 25 July 2016

### Citation:

Perlikowski D, Czyżniewski M, Marczak Ł, Augustyniak A and Kosmala A (2016) Water Deficit Affects Primary Metabolism Differently in Two *Lolium multiflorum*/*Festuca arundinacea* Introgression Forms with a Distinct Capacity for Photosynthesis and Membrane Regeneration. *Front. Plant Sci.* 7:1063. doi: 10.3389/fpls.2016.01063

Understanding how plants respond to drought at different levels of cell metabolism is an important aspect of research on the mechanisms involved in stress tolerance. Furthermore, a dissection of drought tolerance into its crucial components by the use of plant introgression forms facilitates to analyze this trait more deeply. The important components of plant drought tolerance are the capacity for photosynthesis under drought conditions, and the ability of cellular membrane regeneration after stress cessation. Two closely related introgression forms of *Lolium multiflorum*/*Festuca arundinacea*, differing in the level of photosynthetic capacity during stress, and in the ability to regenerate their cellular membranes after stress cessation, were used as forage grass models in a primary metabolome profiling and in an evaluation of chloroplast 1,6-bisphosphate aldolase accumulation level and activity, during 11 days of water deficit, followed by 10 days of rehydration. It was revealed here that the introgression form, characterized by the ability to regenerate membranes after rehydration, contained higher amounts of proline, melibiose, galactaric acid, *myo*-inositol and *myo*-inositol-1-phosphate involved in osmoprotection and stress signaling under drought. Moreover, during the rehydration period, this form also maintained elevated accumulation levels of most the primary metabolites, analyzed here. The other introgression form, characterized by the higher capacity for photosynthesis, revealed a higher accumulation level and activity of chloroplast aldolase under drought conditions, and higher accumulation levels of most photosynthetic products during control and drought periods. The potential impact of the observed metabolic alterations on cellular membrane recovery after stress cessation, and on a photosynthetic capacity under drought conditions in grasses, are discussed.

**Keywords:** chloroplast aldolase, drought, forage grasses, membrane regeneration, photosynthetic activity, primary metabolites

**Abbreviations:** F-1,6-2P, Fructose-1,6-diphosphate; GC, gas chromatography; G-6-P, Glucose-6-phosphate; MS, mass spectrometry; MSTFA, *N*-Methyl-*N*-(trimethylsilyl) trifluoroacetamide; PCA, principal component analysis; pFBA, plastid Fructose-1,6-bisphosphate aldolase; RACE, rapid amplification of cDNA ends; ROS, reactive oxygen species; UTR, untranslated region.

## INTRODUCTION

The sedentary life style of plants exposes them to many unfavorable environmental conditions, limiting their growth and development. Abiotic stresses, such as drought (water deficit), salinity, flooding, and low temperature strongly affect plants during their life cycle. Among those factors water deficit is one of the most important, disturbing plant metabolism, inhibiting their growth and reducing productivity worldwide (Bray et al., 2000; Zhang et al., 2005). During their evolution plants developed many strategies for surviving water deficit, namely: drought avoidance, drought tolerance, drought escape, and recovery after drought cessation (Fang and Xiong, 2015). Numerous anatomical, physiological and molecular components of plant performance, including root parameters, leaf features, osmoprotection system, ROS scavenging system, membrane stability and photosynthetic capacity, influence a plant's response to drought (Fang and Xiong, 2015).

Plant productivity depends mainly on photosynthesis, which is one of the first and most sensitive physiological processes affected by drought (Lawlor, 2002; Munns, 2002). Several studies have been conducted in order to improve our knowledge about drought-induced inhibition of photosynthesis (Cornic, 2000; Lawlor, 2002; Chaves et al., 2003; Flexas et al., 2004). These studies have shown that the inhibitory effects of drought on photosynthesis could be associated with low CO<sub>2</sub> availability, due to limitations of its diffusion through the stomata (stomatal limitations) and/or due to non-stomatal limitations, including both diffusive (reduced mesophyll conductance) and metabolic (photochemical and enzymatic limitations) processes (Lawlor, 2002; Lawlor and Cornic, 2002; Flexas et al., 2004). Nevertheless, there is an ongoing debate on whether drought stress influences photosynthesis more by stomatal or non-stomatal alterations (Flexas and Medrano, 2002). Stomatal closure has been identified as an early response to decreasing soil water potential, or to a decline in leaf turgor due to fall in a relative water content, and as an efficient way to reduce water loss in drying field conditions. Simultaneously, a decreasing stomatal conductance limits carbon uptake into leaves, which affects the photosynthesis during mild to moderate drought (Cornic, 2000; Lawlor, 2002). However, during prolonged and more severe periods of drought metabolic limitations, including a reduction of crucial photosynthetic enzymes' activities, may also significantly reduce CO<sub>2</sub> assimilation (Flexas et al., 2004; Signarbieux and Feller, 2011). It was suggested that these limitations are mainly associated with enzymes of the Calvin cycle (Chaves et al., 2003). This relationship was demonstrated for the inhibition of ribulose-1,5-bisphosphate carboxylase (Rubisco; Chaves et al., 2003), decreases in a total Rubisco activity and protein content (Flexas et al., 2006), and an inhibition of Rubisco activase (Lawlor, 2002). Furthermore, the rate of photosynthesis could be limited not only by a carboxylation of Rubisco but also by a ribulose-1,5-bisphosphate regeneration capacity, which could be reduced mainly by a decreased accumulation of plastid fructose-1,6-bisphosphatase (Kossmann et al., 1995), an inhibition of sedoheptulose-1,7-bisphosphatase (Harrison et al., 1998) and a reduced accumulation of chloroplast fructose-1,6-bisphosphate

aldolase (pFBA; Haake et al., 1998, 1999; Uematsu et al., 2012). As it was demonstrated earlier, alterations in photosynthetic carbon metabolism in response to drought could also be strongly associated with accumulation levels of several classes of primary metabolites, mostly plant hormones, osmoprotectants and ROS scavenging particles which are crucial to develop drought tolerance (Chandler and Robertson, 1994). Osmoprotectants are important soluble metabolites and belong to amino acids and carbohydrates, sharing common characteristics, such as small molecular weight and non-toxic character, thus they can be accumulated in large quantities without being harmful for the cell functioning (Rontein et al., 2002). Under osmotic stress plants produce organic osmolytes from a group of soluble sugars, such as fructose or sucrose, organic alcohols, such as *myo*-inositol, complex sugars, such as trehalose or fructans or amino acids, such as proline and modified amino acids, such as glycine betaine (Kido et al., 2013). These compounds could also function as chaperone-like molecules stabilizing membranes and maintaining the activity and stability of the enzymes crucial for a proper functioning of cell metabolism (Xoconostle-Cazares et al., 2010).

*Festuca arundinacea* (tall fescue) is one of the most drought tolerant grass species in the *Lolium-Festuca* complex. *Lolium multiflorum* (Italian ryegrass) has high yielding capacity but significantly lower levels of tolerance to environmental stresses, such as drought, compared to *F. arundinacea*. *L. multiflorum* and *F. arundinacea* hybridization enables the assembly of complementary characters of both species within a single genotype (Kosmala et al., 2012; Perlikowski et al., 2014). The *L. multiflorum*/*F. arundinacea* introgression forms were shown in our earlier work to be excellent plant materials for dissecting drought tolerance of *F. arundinacea* into several crucial components (Perlikowski et al., 2014). Two introgression forms (4/10 and 7/6) were selected for further research performed at different levels to go deeper into the molecular mechanisms of drought tolerance existing in the group of *Lolium-Festuca* forage grasses. The introgression form 4/10 with better yield performance under simulated drought conditions in the field (14 weeks), and a faster re-growth after stress cessation, was also characterized by stronger membrane regeneration during recovery after 11 days of drought in simulated pot conditions. It was manifested by the electrolyte leakage parameter, describing the level of membrane stability. This parameter increased significantly on the 11th day of drought application in the two introgression forms, although after re-watering it returned to the values calculated for the conditions before drought initiation only in the form 4/10. On the other hand, the form 7/6 was characterized by a lower yield potential after 14 weeks of drought in the field, and a lower ability of re-growth after rehydration. However, this form was also shown to possess a greater level of photosynthesis capacity during 11 days of drought treatment in pot conditions, compared to the form 4/10, as manifested by CO<sub>2</sub> assimilation level [ $\mu\text{mol}(\text{CO}_2) \text{ m}^{-2} \text{ s}^{-1}$ ] (values marked below with the same letter did not differ statistically at  $P = 0.05$ , according to Tukey HSD test). This level was significantly higher in the form 7/6 on the 11th day of drought treatment ( $7.42 \pm 0.16b$ ), compared to the

genotype 4/10 ( $6.58 \pm 0.17$ ; Perlikowski et al., 2014). Our earlier work (Perlikowski et al., 2014) demonstrated that more efficient photosynthesis during drought in the form 7/6 was, with a high probability, not associated with the photoactivity performance, since no differences between the analyzed introgression forms in the level of chlorophyll fluorescence parameters were observed under water deficit. Furthermore, it was also shown in our previous work that under drought conditions  $\text{CO}_2$  assimilation rate was not limited by stomatal aperture; both introgression forms significantly reduced stomatal aperture under drought, compared to the control conditions but with a closely similar level of stomatal conductance the form 7/6 revealed a significantly higher level of  $\text{CO}_2$  assimilation rate, compared to the form 4/10. We suggested that this greater capacity of photosynthesis could be due to a higher efficiency of the Calvin cycle in that introgression form. After pre-screening of protein profiles based on 2-D maps, it was found that the accumulation level of pFBA (EC 4.1.2.13) was higher in the form 7/6 (Perlikowski et al., 2014). These two closely related *L. multiflorum*/*F. arundinacea* introgression forms, 4/10 and 7/6, were applied in the research presented herein.

In this study, we hypothesize that (i) the introgression form 7/6, with more intensive  $\text{CO}_2$  assimilation level during drought, will be characterized by higher pFBA accumulation and activity levels, which could be a crucial component of non-stomatal machinery involved in a regulation of photosynthetic efficiency in the *Lolium-Festuca* forage grasses; (ii) this phenomenon will also be accompanied by a higher accumulation level of primary photosynthetic metabolites in this form. Moreover, we hypothesize that (iii) a stronger regeneration capacity of the introgression form 4/10, including a membrane regeneration process after stress cessation, could be associated with higher accumulation levels of key metabolites, including osmoprotectants responsible for a protection of crucial proteins, and other important cell components. Thus, the research presented herein performed on the two introgression forms — 4/10 and 7/6 involved: (i) Western blot experiments to confirm pFBA accumulation level during water deficit and recovery periods, accompanied by pFBA activity measurements and (ii) a primary metabolite profiling under stress conditions and after stress cessation, using GC - MS.

## MATERIALS AND METHODS

### Plant Materials

Plant materials used in the present research involved two *L. multiflorum*/*F. arundinacea* introgression forms (genotypes 4/10 and 7/6) obtained after four rounds of backcrossing of *L. multiflorum* (4x)  $\times$  *F. arundinacea* (6x) hybrid to *L. multiflorum* (4x). These plants were selected earlier from a larger population in the field conditions with respect to their drought tolerance, as described by Perlikowski et al. (2014). After this selection, the two forms, each one in four biological replicates, were transferred to pots (1.75 dm<sup>3</sup>), containing a sand:peat (1:3) mixture. The experiment of 11 days of water deficit, followed by 10 days of re-watering was performed in a

growth chamber at a temperature of 22/17°C (16 h day/8 h night, light of 400  $\mu\text{mol}(\text{quanta}) \text{m}^{-2} \text{s}^{-1}$ , HPS “Agro” lamps, Philips, Brussels, Belgium), 30% relative air humidity and watering completed. The level of soil water content decreased from 63% of field water capacity observed in control conditions down to approximately 3% on the 11th day of stress duration. After 10 days of re-watering this capacity increased to the value observed before drought treatment (Perlikowski et al., 2014). The physiological measurements summarized in the introduction, were performed during this experiment. The leaf tissue sample (100 mg) was collected before drought treatment (control), at three different time-points of drought (after 3, 6, and 11 days of drought), and after 10 days of subsequent re-watering, every time from each replicate, and frozen in liquid nitrogen.

### Identification of Aldolase cDNA Sequences

Full length cDNA sequences encoding chloroplast 1,6-bisphosphate aldolase (pFBA) were obtained by RACE reaction using commercial kit (5'/3' RACE Kit, 2nd Generation - ROCHE®). Initial primers (forward primer – TTCGAGGAG ACCCTCTACCA; reverse primer – GGCTACAGTGCCTCT CAAG) were designed on the basis of pFBA mRNA sequence of *Brachypodium distachyon* available in NCBI database [gi|357157398|ref|XM\_003577737.1|]. Special primers for RACE reaction were designed on the basis of sequenced initial fragment of pFBA cDNA:

SP-F1 - GACTGTAGATGGCAAGAAGATTGTTGAC  
 SP-F2 - CCAATTGTTGAGCCTGAGATCATG  
 SP-R1 - CTACTAGCACTCTCTCCATAGGTAGATA  
 SP-R2 - ATCAGTAGCTGTAGTTCTTGACGAACAT  
 SP-R3 - TTCTCTGGAGGAGCTTGAGAGTGTA.

The PCR and RACE products were purified by QIAEXII Gel Extraction Kit (Qiagen), and ligated into the pGEM-T Easy vector (Promega). Vectors containing the ligated product were transformed into *Escherichia coli* strain XL1 Blue, and multiplied plasmids from clones selected with X-Gal and IPTG were extracted using QIAprep Spin Miniprep Kit (Qiagen). The obtained plasmids were sequenced (Molecular Biology Techniques Laboratory, Faculty of Biology, Adam Mickiewicz University, Poznań) using SP6 and T7 primers. The obtained cDNA sequences were aligned with BioEdit software (ver 7.2.5).

### Analysis of Aldolase Accumulation Level

To estimate a pFBA protein accumulation level a Western blot analysis was performed, with the antibody directed against pFBA. The antibody was produced by Agrisera® company<sup>1</sup> using a rabbit host immunized with a highly specific pFBA 15 amino acid peptide (TFEVAQKVWAETFY). The peptide was selected on the basis of comparison between

<sup>1</sup>www.agrisera.com

pFBA sequence of the two analyzed introgression forms, and available in database sequence of a cytosolic enzyme of *B. distachyon* (XM\_003564823.1) to avoid a cross-reaction of the anti-pFBA antibody with a cytosolic FBA. The detailed protocols for a protein extraction and Western blotting were as described by Pawłowicz et al. (2012). Briefly, 10 µg of chloroplast proteins from each time-point and introgression form in three biological replicates and standard samples were separated by 12% SDS-polyacrylamide gel electrophoresis and electroblotted onto nitrocellulose membranes (Bio-Rad). Immunodetection was performed with a rabbit polyclonal antibody (diluted 1:4000; Agrisera). The antigen-antibody complexes were detected using a chemiluminescent detection system with a secondary anti-rabbit IgG-horseradish peroxidase conjugate (diluted 1:20 000; Sigma) and a chemiluminescent substrate (Westar Supernova – Cyanogen) and the products intensities were estimated using ImageJ software.

## Analysis of Aldolase Activity

The pFBA activity was measured according to a modified Sibley-Lehninger method (Sibley and Lehninger, 1949; Willard and Gibbs, 1968). A protein extract from chloroplasts was prepared according to a modified method used by Kosmala et al. (2012). Briefly, 1g of frozen leaf material in three replicates for each sample, was ground in a liquid nitrogen, suspended in 4 ml of chloroplast isolation buffer (Sigma-Aldrich), shaken, filtered through a mesh 100 nylon (Sigma-Aldrich) and centrifuging 3 min at 200 g at 5°C. The collected supernatant was subsequently centrifuged for 15 min at 900 g at 5°C, and the washed chloroplast pellet was suspended in 2 ml 0.1 M phosphate buffer (0.1 M Na<sub>2</sub>HPO<sub>4</sub>) with 3% Triton X100 and shaken 5 min in 1000 rpm. The collected supernatant was used to determine the pFBA activity. To 2 ml tubes 50 µl of 0.06 M fructose-1,6-bisphosphate and 140 µl of incubation buffer (0.05 M 2,4,6 trimethylpyridine, 0.08 M hydrazine sulfate, 0.3 mM sodium iodoacetate) pH 7.4 were added and pre-incubated in water bath during 3 min at 30°C. The volume of 100 µl of chloroplast extract was added and incubated at 30°C for 2 h. After the incubation, tubes were chilled and 300 µl of 10% trichloroacetic acid was added to stop the reaction. For each sample one tube was treated as a reagent blank and was filled with 300 µl of 10% trichloroacetic acid before proceeding. After centrifugation at 10000 g, 100 µl of collected supernatant was pre-incubated with 100 µl of 0.75 M NaOH at room temperature for 10 min and then incubated at 30°C water bath for 10 min with addition of 100 µl of 0.1% 2,4-dinitrophenylhydrazine. After that step, samples were mixed with 700 µl of 0.75 M NaOH and the absorbance measurements were performed with reference to a reagent blank using a spectrophotometer with 540 nm wavelength. A standard curve was prepared as follows: 2 ml tubes in two replicates were filled in order with 25, 50, 75, and 100 µl of 0.01 mM D-glyceraldehyde and filled up with water to a final volume of 100 µl. In the next step, 100 µl of 2,4-dinitrophenylhydrazine solution was added and samples were incubated at 30°C water bath for 10 min. After incubation, 800 µl of NaOH was added and after 3 min of incubation the absorbance was measured

with 540 nm with reference to a blank sample (100 µl of water plus reagents). The amount of produced trioses in the pFBA assay was read according to a standard curve and presented after calculation as µg of glyceraldehyde produced by 1 g of plant sample during 1 h.

## Metabolite Profiling

Analysis of primary metabolites accumulation was performed with slight modifications according to the protocol described earlier by Wojakowska et al. (2015). This protocol is presented briefly in the following sub-sections.

## Materials and Reagents

Solvents used for extraction and GC-MS analyses were MS grade methanol, methylene chloride, isopropanol, ribitol; derivatization reagents for GC-MS analyses were - MSTFA, O-methylhydroxylamine hydrochloride, pyridine and alkanes (C10–C36) used as retention index standards purchased from Sigma-Aldrich (Poznań, Poland). A deionized water was purified by Milli-Q system Direct Q3 (Millipore, Bedford, MA, USA). A homogenization was performed with the MM400 (Retsch GmbH, Haan, Germany) homogenizer. Centrifugation was done with the EBA21 centrifuge (Hettich, Tuttlingen, Germany).

## Extraction of Metabolites

The amount of 100 mg of dried, powdered leaf material sample was transferred to 2 ml plastic tubes with two stainless steel balls, and 1.5 ml of 80% methanol in deionized water was added. Ribitol (25 µl of 1 mg/ml solution) was added to each sample as the internal standard. The samples were homogenized for 10 min at 1800 rpm, sonicated for 15 min and centrifuged for 15 min at 12000 rpm, followed by filtering through PTFE syringe filters 0.45 µm GHP ACRODISC 1 (Waters, Milford, CT, USA). The volume of 300 µl of each sample was transferred to a new tube and evaporated in a SpeedVac concentrator. A dried extract was then derivatized with 50 µl of methoxyamine hydrochloride in pyridine (20 mg/ml) at 37°C for 90 min with agitation. The second step of derivatization was performed by adding 80 µl of MSTFA and an incubation at 37°C for 30 min with agitation. Samples were subjected to GC/MS analysis directly after a derivatization. Each sample was prepared in four replicates. The compounds were considered “identified” when they met the identification criteria established by the GC software used (LECO ChromaTOF), namely: identity score higher than 700, Mass Threshold higher than 10 and a matched retention index.

## GC/MS Analysis

The analysis for separation was performed using the Agilent 7890A gas chromatograph (Agilent Technologies) connected to Pegasus 4D GCxGC-TOFMS mass spectrometer (Leco). A DB-5 bonded-phase fused-silica capillary column of 30 m length, 0.25 mm inner diameter and 0.25 µm film thickness (J&W Scientific Co., USA). The GC oven temperature program was as follows: 2 min at 70°C, raised by 8°C/min to 300°C and held for 16 min at 300°C. The total time of GC analysis was 46.75 min.



Helium was used as the carrier gas at a flow rate of 1 ml/min. The volume of 1  $\mu$ l of each derivatized sample was injected in a splitless mode. The initial PTV (Programmed Temperature Vaporization) injector temperature was 20°C for 0.1 min and then raised by 600°C/min to 350°C. The septum purge flow rate was 3 ml/min and the purge was turned on after 60 s. The transfer line and ion source temperatures were set to 250°C. In-source fragmentation was performed with 70 eV energy. Mass spectra were recorded in the mass range 35–650 m/z.

## Analysis of Mass Spectra

Data acquisition, automatic peak detection, mass spectrum deconvolution, retention index calculation and library search were done by Leco ChromaTOF-GC software (v4.51.6.0). To eliminate retention time shift and to determine the retention indexes (RI) for each compound, the alkane series mixture (C-10 to C-36) was injected into the GC/MS system. The metabolites were automatically identified by library search (Replib, Mainlib, Fiehn library) with a similarity index above 700 and retention index  $\pm 10$ . All known artifact peaks including alkanes, plasticizers, column bleed, MSTFA artifact and reagent peaks were not considered in the final results. To obtain accurate peak areas for the deconvoluted components, unique quantification masses for each component were specified and the samples were reprocessed. The obtained metabolite data was normalized relatively to the quant mass peak of internal standard (Ribitol – 217) in each sample before statistical analysis.

## Statistical Analysis

The normalized mass spectral intensity Log-transformed data (base 2) was subjected to statistical analysis. Two-way analysis of variance (ANOVA) with a genotype and treatment as classification factors, Fisher's least significant difference (LSD) and PCA was made using STATISTICA 10 software (StatSoft, Tulsa, OK, USA). The PCA was carried out by eigenvalue decomposition of data correlation matrix. The significant effects of genotype, time and genotype  $\times$  time interaction were selected using the family wise error rate less than 1%. Fisher's LSD of samples at 1% was used. Heatmaps for a difference between means of time-points and the control were prepared.

## RESULTS

### The Accumulation Level and Activity of Chloroplast Fructose Bisphosphate Aldolase during Drought and Rehydration

The RACE analysis performed on total RNA extracted from the two introgression forms allowed the identification of two pFBA mRNA sequences in each form (**Supplementary Figure S1**). The overall length of the identified sequences with 3' and 5' UTR varied from 1423 to 1430 nucleotides, whereas the length of the coding region was the same for each identified sequence

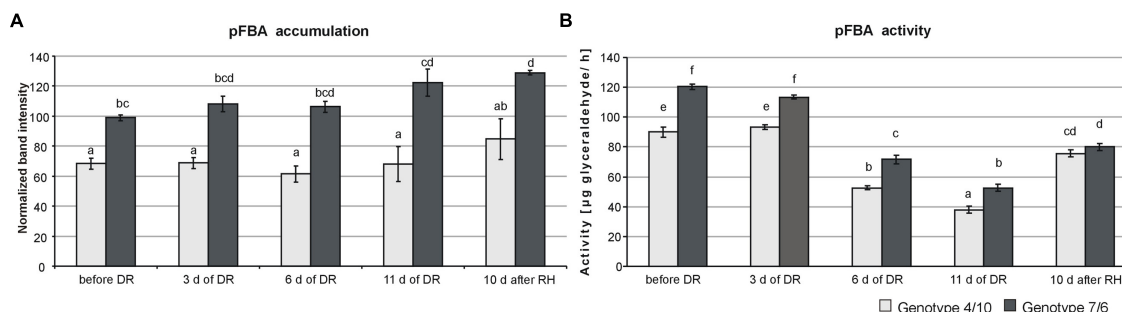
and covered 1164 nucleotides. This coding region showed 99% of similarity to *B. distachyon* sequence used earlier for primers designed to clone pFBA from the analyzed introgression forms (**Supplementary Figure S1**). The identified sequences were characterized by several single nucleotide polymorphisms within a coding region, and modifications within 3' and 5' UTR (**Supplementary Figure S1**). The predicted protein sequence for the analyzed mRNAs covered 388 amino acids, including 37 amino acids of the chloroplastic transit sequence (**Supplementary Figure S2**). The predicted molecular mass of the protein was 42.05 and 38.48 kDa with and without a transit sequence, respectively. Its predicted isoelectric point was 5.51, and between the different predicted here protein sequences three amino acid modifications were found (**Supplementary Figure S2**).

The proteomic assays revealed a significantly higher accumulation level (**Figure 1A**) and total activity (**Figure 1B**) of pFBA in the 7/6 introgression form at all the time-points of drought treatment, compared to the 4/10 form. However, in both forms a significant decrease of pFBA activity after 6 and 11 days of drought was observed, compared to the control conditions and initial days of stress duration. On the other hand, after 10 days of rehydration an increase of enzyme activity was revealed but without significant differences between the two forms (**Figure 1B**).

### The Accumulation Level of Primary Metabolites during Drought and Rehydration Metabolite Accumulation Dynamics

A total of 937 different metabolite compounds were identified, and 66 were selected for further analysis. These 66 were present in all the biological replicates, and had a similarity index value above 700 and/or were manually identified using a comparison of their retention time with Golm metabolite VAR5 library data<sup>2</sup>, and after one-way ANOVA pre-selection had *p*-value lower than 0.01 (**Supplementary Table S1**). The analyzed metabolites were further divided into nine classes, including amino acids, amines, sugars, sugar acids, sugar alcohols, phosphoryl compounds, organic acids, alcohols, and fatty acids (**Figure 2**). The further statistical analysis revealed that 50 compounds presented significant genotype dependent differences, 63 time-point dependent differences and 60 compounds revealed a significant interaction between a genotype and a time-point (**Supplementary Table S1**). The introgression form 7/6 was characterized by a significantly higher accumulation level of primary metabolites, including 47 in the control conditions and 32 after 11 days of drought, compared to the 4/10 form. This latter plant expressed a higher accumulation level of six metabolites in the control conditions and 10 after 11 days of drought, compared to the form 7/6. Only after rehydration did the form 4/10 reveal 28 metabolites with a significantly higher accumulation level than the form 7/6. At this time-point the form 7/6 expressed only 14

<sup>2</sup><http://gmd.mpimp-golm.mpg.de>



**FIGURE 1 |** The activity and accumulation levels of chloroplast fructose biphosphate aldolase (pFBA) at five time-points: before drought, after 3, 6 and 11 days of drought (DR), and 10 days of re-watering (RH) in *Lolium multiflorum*/*Festuca arundinacea* introgression forms. The bars represent a calculated mean value of enzyme activity expressed in  $\mu\text{g}$  glyceraldehyde produced by pFBA during 1 h of incubation at  $30^\circ\text{C}$  (A) and a mean of pFBA bands intensities values (B). Error bars represent standard deviations. The letters indicate groups of means that do not differ significantly at a significance level of 0.01 (Fisher's LSD-test).

metabolites with higher accumulation levels than the form 4/10 (Figure 3A).

The analysis of changes in accumulation levels between the time-points in these two introgression forms revealed that during drought duration the patterns of metabolite dynamics were different in both plants. Between the control conditions and drought time-points more metabolites were significantly down-regulated in the 7/6 form, especially at the beginning of drought treatment between – the control and the 3rd day of stress period, with 36 metabolites decreasing their accumulation levels. The abundance of these metabolites started to be up-regulated between the 3rd and 11th day of drought, with 35 metabolites increasing their abundance (Figures 3B and 4). Different patterns of metabolite dynamics were observed in the 4/10 form. More metabolites were up-regulated than down-regulated between all the time-points, but this phenomenon was most visible between the control and 11th day of drought period, with 34 metabolites increasing their abundance (Figures 3C and 4). In the same period, only 21 metabolites increased their accumulation level in the form 7/6 (Figures 3B and 4). Seventeen of these metabolites overlapped in both introgression forms (Figure 4). After rehydration, 44 metabolites presented a higher accumulation level, compared to the control conditions, in the form 4/10 (Figure 3C). Twenty-six of them belonged to the group of metabolites which significantly increased their abundance after 11 days of drought period and remained this elevated accumulation level or even increased it after rehydration. This trend was not observed in the 7/6 form, with only eight metabolites remaining the elevated abundance after rehydration (Figure 4). Moreover, 18 more metabolites, mainly carbohydrates, which were not accumulated under drought conditions, increased their abundance significantly after rehydration in the 4/10 form. This phenomenon was not observed in 7/6 form (Figure 4). These relations were also visible in the PCA, where the Principal Component 1 accounted for 47% of the variance, clearly separated the two forms with respect to the control values for the analyzed metabolites and indicated that during drought duration and rehydration the patterns of metabolite dynamics were different in both plants.

For this separation, the highest contribution revealed not only the metabolites from the correlation groups number 4, 5, and 6 with most carbohydrates and both substrates and products of photosynthesis but also from the group number 3 with amino acids (Figure 5 and Supplementary Table S2).

### Amino Acids and Amines

The accumulation of amino acids from the correlation group number 3 (Figure 5) revealed a significant impact on genotypes and time-points separation according to the PCA Component 2 values (Supplementary Table S2). Although, a clear separation of genotypes was not in fact observed, it was simultaneously noticed that after 6 and 11 days of drought treatment in both introgression forms the accumulation level of most amino acids significantly increased, compared to control values (Figures 2, 4 and 5). A higher accumulation level of aspartic acid, isoleucine, lysine, methionine, tryptophan and tyrosine in response to drought application in both introgression forms, and asparagine only in the form 7/6, was revealed. On the other hand, alanine, arginine, proline and valine increased their accumulation level after 11 days of drought only in the form 4/10 (Figures 2 and 4). These results demonstrated that the proline accumulation level unexpectedly decreased at the beginning of drought (the 3rd and the 6th day of drought) in both analyzed grass forms. However, when drought treatment progressed, proline abundance returned to the values observed in the control conditions in the form 7/6, and was even higher in the form 4/10 (Figure 2). The accumulation pattern of dimethylglycine-like amine was different between the two forms. In the form 4/10, its increased accumulation level after 11 days of drought without further change after rehydration, was revealed. Contrary, in the 7/6 form, a progressive decrease in dimethylglycine-like amine accumulation during drought period, compared to the control conditions, and a slight increase after rehydration to the level observed in the 4/10 form, was noticed (Figure 2).

### Carbohydrates

With respect to the accumulation profiles of the identified carbohydrates, including substrates and products of

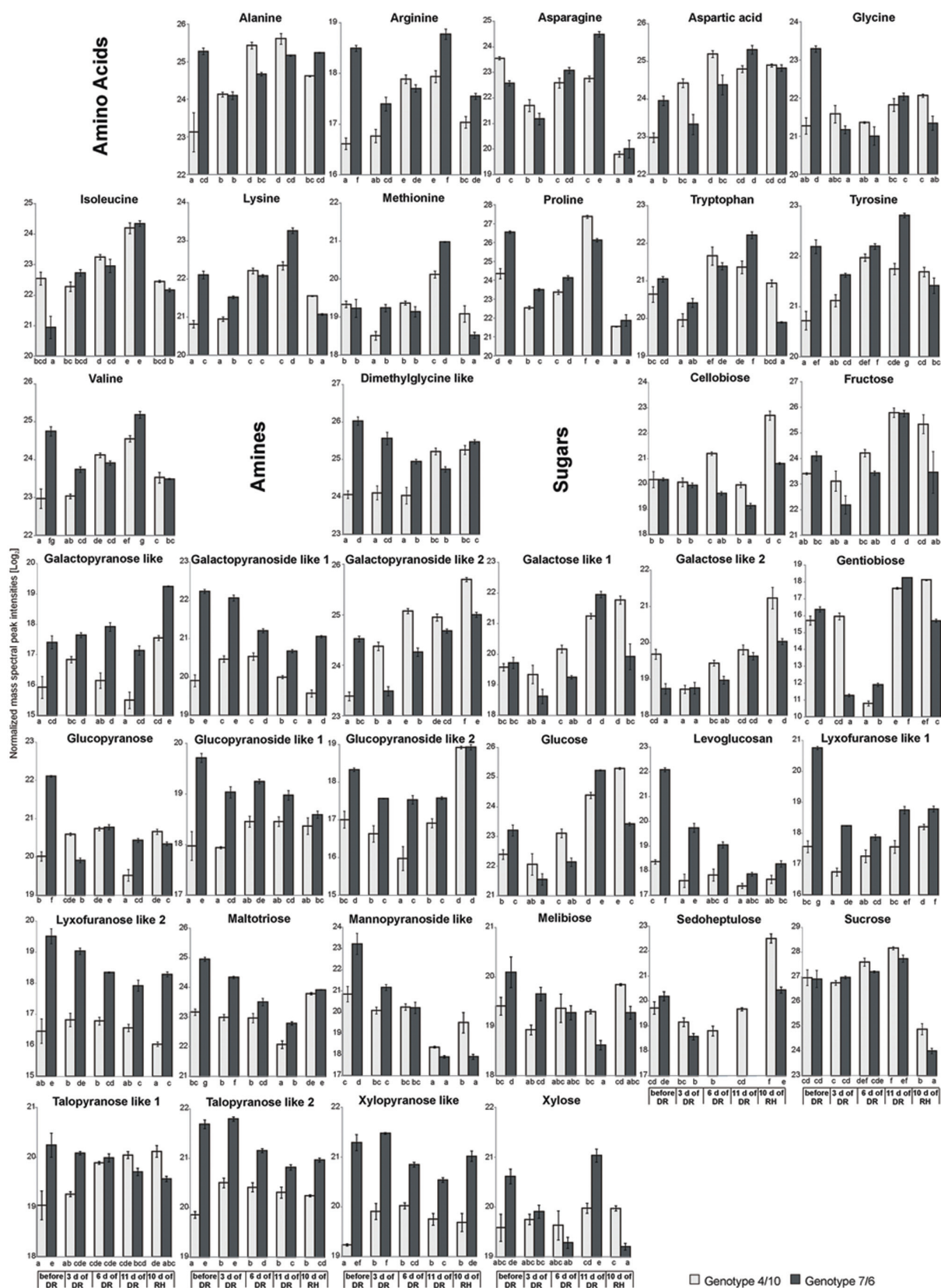
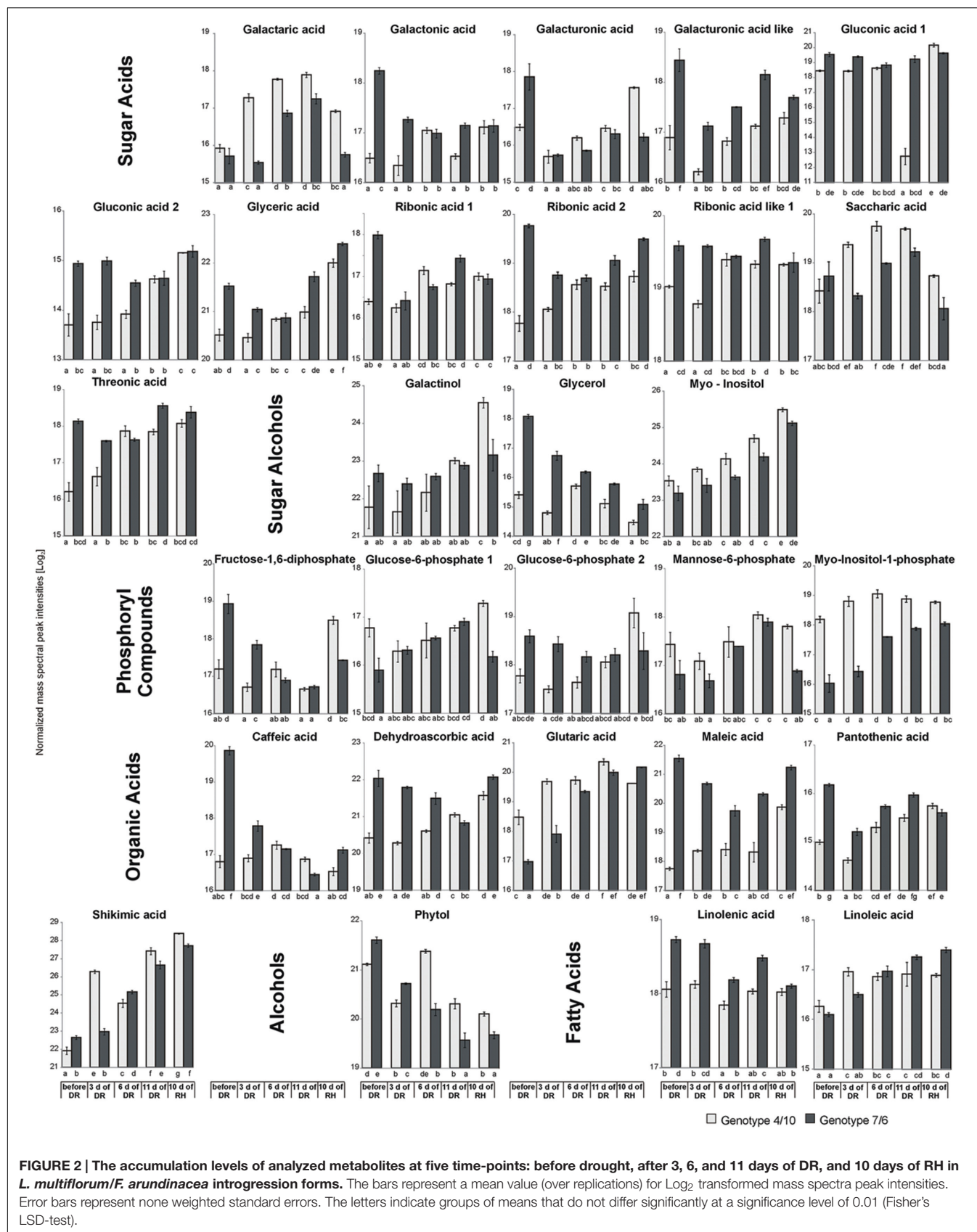
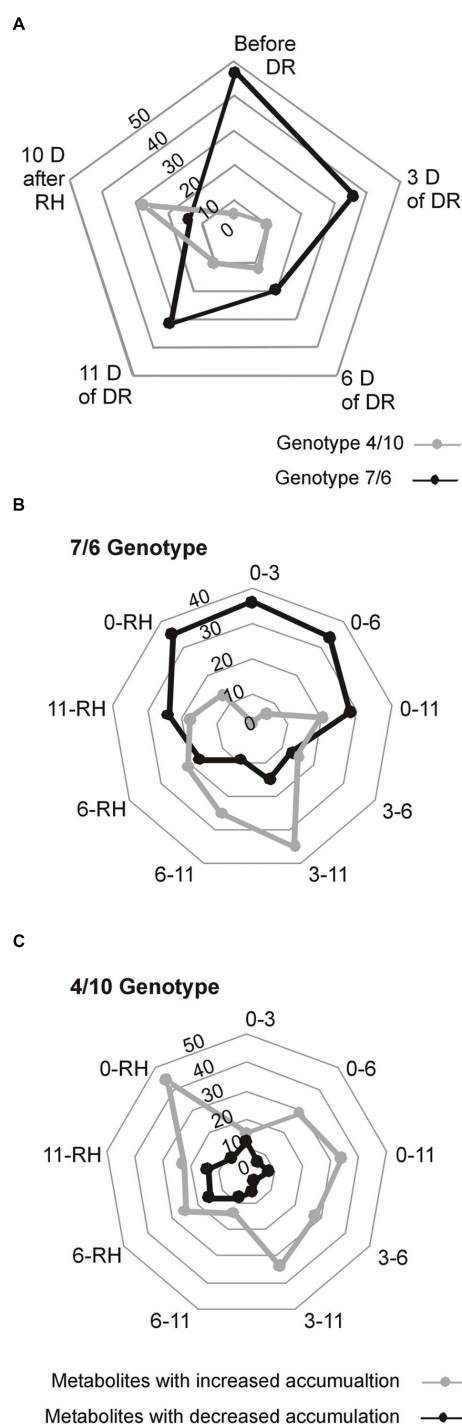


FIGURE 2 | Continued



**FIGURE 2 |** The accumulation levels of analyzed metabolites at five time-points: before drought, after 3, 6, and 11 days of DR, and 10 days of RH in *L. multiflorum*/*F. arundinacea* introgression forms. The bars represent a mean value (over replications) for Log<sub>2</sub> transformed mass spectra peak intensities. Error bars represent none weighted standard errors. The letters indicate groups of means that do not differ significantly at a significance level of 0.01 (Fisher's LSD-test).





**FIGURE 3 | The overall statistical comparison of differences in accumulation levels of the analyzed metabolites at five time-points: before drought, after 3, 6, and 11 days of DR, and 10 days of RH in *L. multiflorum*/*F. arundinacea* introgression forms. (A)** Numbers of metabolites with a significantly higher accumulation level between the analyzed forms at the particular time-points. **(B)** Numbers of metabolites with a significant increase or decrease in an accumulation level between time-points in the 7/6 form. **(C)** Numbers of metabolites with a significant increase or decrease in an accumulation level between time-points in the 4/10 form.

photosynthesis, it was revealed that under the control conditions the form 7/6 had a higher accumulation level of F-1,6-2P, G-6-P 2 and glucose, compared to the form 4/10. The accumulation levels of sucrose and fructose did not differ significantly between the two forms, and only G-6-P 1 showed a higher abundance in the 4/10 form. However, after 11 days of drought the accumulation of F-1,6-2P and G-6-P 2 decreased, and G-6-P 1 and mannose-6-phosphate increased in the 7/6 form to the levels observed in the 4/10 form (**Figure 2**). For the main photosynthesis products: glucose, fructose, and sucrose a significant increase of accumulation level after 11 days of drought in the analyzed introgression forms was observed. No significant differences between the plants in fructose and sucrose accumulation levels were revealed, and only glucose showed a significantly higher abundance in the 7/6 form (**Figure 2**). After rehydration, in the 4/10 form all these compounds demonstrated higher accumulation levels, compared to the 7/6 form.

Other identified here carbohydrates presenting significant differences in an accumulation level between the analyzed introgression forms were *myo*-inositol and *myo*-inositol-1-phosphate. Their abundance increased in response to drought in both introgression forms, but was significantly higher in the form 4/10 (**Figure 2**). The accumulation level of gentiobiose increased after 11 days of drought period in both forms being higher in the 7/6 form. Melibiose and glycerol revealed a gradual decrease of their abundance during drought period in the form 7/6 and more or less stable level in the 4/10 (**Figure 2**).

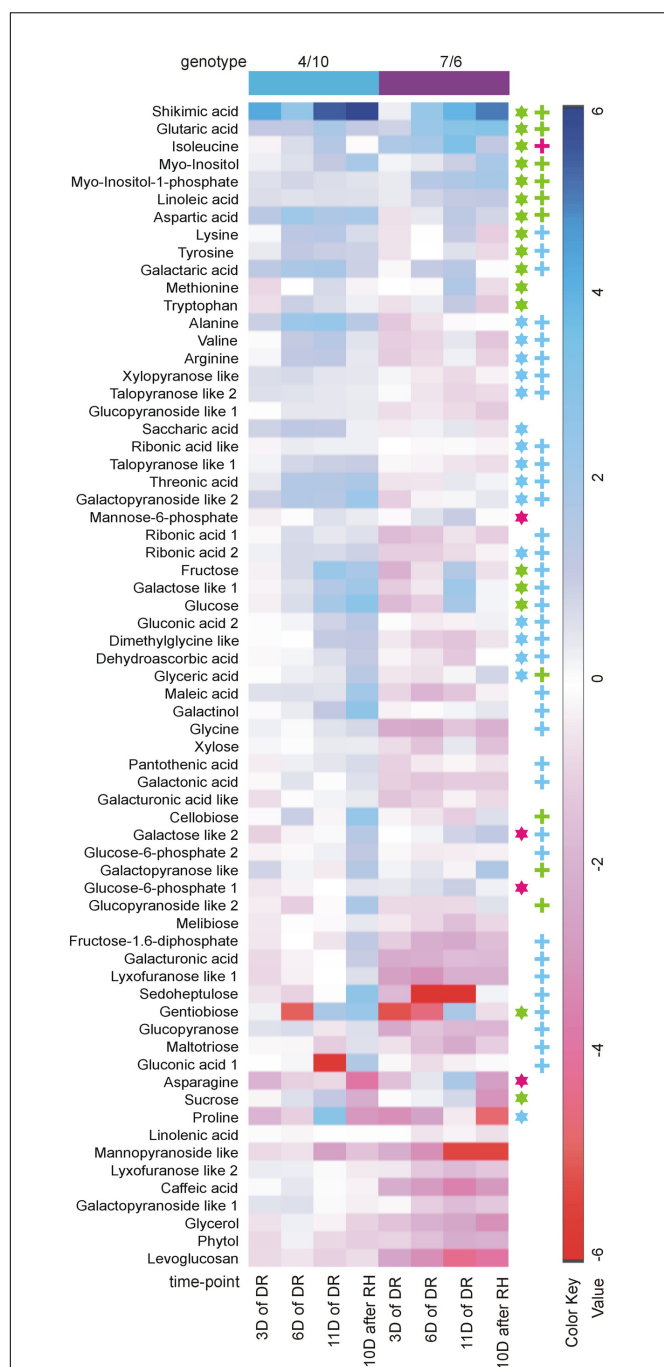
### Other Compounds

Sugar acids, such as galactaric acid, gluconic acid 2, glyceric acid, ribonic acid 2, ribonic acid like 1, saccharic acid and threonic acid revealed a significant increase of accumulation level after 11 days of drought in the 4/10 form, while in the form 7/6 rather a stable or even a decreased accumulation level in majority of sugar acids, was observed. Only galactaric acid showed an increased accumulation level during drought in this introgression form (**Figure 2**). The analyzed accumulation patterns for organic acids revealed differences between two forms. Five of six analyzed organic acids presented a higher accumulation level under drought treatment and rehydration in the form 4/10. In the form 7/6, caffeic acid, dehydroascorbic acid, and maleic acid showed a significant reduction in an accumulation level under drought. Only glutaric acid and shikimic acid accumulated significantly during drought treatment in this introgression form (**Figure 2**).

## DISCUSSION

### Primary Metabolite Accumulation Profiles with Respect to Chloroplast Fructose Bisphosphate Aldolase Activity and Photosynthetic Capacity

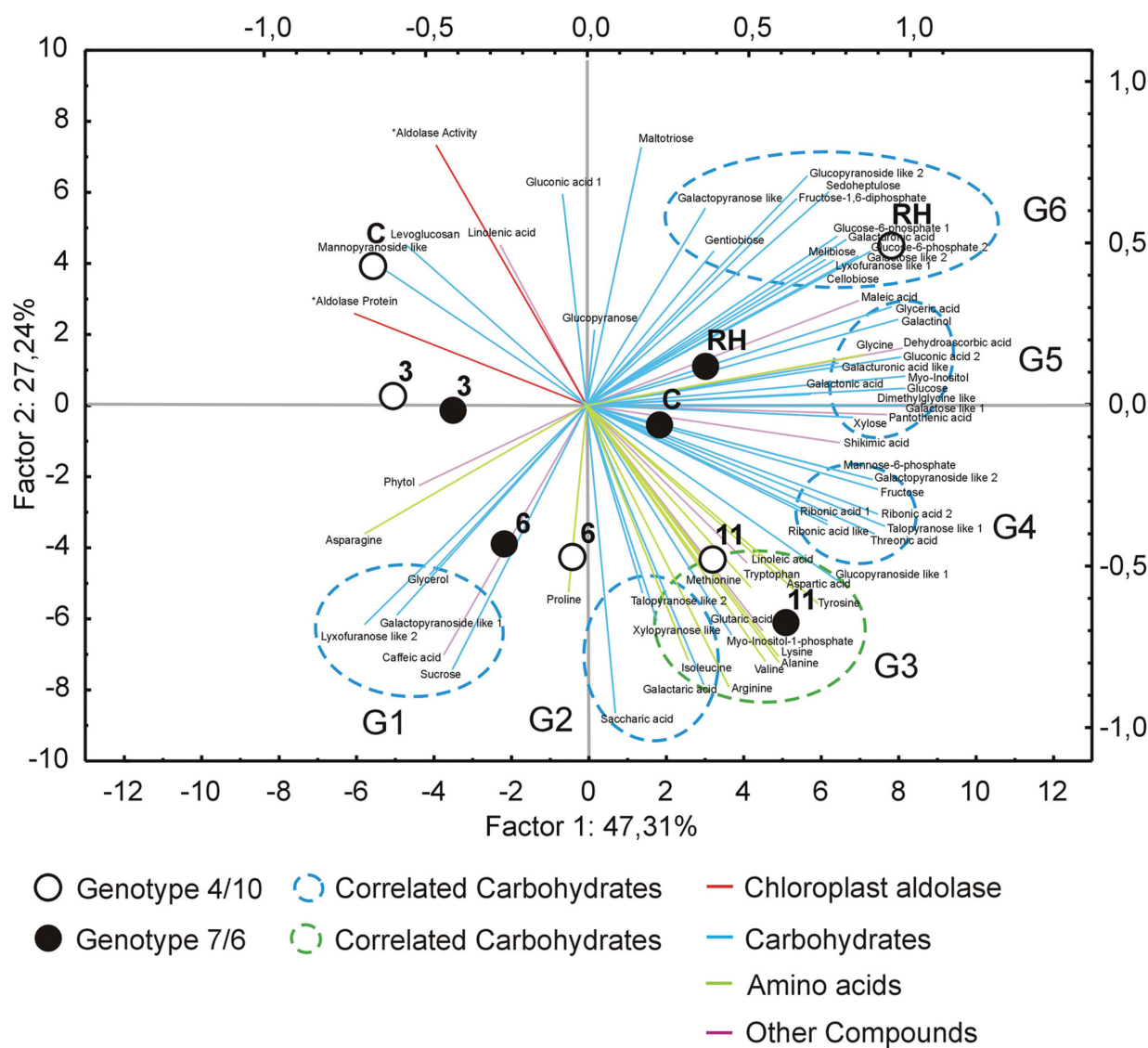
Under mild water deficit conditions stomata closure is the main physiological factor responsible for inhibiting photosynthesis through a reduction of CO<sub>2</sub> availability for the assimilation



**FIGURE 4 | The accumulation levels of the analyzed metabolites (relative to control values, calculated for mean Log<sub>2</sub> transformed mass spectral peak intensities) at four time-points of the experiment: after 3, 6, and 11 days (D) of DR, and after 10 days of RH in the 4/10 and 7/6 *L. multiflorum*/F. *arundinacea* introgression forms.** The values lower than the control are shown in shades of red, and the values higher than the control are shown in shades of blue. The asterisks indicate metabolites with a significant increase of an accumulation level after 11 days of drought: blue in the 4/10 form, magenta in the 7/6 form and green in both introgression forms. The crosses indicate metabolites with a significantly higher accumulation level after 10 days of re-watering: blue in the 4/10 form, magenta in the 7/6 form and green in both introgression forms.

process (Lawlor, 2002). However, an important recent topic for discussion has become the mechanisms of non-stomatal limitations of photosynthesis associated also with metabolic factors (Flexas and Medrano, 2002), such as the Calvin cycle enzymes, among which pFBA accumulation level plays a crucial role (Haake et al., 1998, 1999; Uematsu et al., 2012). The pFBA is a key enzyme of the regeneration phase of this cycle, and its activity may be important for the regulation of photosynthesis intensity (Haake et al., 1998; Raines, 2003). It initiates the third regeneration phase of the Calvin cycle by catalyzing the reversible conversion of glyceraldehydes-3-phosphate and dihydroxyacetone phosphate to fructose-1,6-bisphosphate (F-1,6-2P), and condensation of sedoheptulose-1,7-bisphosphate from erythrose-4-phosphate and dihydroxyacetone phosphate (Raines and Lloyd, 2007). It has been proved in several studies that a changed accumulation level of pFBA in plants could influence the efficiency of photosynthesis (Haake et al., 1998, 1999; Uematsu et al., 2012). However, this relationship has not been confirmed for the forage grasses. Here, we try to fill this gap in our knowledge. Photosynthesis is the main source of substrates for carbohydrates, amino acids, and other cellular compounds production. These metabolites serve not only as a cell energy reservoir but often also as signaling molecules important for growth, development, and response to unfavorable conditions (Dolferus, 2014). Thus, an inhibition of photosynthesis usually has its reflection in growth, biomass production, and metabolite accumulation disturbances (Chaves et al., 2010). Perennial grasses accumulate large quantities of soluble carbohydrates in their leaves, simultaneously having a low amount of starch (Pollock and Cairns, 1991; Cairns, 2002). The observed higher accumulation level of most primary metabolites under control conditions and drought treatment, including glucose in the form 7/6, was probably due to a higher efficiency of photosynthesis in this form. This validates one of our hypotheses formulated in the introduction section. On the other hand, a higher accumulation level of particular primary metabolites, including all the analyzed substrates and products of photosynthesis after re-hydration accompanied a higher potential of recovery in the 4/10 form (Perlikowski et al., 2014).

Our results indicated that water deficit had a significant impact on the pFBA activity in the two analyzed introgression forms. However, after further re-watering that activity was closer to the control values (before stress application) in the 4/10 form, indicating its higher potential of metabolism recovery, after stress cessation. It was previously observed that the impaired expression of pFBA in transgenic lines of *Solanum tuberosum* (potato) affected plant growth, and negatively influenced carbohydrates accumulation, including phosphoryl substrates for photosynthesis and fructose, glucose, sucrose, and starch in ambient conditions. A lower level of pFBA accumulation also influenced the activity of other Calvin cycle enzymes, such as plastid fructose-1,6-bisphosphatase (Haake et al., 1998, 1999). However, the impact of decreased pFBA activity on photosynthesis rate and metabolite accumulation was observed only when less than 50% of pFBA activity was removed, but even under those conditions the accumulation levels of most important hexose phosphates were not impaired (Haake et al.,

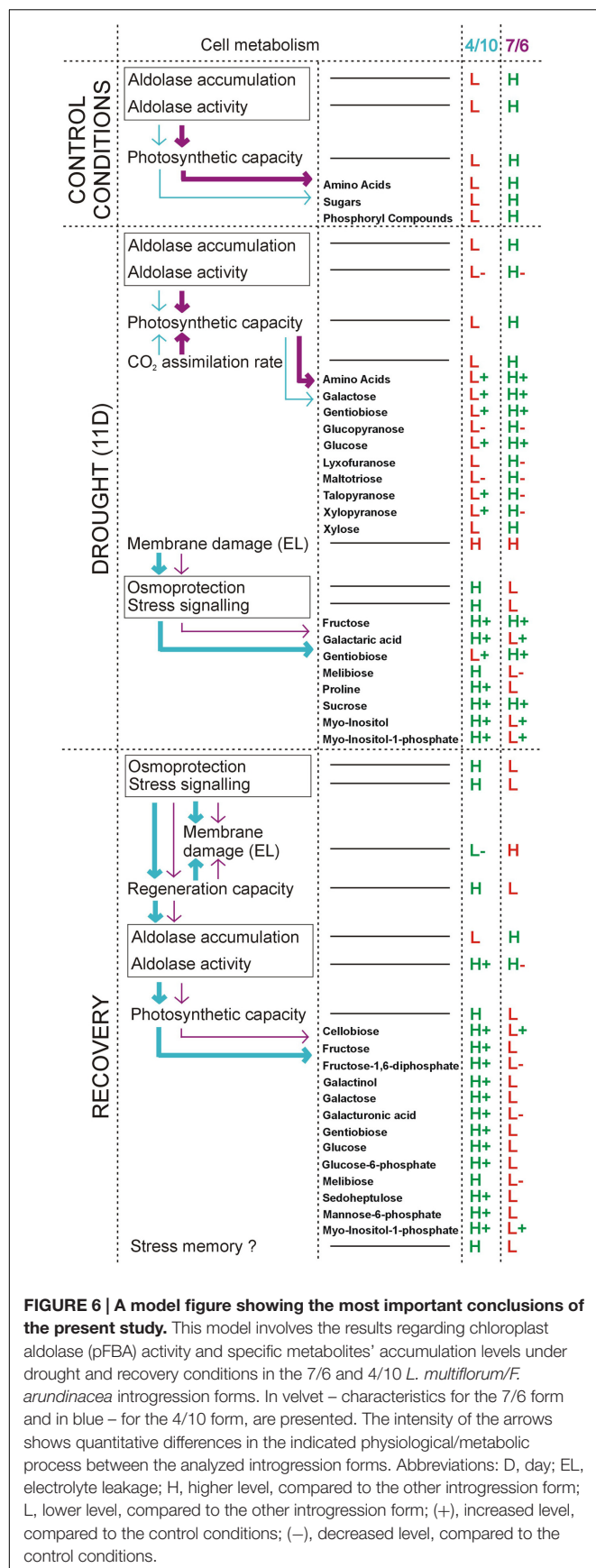


**FIGURE 5 | The PCA of the analyzed metabolites accumulation level using the first two principal components with the highest variance.** Close circles represent the 7/6 *L. multiflorum*/*F. arundinacea* introgression form, open circles – the 4/10 form. The time points are labeled, as follows: C, control; 3, the 3rd day of drought; 6, the 6th day of drought; 11, the 11th day of drought; RH, 10 days after re-hydration. Blue lines represent carbohydrate metabolites, green – amino acids, purple – other metabolites. Dotted line circles represent correlation groups (G1–G6).

1998). In the present study, the impact of higher activity of pFBA on a higher photosynthesis efficiency, and a higher metabolite accumulation level in the 7/6 form, could be suggested, but in fact the difference between the analyzed plants for the activity of this enzyme was less than 25%. Also, the research performed on transgenic *Nicotiana tabacum* (tobacco) with overexpression of pFBA (Uematsu et al., 2012) clearly showed that an increased accumulation level of pFBA in plant tissues enhanced plant growth, biomass production, CO<sub>2</sub> assimilation rate and ribulose-1,5-bisphosphate accumulation. Both analyzed here introgression forms reduced significantly their stomatal conductance during drought period, compared to the control conditions, and this process was associated with a reduction of CO<sub>2</sub> assimilation

levels in these two plants. However, as proved also in our earlier work, with closely similar levels of stomatal conductance during drought, the form 7/6 revealed simultaneously a significantly higher level of CO<sub>2</sub> assimilation rate, compared to the form 4/10 (Perlikowski et al., 2014). In the current study, we demonstrated that this phenomenon was accompanied by higher accumulation and activity levels of pFBA in the form 7/6. Further research is required on the aldolase gene expression level using qRT-PCR as well as on the expression of genes coding other enzymes of the Calvin cycle. As far as we know, this report is the first one for the *Lolium-Festuca* forage grasses demonstrating crucial components of non-stomatal regulation mechanisms of CO<sub>2</sub> assimilation level and photosynthetic capacity under drought conditions.





## Primary Metabolite Accumulation Profiles in Response to Water Deficit Conditions

As per our hypothesis, water deficit revealed a significant impact on the primary metabolism of *Lolium-Festuca* grasses. The differences in the observed alterations in metabolite profiles were well visible between the analyzed introgression forms, mainly with respect to osmoprotectant and signaling molecules.

### Carbohydrates and Their Derivatives

Within a group of osmotic active compounds, soluble carbohydrates are especially important. Except their obvious central role at various levels of plant cell metabolism, these compounds are also associated with a wide range of response and signaling pathways altered by environmental stimuli (Hare et al., 1998; Fang and Xiong, 2015). Their small size, neutral character and biochemical compatibility allow them to maintain cellular water potential and interact with other cellular compound, stabilizing protein and membrane structures (Hoekstra et al., 2001; Xoconostle-Cazares et al., 2010). In the previous section, the accumulation profiles of the identified substrates and products of photosynthesis, were discussed. Here, in this paragraph, more potential functions of glucose, fructose, and sucrose in plant cell response to drought treatment, are considered. A higher accumulation level of glucose noticed in the introgression form 7/6 on the 11th day of drought period could also result in a higher ROS production in this form (Russell et al., 2002), and its higher exposure to potential oxidative damage, reducing the level of regeneration. However, this research aspect requires further work. An increased accumulation level of sucrose, fructose, and glucose was previously observed in forage grasses, such as *F. arundinacea* (Zwicke et al., 2015) and *L. perenne* (Foito et al., 2009), and in other plant species, such as *Oryza sativa* (rice; Ambavaram et al., 2014), *Triticum aestivum* (wheat; Izanloo et al., 2008), and *S. tuberosum* (Yang et al., 2015) under dehydration conditions, indicating that the mechanism of water deficit tolerance could be associated with a higher accumulation of soluble carbohydrates (Izanloo et al., 2008; Ambavaram et al., 2014). On the other hand, the research performed on *L. perenne* under drought treatment showed that sucrose, glucose, and fructose accumulation levels did not change significantly in leaves under stress treatment (Amiard et al., 2003). Glucose and sucrose are among the final products of carbon assimilation in plants, as well as precursors of the majority of organic compounds used in metabolic pathways (Ap Rees, 1995). Changing environmental conditions, including stress events such as water deficit often promote the accumulation of sucrose and products of its metabolism in the vegetative tissues. These compounds can serve as osmoprotectants (Kido et al., 2013), replacing dissipating water, especially under desiccation conditions (Hoekstra et al., 2001). The sucrose accumulation level is a result of sucrose biosynthesis or degradation of sugar polymers, such as starch. Both of these processes can be stimulated by drought (Zwicke et al., 2015). Overall, it was proved that sucrose could be associated with a dehydration avoidance mechanism, and can



replace the functions of other osmoprotectants, such as trehalose or raffinose (Foito et al., 2009), which were not identified here.

Other well-recognized osmoprotectants revealed here involved gentiobiose, melibiose, and glycerol (Bartels and Sunkar, 2005), which presented genotype dependent differences in accumulation levels under drought conditions. The important metabolites identified here are also *myo*-inositol and its related compounds, which can be used as substrates for the synthesis other osmoprotectants, such as raffinose family oligosaccharides (Karner et al., 2004), galactinol (ElSayed et al., 2014) or d-ononitol (Sheveleva et al., 1997), and a wide spectrum of lipid signaling compounds, such as phosphatidylinositol, phosphatidylinositol-phosphate, polyphosphoinositides, *myo*-inositol phosphate or sphingolipid related molecules - crucial components of various metabolic pathways involved in a control of gene expression, hormonal regulation and response reactions to stress conditions (Liu et al., 2013; Zhai et al., 2015). *Myo*-inositol belongs to a sugar alcohol group of metabolites in which hydroxyl group can substitute the hydroxyl group of water during interaction with membrane lipids and proteins, maintaining their structure and properties (ElSayed et al., 2014). The increased accumulation of *myo*-inositol and *myo*-inositol-1-phosphate in the 4/10 form was positively correlated with a higher accumulation level of phosphatidylinositol in this form during drought treatment, as described earlier by Perlikowski et al. (2016). It was shown earlier that the overexpression of *myo*-inositol-1-phosphate synthase, which is a key enzyme restricting amount of produced *myo*-inositol, increased the level of tolerance to osmotic stresses in plants, such as rice, tobacco, and potato (Yang et al., 2008; Goswami et al., 2014; Zhai et al., 2015). Also, in the other experiments, in drought and salt treated tobacco *myo*-inositol was accumulated in higher amounts (Sheveleva et al., 1997). In this work, an increased abundance of *myo*-inositol did not perfectly reflect the accumulation level of its potential product – galactinol under drought conditions. On the other hand, the increase of *myo*-inositol and galactinol accumulation levels was not observed earlier in *L. perenne* under drought conditions (Amiard et al., 2003). The accumulated carbohydrates can above all serve as carbon storage for recovery period after stress cessation (Hare et al., 1998). A higher accumulation level of *myo*-inositol, *myo*-inositol-1-phosphate and melibiose after 11 days of drought in the form 4/10 could positively influence a faster recovery of this form during rehydration.

Shikimic acid is a key compound used in biosynthetic pathway of aromatic amino acid production, such as phenylalanine, tyrosine, and tryptophan (Kojima et al., 2015). Shikimic acid was previously observed to accumulate under drought in potato leaflets (Yang et al., 2015) and *L. perenne* leaves (Foito et al., 2009). An accumulation of known antioxidants, such as threonic acid and dehydroascorbic acid (Debolt et al., 2007) in the form 4/10 under drought and further recovery might be associated with more efficient regeneration mechanism in this form, associated with more efficient scavenging of ROS. In the form 7/6, those compounds were not accumulated highly during drought treatment, compared to the control

conditions or even their accumulation levels decreased in drought.

## Amino Acids

Amino acids represent a highly significant group of metabolites, and it was demonstrated that their accumulation levels have a significant impact on the expression of plant drought tolerance (Barchet et al., 2014). It was previously observed that a particular amino acid content increased under drought conditions in plants. In creosotebush (*Larrea divaricata*) a significant increase of alanine, arginine, histidine, isoleucine, valine, glutamic acid, phenylalanine, and proline abundance was observed (Saunier et al., 1968). Also potato plants exposed to drought treatment accumulated higher amounts of glutamine/glutamic acid, serine, threonine, proline, phenylalanine, isoleucine, leucine, and valine (Yang et al., 2015). The previously observed higher accumulation level of asparagine in *O. sativa* cultivars under drought conditions was found to be negatively correlated with drought tolerance, and was characteristic for susceptible genotypes with a lower water use efficiency (Degenkolbe et al., 2013). The amino acid accumulation in plants exposed to dehydration conditions could be due as well to a protein hydrolysis (Saunier et al., 1968), and may be also associated with nitrogen storage for further metabolism remobilization during recovery period after stress cessation (Sicher and Barnaby, 2012).

One of the most common reactions of many plants upon experiencing dehydration conditions, is an accumulation of proline and glycine betaine (Bandurska and Jóźwiak, 2010). Proline, with its neutral character, does not negatively affect the cell environment and could be accumulated in higher amounts, since it functions as an osmotic adjustment molecule and protects other cellular compounds, such as proteins and lipids against denaturation and peroxidation (Bandurska, 2001). Proline has a high water potential, and its hydrophobic site can interact with hydrophobic parts of other proteins, while a hydrophilic site with high affinity to water allows it to maintain a high water potential, solubility and a native structure of proteins under dehydration conditions (Hoekstra et al., 2001). It was previously reported that related genotypes of *F. arundinacea* (Abernethy and McManus, 1998), *F. rubra* and *L. perenne* (Bandurska and Jóźwiak, 2010; Salehi et al., 2014) accumulated higher levels of proline and glycine betaine during a water deficit period, compared to well-watered plants. On the other hand, it was proposed earlier that proline had a lower contribution to the osmotic adjustment under drought stress in forage grasses, compared to soluble carbohydrates (Barker et al., 1993). Glycine betaine accumulated in plants under water deficit conditions, functioning as an osmoprotector and stabilizing cell compounds structure and activity (Sakamoto and Murata, 2002). It was previously reported that a progressive drought treatment induced an accumulation of glycine betaine in *Hordeum vulgare* (barley; Zuniga et al., 1989) and *F. arundinacea* (Abernethy and McManus, 1998). Here, glycine betaine was not identified, although one of its relatives, a precursor dimethylglycine-like was found among the analyzed metabolites. Probably, a higher accumulation level of proline and

dimethylglycine-like under the advanced drought conditions (the 11th day of drought) in the form 4/10 could positively affect a regeneration capacity of this form during recovery period, after rehydration.

## Primary Metabolite Accumulation Profiles and Plant Regeneration after Stress Cessation

A recovery after cessation of water deficit conditions refers to a plant's ability for re-growth, and to produce a fresh biomass during tissue regeneration following severe turgor loss and dehydration (Luo, 2010). In previous work, it was shown that the form 4/10, after 10 days of rehydration, was characterized by a significantly decreased level of electrolyte leakage, compared to its level after 11 days of drought, whereas in the 7/6, an elevated level of electrolyte leakage was also observed after a rehydration period. This indicated that the form 4/10 had more efficient regeneration mechanisms after stress cessation (Perlikowski et al., 2014). During recovery, soluble compounds stabilizing membrane surface are replaced by water in rehydrating membranes allowing them to slowly regain their function, and preventing any quick rupture of membranes caused by water flow (Hoekstra et al., 2001). It has been noticed that the accumulation of osmotically active compounds during a drought period was not fully reversible after the stress cessation, and this was associated with a stress memory. The rate of possible return of these compounds to the control values depended mainly on the strength of earlier stress treatment and the level of plant stress tolerance (Morgan, 1984). Thus, we assume that observed elevated levels, after rehydration, of most crucial compounds of photosynthetic pathway and osmoprotectants in the 4/10 form could be, at least partially, associated also with a stress memory, and with a stronger physiological performance of this form under recovery conditions, compared to the form 7/6. Sucrose decreased its abundance in both introgression forms after rehydration, and this could be associated with metabolic demands to restore normal cell activity and growth after stress cessation (Zwicke et al., 2015). Osmotic adjustment is considered as one of the most important mechanisms of plant tolerance to water deficit conditions, and also plays a crucial role in a plant recovery after stress cessation (Morgan, 1984). The impact of drought stress on plant development depends mainly on the stress intensity and duration, but also on genotype specific traits and earlier plant pre-hardening in stressful conditions. Although all the mechanisms driving a stress memory are still poorly understood, the evidence exists that the accumulation of some signaling molecules during drought conditions could be, at least partially, associated with this phenomenon (Bruce et al., 2007).

## CONCLUSIONS

The results obtained in this study clearly indicate that accumulation and activity levels of pFBA can influence the capacity of photosynthesis in the *L. multiflorum*/*F. arundinacea* introgression forms, due to the efficiency of the Calvin cycle. The

phenomenon, described in our paper, is the first example of non-stomatal mechanisms involved in a regulation of photosynthetic rate during prolonged drought treatment in the *Lolium-Festuca* forage grasses. The higher activity and accumulation levels of pFBA in the 7/6 form was accompanied by a higher accumulation level of most photosynthesis products in this form, during control and drought periods. On the other hand, the form 4/10 demonstrated higher accumulation levels of stress tolerance marker metabolites, such as proline, melibiose, galactaric acid, *myo*-inositol and *myo*-inositol-1-phosphate under drought conditions. Their accumulation could be associated with more efficient capacity of membrane regeneration in this form, after stress cessation. Furthermore, during rehydration, the 4/10 introgression form also maintained elevated levels of most metabolites, analyzed herein. It cannot be excluded though that these alterations to metabolism could be involved in a stress memory mechanism in forage grasses, however, this research aspect requires further work. The most important conclusions of this study are presented also graphically in a model figure (Figure 6).

## AUTHOR CONTRIBUTIONS

DP and AK designed the experiments. DP, MC, LM, and AA conducted the experimental work. DP and AK drew main conclusions. DP carried out the statistical analysis. DP and AK prepared the first version of the manuscript, but all the authors contributed in further writing, and finally read and approved the manuscript.

## FUNDING

The plant selection and physiological analysis were performed within projects funded by the Polish Ministry of Agriculture and Rural Development (no. 84; 2011 and 2012).

## ACKNOWLEDGMENT

We thank Prof. Neil Jones from Aberystwyth University for a critical reading of this manuscript.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.01063>

**FIGURE S1 |** The alignment of chloroplast fructose biphosphate aldolase (pFBA) cDNA sequences with 3' and 5' UTR of 4/10 and 7/6 *Lolium multiflorum*/*Festuca arundinacea* introgression forms with pFBA cDNA sequence of *Brachypodium distachyon* (B.d.; XM\_003577737.1). Gray highlighted text indicates 'start' and 'stop' positions of the coding sequence.

**FIGURE S2 |** The alignment of chloroplast fructose biphosphate aldolase (pFBA) predicted protein sequences of 4/10 and 7/6 *L. multiflorum*/*F. arundinacea* introgression forms with pFBA protein sequence of *Brachypodium distachyon* (B.d.; XM\_003577737.1). Black highlighted text indicates a predicted chloroplast transition sequence.

**TABLE S1 | The normalized mass spectral intensities for three biological replicates, data base annotations and ANOVA results of the analyzed metabolites.** Calculations were performed for five time-points of experiment: before drought, after three, six and 11 days of drought, and after ten days of re-watering in *Lolium multiflorum*/*Festuca arundinacea* introgression forms, based on [Log<sub>2</sub>] transformed normalized mass spectra peak intensities.

**TABLE S2 | Loadings of all the PCA components for the analyzed metabolites.** Calculations were performed for five time-points of experiment: before drought, after three, six and 11 days of drought, and after ten days of re-watering in *Lolium multiflorum*/*Festuca arundinacea* introgression forms, based on [Log<sub>2</sub>] transformed normalized mass spectra peak intensities.

## REFERENCES

- Abernethy, G. A., and McManus, M. T. (1998). Biochemical responses to an imposed water deficit in mature leaf tissue of *Festuca arundinacea*. *Environ. Exper. Bot.* 40, 17–28. doi: 10.1016/S0098-8472(98)00017-3
- Ambavaram, M. M., Basu, S., Krishnan, A., Ramegowda, V., Batlang, U., Rahman, L., et al. (2014). Coordinated regulation of photosynthesis in rice increases yield and tolerance to environmental stress. *Nat. Commun.* 5:5302. doi: 10.1038/ncomms6302
- Amiard, V., Morvan-Bertrand, A., Billard, J. P., Huault, C., Keller, F., and Prud'homme, M. P. (2003). Fructans, but not the sucrosyl-galactosides, raffinose and loliose, are affected by drought stress in perennial ryegrass. *Plant Physiol.* 132, 2218–2229. doi: 10.1104/pp.103.022335
- Ap Rees, T. (1995). "What next?" in *Sucrose Metabolism, Biochemistry, Physiology and Molecular Biology*, eds G. L. Salerno and E. J. Echeverria (Rockville, MD: American Society of Plant Physiologists), 278–283.
- Bandurska, H. (2001). Does proline accumulated in leaves of water deficit stressed barley plants confine cell membrane injuries? II. Proline accumulation during hardening and its involvement in reducing membrane injuries in leaves subjected to severe osmotic stress. *Acta Physiol. Plant.* 23, 483–490. doi: 10.1007/s11738-001-0059-0
- Bandurska, H., and Józwiak, W. (2010). A comparison of the effects of drought on proline accumulation and peroxidases activity in leaves of *Festuca rubra* L. and *Lolium perenne* L. *Acta Soc. Bot. Pol.* 79, 111–116. doi: 10.5586/asbp.2010.015
- Barchet, G. L., Dauwe, R., Guy, R. D., Schroeder, W. R., Soolanayakanahally, R. Y., Campbell, M. M., et al. (2014). Investigating the drought-stress response of hybrid poplar genotypes by metabolite profiling. *Tree Physiol.* 34, 1203–1219. doi: 10.1093/treephys/tpt080
- Barker, D. J., Sullivan, C. Y., and Moser, L. E. (1993). Water deficit effects on osmotic potential, cell wall elasticity, and proline in five forage grasses. *Agron. J.* 85, 270–275. doi: 10.2134/agronj1993.00021962008500020020x
- Bartels, D., and Sunkar, R. (2005). Drought and salt tolerance in plants. *Crit. Rev. Plant Sci.* 24, 23–58. doi: 10.1080/07352680590910410
- Bray, E. A., Bailey-Serres, J., and Weretilnyk, E. (2000). "Responses to abiotic stresses," in *Biochemistry and Molecular Biology of Plants*, eds W. Gruissem, B. Buchanan, and R. Jones (Rockville, MD: American Society of Plant Physiologists), 1158–1249.
- Bruce, T. J. A., Matthes, M. C., Napier, J. A., and Pickett, J. A. (2007). Stressful "memories" of plants: evidence and possible mechanisms. *Plant Sci.* 173, 603–608. doi: 10.1016/j.plantsci.2007.09.002
- Cairns, A. (2002). Starch accumulation in temperate forage grasses. *IGER Innovations* 6, 6–9.
- Chandler, P. M., and Robertson, M. (1994). Gene expression regulated by abscisic acid and its relation to stress tolerance. *Annu. Rev. Plant Physiol.* 45, 113–141. doi: 10.1146/annurev.pp.45.060194.000553
- Chaves, M. M., Maroco, J. P., and Pereira, J. S. (2003). Understanding plant responses to drought – from genes to the whole plant. *Funct. Plant Biol.* 30, 239–264. doi: 10.1071/FP02076
- Chaves, M. M., Zarrouk, O., Francisco, R., Costa, J. M., Santos, T., Regalado, A. P., et al. (2010). Grapevine under deficit irrigation: hints from physiological and molecular data. *Ann. Bot.* 105, 661–676. doi: 10.1093/aob/mcq030
- Cornic, G. (2000). Drought stress inhibits photosynthesis by decreasing stomatal aperture—not by affecting ATP synthesis. *Trends Plant Sci.* 5, 187–188. doi: 10.1016/S1360-1385(00)01625-3
- Debolt, S., Melino, V., and Ford, C. M. (2007). Ascorbate as a biosynthetic precursor in plants. *Ann. Bot.* 99, 3–8. doi: 10.1093/aob/mcl236
- Degenkolbe, T., Do, P. T., Kopka, J., Zuther, E., Hinch, D. K., and Köhl, K. I. (2013). Identification of drought tolerance markers in a diverse population of rice cultivars by expression and metabolite profiling. *PLoS ONE* 8:e63637. doi: 10.1371/journal.pone.0063637
- Dolferus, R. (2014). To grow or not to grow: a stressful decision for plants. *Plant Sci.* 229, 247–261. doi: 10.1016/j.plantsci.2014.10.002
- ElSayed, A. I., Rafudeen, M. S., and Golldack, D. (2014). Physiological aspects of raffinose family oligosaccharides in plants: protection against abiotic stress. *Plant Biol. (Stuttg)* 16, 1–8. doi: 10.1111/plb.12053
- Fang, Y., and Xiong, L. (2015). General mechanisms of drought response and their application in drought resistance improvement in plants. *Cell Mol. Life. Sci.* 72, 673–689. doi: 10.1007/s00018-014-1767-0
- Flexas, J., Bota, J., Loreto, F., Cornic, G., and Sharkey, T. D. (2004). Diffusive and metabolic limitations to photosynthesis under drought and salinity in C(3) plants. *Plant Biol. (Stuttg)* 6, 269–279. doi: 10.1055/s-2004-820867
- Flexas, J., and Medrano, H. (2002). Drought-inhibition of photosynthesis in C3 plants: stomatal and non-stomatal limitations revisited. *Ann. Bot.* 89, 183–189. doi: 10.1093/aob/mcf027
- Flexas, J., Ribas-Carbó, M., Bota, J., Galmés, J., Henkle, M., Martínez-Cañellas, S., et al. (2006). Decreased Rubisco activity during water stress is not induced by decreased relative water content but related to conditions of low stomatal conductance and chloroplast CO<sub>2</sub> concentration. *New Phytol.* 172, 73–82. doi: 10.1111/j.1469-8137.2006.01794.x
- Foito, A., Byrne, S. L., Shepherd, T., Stewart, D., and Barth, S. (2009). Transcriptional and metabolic profiles of *Lolium perenne* L. genotypes in response to a PEG-induced water stress. *Plant Biotechnol. J.* 7, 719–732. doi: 10.1111/j.1467-7652.2009.00437.x
- Goswami, L., Sengupta, S., Mukherjee, S., Ray, S., Mukherjee, R., and Majumder, A. L. (2014). Targeted expression of L-myo-inositol 1-phosphate synthase from *Porteresia coarctata* (Roxb.) Tateoka confers multiple stress tolerance in transgenic crop plants. *J. Plant Biochem. Biotechnol.* 23, 316–330. doi: 10.1007/s13562-013-0217-7
- Haake, V., Geiger, M., Walch-Liu, P., Of Engels, C., Zrenner, R., and Stitt, M. (1999). Changes in aldolase activity in wild-type potato plants are important for acclimation to growth irradiance and carbon dioxide concentration, because plastid aldolase exerts control over the ambient rate of photosynthesis across a range of growth conditions. *Plant J.* 17, 479–489. doi: 10.1046/j.1365-313X.1999.00391.x
- Haake, V., Zrenner, R., Sonnewald, U., and Stitt, M. (1998). A moderate decrease of plastid aldolase activity inhibits photosynthesis, alters the levels of sugars and starch, and inhibits growth of potato plants. *Plant J.* 14, 147–157. doi: 10.1046/j.1365-313X.1998.00089.x
- Hare, P. D., Cress, W. A., and Van Staden, J. (1998). Dissecting the roles of osmolyte accumulation during stress. *Plant Cell Environ.* 21, 535–553. doi: 10.1046/j.1365-3040.1998.00309.x
- Harrison, E. P., Willingham, N. M., Lloyd, J. C., and Raines, C. A. (1998). Reduced sedoheptulose-1,7-bisphosphatase levels in transgenic tobacco lead to decreased photosynthetic capacity and altered carbohydrate accumulation. *Planta* 204, 27–36. doi: 10.1007/s004250050226
- Hoekstra, F. A., Golovina, E. A., and Buitink, J. (2001). Mechanisms of plant desiccation tolerance. *Trends Plant Sci.* 6, 431–438. doi: 10.1016/S1360-1385(01)02052-0
- Izanloo, A., Condon, A. G., Langridge, P., Tester, M., and Schnurbusch, T. (2008). Different mechanisms of adaptation to cyclic water stress in two South Australian bread wheat cultivars. *J. Exp. Bot.* 59, 3327–3346. doi: 10.1093/jxb/ern199
- Karner, U., Peterbauer, T., Raboy, V., Jones, D. A., Hedley, C. L., and Richter, A. (2004). Myo-Inositol and sucrose concentrations affect the accumulation of raffinose family oligosaccharides in seeds. *J. Exp. Bot.* 55, 1981–1987. doi: 10.1093/jxb/erh216
- Kido, E. A., Ferreira Neto, J. R., Silva, R. L., Belarmino, L. C., Bezerra Neto, J. P., Soares-Cavalcanti, N. M., et al. (2013). Expression dynamics and genome distribution of osmoprotectants in soybean: identifying important components



- to face abiotic stress. *BMC Bioinformatics* 14(Suppl. 1):S7. doi: 10.1186/1471-2105-14-S1-S7
- Kojima, M., Kimura, N., and Miura, R. (2015). Regulation of primary metabolic pathways in oyster mushroom mycelia induced by blue light stimulation: accumulation of shikimic acid. *Sci. Rep.* 5:8630. doi: 10.1038/srep08630
- Kosmala, A., Perlikowski, D., Pawłowicz, I., and Rapacz, M. (2012). Changes in the chloroplast proteome following water deficit and subsequent watering in a high and a low drought tolerant genotype of *Festuca arundinacea*. *J. Exp. Bot.* 63, 6161–6172. doi: 10.1093/jxb/ers265
- Kossmann, J., Sonnewald, U., and Willmitzer, L. (1995). Reduction of the chloroplast fructose-1,6-bisphosphatase in transgenic potato plants impairs photosynthesis and plant growth. *Plant J.* 6, 637–650. doi: 10.1046/j.1365-313X.1994.6050637.x
- Lawlor, D. W. (2002). Limitation to photosynthesis in water-stressed leaves: stomata vs. metabolism and the role of ATP. *Ann. Bot.* 89, 871–885. doi: 10.1093/aob/mcf110
- Lawlor, D. W., and Cornic, G. (2002). Photosynthetic carbon assimilation and associated metabolism in relation to water deficits in higher plants. *Plant Cell Environ.* 25, 275–294. doi: 10.1046/j.0016-8025.2001.00814.x
- Liu, X., Zhai, S., Zhao, Y., Sun, B., Liu, C., Yang, A., et al. (2013). Overexpression of the phosphatidylinositol synthase gene (ZmPIS) conferring drought stress tolerance by altering membrane lipid composition and increasing ABA synthesis in maize. *Plant Cell Environ.* 36, 1037–1055. doi: 10.1111/pce.12040
- Luo, L. J. (2010). Breeding for water-saving and drought-resistance rice (WDR) in China. *J. Exp. Bot.* 61, 3509–3517. doi: 10.1093/jxb/erq185
- Morgan, J. M. (1984). Osmoregulation and water stress in higher plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 35, 299–319. doi: 10.1146/annurev.pp.35.060184.001503
- Munns, R. (2002). Comparative physiology of salt and water stress. *Plant Cell Environ.* 25, 239–250. doi: 10.1046/j.0016-8025.2001.00808.x
- Pawłowicz, I., Kosmala, A., and Rapacz, M. (2012). Expression pattern of the psbO gene and its involvement in acclimation of the photosynthetic apparatus during abiotic stresses in *Festuca arundinacea* and *F. pratensis*. *Acta Physiol. Plant* 34, 1915–1924. doi: 10.1007/s11738-012-0992-0
- Perlikowski, D., Kiersznowska S., Sawikowska, A., Krajewski, P., Rapacz, M., Eckhardt, A., et al. (2016). Remodeling of leaf cellular glycerolipid composition under drought and re-hydration conditions in grasses from the *Lolium-Festuca* complex. *Front. Plant Sci.* 7:1027. doi: 10.3389/fpls.2016.01027
- Perlikowski, D., Kosmala, A., Rapacz, M., Kościelniak, J., Pawłowicz, I., and Zwierzykowski, Z. (2014). Influence of short-term drought conditions and subsequent re-watering on the physiology and proteome of *Lolium multiflorum*/*Festuca arundinacea* introgression forms with contrasting levels of tolerance to long-term drought. *Plant Biol.* 16, 385–394. doi: 10.1111/plb.12074
- Pollock, C. J., and Cairns, A. J. (1991). Fructan metabolism in grasses and cereals. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 42, 77–101. doi: 10.1146/annurev.pp.42.060191.000453
- Raines, C. A. (2003). The Calvin cycle revisited. *Photosynth. Res.* 75, 1–10. doi: 10.1023/A:1022421515027
- Raines, C. A., and Lloyd, J. C. (2007). C3 carbon reduction cycle. *Encycl. Life Sci.* doi: 10.1002/9780470015902.a0001314.pub2
- Rontein, D., Basset, G., and Hanson, A. D. (2002). Metabolic engineering of osmoprotectant accumulation in plants. *Metab. Eng.* 4, 49–56. doi: 10.1006/mben.2001.0208
- Russell, J. W., Golovoy, D., Vincent, A. M., Mahendru, P., Olzmann, J. A., Mentzer, A., et al. (2002). High glucose-induced oxidative stress and mitochondrial dysfunction in neurons. *FASEB J.* 16, 1738–1748. doi: 10.1096/fj.01-1027com
- Sakamoto, A., and Murata, N. (2002). The role of glycine betaine in the protection of plants from stress: clues from transgenic plants. *Plant Cell Environ.* 25, 163–171. doi: 10.1046/j.0016-8025.2001.00790.x
- Salehi, M., Salehi, H., xNiazi, H., and Ghobadi, C. (2014). Convergence of goals: phylogenetical, morphological, and physiological characterization of tolerance to drought stress in tall fescue (*Festuca arundinacea* Schreb.). *Mol. Biotechnol.* 56, 248–257. doi: 10.1007/s12033-013-9703-3
- Saunier, R. E., Hull, H. M., and Ehrenreich, J. H. (1968). Aspects of the drought tolerance in creosotebush (*Larrea divaricata*). *Plant Physiol.* 43, 401–404. doi: 10.1104/pp.43.3.401
- Sheveleva, E., Chmara, W., Bohnert, H. J., and Jensen, R. G. (1997). Increased salt and drought tolerance by d-ononitol production in transgenic *Nicotiana tabacum* L. *Plant Physiol.* 115, 1211–1219.
- Sibley, J. A., and Lehninger, A. L. (1949). Determination of aldolase in animal tissues. *J. Biol. Chem.* 177, 859–872.
- Sicher, R. C., and Barnaby, J. Y. (2012). Impact of carbon dioxide enrichment on the responses of maize leaf transcripts and metabolites to water stress. *Physiol. Plant* 144, 238–253. doi: 10.1111/j.1399-3054.2011.01555.x
- Signarbieux, C., and Feller, U. (2011). Non-stomatal limitations of photosynthesis in grassland species under artificial drought in the field. *Environ. Exp. Bot.* 71, 192–197. doi: 10.1016/j.envexpbot.2010.12.003
- Uematsu, K., Suzuki, N., Iwamae, T., Inui, M., and Yukawa, H. (2012). Increased fructose 1,6-bisphosphate aldolase in plastids enhances growth and photosynthesis of tobacco plants. *J. Exp. Bot.* 63, 3001–3009. doi: 10.1093/jxb/ers004
- Willard, J. M., and Gibbs, M. (1968). Role of aldolase in photosynthesis. 2. demonstration of aldolase types in photosynthetic organisms. *Plant Physiol.* 43, 793–798. doi: 10.1104/pp.43.5.793
- Wojakowska, A., Marczak, Ł., Jelonek, K., Polanski, K., Widlak, P., and Pietrowska, M. (2015). An optimized method of metabolite extraction from formalin-fixed paraffin-embedded tissue for GC/MS analysis. *PLoS ONE* 10:e0136902. doi: 10.1371/journal.pone.0136902
- Xoconostle-Cazares, B., Ramirez-Ortega, F. A., Flores-Elenes, L., and Ruiz-Medrano, R. (2010). Drought Tolerance in Crop Plants. *Am. J. Plant Physiol.* 5, 241–256. doi: 10.3923/ajpp.2010.241.256
- Yang, J., Fleisher, D. H., Sicher, R. C., Kim, J., Baligar, V. C., and Reddy, V. R. (2015). Effects of CO<sub>2</sub> enrichment and drought pretreatment on metabolite responses to water stress and subsequent rehydration using potato tubers from plants grown in sunlit chambers. *J. Plant Physiol.* 189, 126–136. doi: 10.1016/j.jplph.2015.10.004
- Yang, L., Tang, R., Zhu, J., Liu, H., Mueller-Roeber, B., Xia, H., et al. (2008). Enhancement of stress tolerance in transgenic tobacco plants constitutively expressing AtPlk2b, an inositol polyphosphate 6-/3-kinase from *Arabidopsis thaliana*. *Plant Mol. Biol.* 66, 329–343. doi: 10.1007/s11103-007-9267-3
- Zhai, H., Wang, F., Si, Z., Huo, J., Xing, L., An, Y., et al. (2015). A myo-inositol-1-phosphate synthase gene, IbMIPS1, enhances salt and drought tolerance and stem nematode resistance in transgenic sweet potato. *Plant Biotechnol. J.* 14, 592–602. doi: 10.1111/pbi.12402
- Zhang, X., Wu, N., and Li, C. (2005). Physiological and growth responses of *Populus davidiana* ecotypes to different soil water contents. *J. Arid Environ.* 60, 567–579. doi: 10.1016/j.jaridenv.2004.07.008
- Zuniga, G. E., Argandona, V. H., and Corcuera, L. K. J. (1989). Distribution of glycinebetaine and proline in water stressed and unstressed barley leaves. *Phytochemistry* 28, 419–420. doi: 10.1016/0031-9422(89)80024-X
- Zwicke, M., Picon-Cochard, C., Morvan-Bertrand, A., Prud'homme, M. P., and Voltaire, F. (2015). What functional strategies drive drought survival and recovery of perennial species from upland grassland? *Ann. Bot.* 116, 1001–1015. doi: 10.1093/aob/mcv037

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Perlikowski, Czyżniewski, Marczak, Augustyniak and Kosmala. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Transcriptome Profiling of Buffalograss Challenged with the Leaf Spot Pathogen *Curvularia inaequalis*

Bimal S. Amaradasa<sup>1</sup> and Keenan Amundsen<sup>2\*</sup>

<sup>1</sup> Department of Plant Pathology, University of Nebraska–Lincoln, Lincoln, NE, USA, <sup>2</sup> Department of Agronomy and Horticulture, University of Nebraska–Lincoln, Lincoln, NE, USA

## OPEN ACCESS

### Edited by:

Sergio Lanteri,  
University of Turin, Italy

### Reviewed by:

Abu Hena Mostafa Kamal,  
University of Texas at Arlington, USA  
Julio Vega-Arreguin,  
National Autonomous University  
of Mexico, Mexico

### \*Correspondence:

Keenan Amundsen  
kamundsen2@unl.edu

### Specialty section:

This article was submitted to  
Crop Science and Horticulture,  
a section of the journal  
Frontiers in Plant Science

Received: 29 December 2015

Accepted: 09 May 2016

Published: 25 May 2016

### Citation:

Amaradasa BS and Amundsen K  
(2016) Transcriptome Profiling  
of Buffalograss Challenged with  
the Leaf Spot Pathogen *Curvularia*  
*inaequalis*. *Front. Plant Sci.* 7:715.  
doi: 10.3389/fpls.2016.00715

Buffalograss (*Bouteloua dactyloides*) is a low maintenance U. S. native turfgrass species with exceptional drought, heat, and cold tolerance. Leaf spot caused by *Curvularia inaequalis* negatively impacts buffalograss visual quality. Two leaf spot susceptible and two resistant buffalograss lines were challenged with *C. inaequalis*. Samples were collected from treated and untreated leaves when susceptible lines showed symptoms. Transcriptome sequencing was done and differentially expressed genes were identified. Approximately 27 million raw sequencing reads were produced per sample. More than 86% of the sequencing reads mapped to an existing buffalograss reference transcriptome. *De novo* assembly of unmapped reads was merged with the existing reference to produce a more complete transcriptome. There were 461 differentially expressed transcripts between the resistant and susceptible lines when challenged with the pathogen and 1552 in its absence. Previously characterized defense-related genes were identified among the differentially expressed transcripts. Twenty one resistant line transcripts were similar to genes regulating pattern triggered immunity and 20 transcripts were similar to genes regulating effector triggered immunity. There were also nine up-regulated transcripts in resistance lines which showed potential to initiate systemic acquired resistance (SAR) and three transcripts encoding pathogenesis-related proteins which are downstream products of SAR. This is the first study characterizing changes in the buffalograss transcriptome when challenged with *C. inaequalis*.

**Keywords:** buffalograss (*Bouteloua dactyloides*), *Curvularia inaequalis*, leaf spot, defense-related genes, transcriptome, next-generation sequencing

## INTRODUCTION

Buffalograss (*Bouteloua dactyloides*) is a U. S. native, warm-season turfgrass species with exceptional drought, heat, and cold tolerance (Beetle, 1950; Reeder, 1971). Buffalograss requires less fertility, pesticides and water to maintain an acceptable quality level compared to traditional turfgrass species (Riordan et al., 1993). Replacing traditional turfgrass species with buffalograss may help to conserve water, especially in the semi-arid and arid regions of the USA (Riordan et al., 1993). Buffalograss is tolerant of many diseases, but leaf spot can cause decline or death of buffalograss turf. Leaf spot is caused by several species belonging to the *Curvularia*, *Bipolaris*,

and *Cercospora* genera (Smith et al., 1989; Smiley et al., 2005). In Nebraska, *Curvularia inaequalis* (Shear) Boedijn and *Bipolaris spicifera* (Bainier) Subram (teleomorph: *Cochliobolus spicifer* Nelson) are commonly isolated from buffalograss with leaf spot symptoms (Amaradasa and Amundsen, 2014b). On lawns, leaf spot initiates as dark brown leaf spots followed by leaf tip dieback and eventual blighting of entire tillers. As the disease progresses, patches of leaf decline and canopy thinning occur. Leaf spot symptoms of *C. inaequalis* and *B. spicifera* are identical and therefore it is not possible to distinguish the causal organism by disease symptoms alone. Disease development commonly occurs when temperatures are 30°C and above. Disease severity increases when buffalograss is under stress by adverse weather conditions such as high temperatures, high humidity, drought, excess rain, and cloud cover. Since buffalograss is often considered a low-maintenance turfgrass, the use of fungicides is usually not preferred by homeowners and lawn care managers. Incorporating host resistance through plant breeding is one way to combat leaf spot disease. Conventional breeding for disease resistance is difficult and time consuming, and is based on inoculation, rating for incidence and severity of disease, and selection of resistant genotypes. Identification of genes that confer leaf spot resistance would enable molecular-based strategies to improve the efficiency of breeding for resistant cultivars.

Today, comparative genetic studies using next generation sequencing (NGS) technology are common for characterizing gene functions in plants and other organisms. NGS technology can be used to sequence both genomic DNA and total RNA (RNA-seq). The large number of short sequencing reads produced by this technology is highly cost effective and can be used to assemble a genome or transcriptome *de novo* or can be mapped to a reference to determine differentially expressed genes. Buffalograss has a basic chromosome number of 10 and exists as a ploidy series of diploids, tetraploids, pentaploids, and hexaploids (Johnson et al., 1998). This large repetitive genome makes whole genome sequencing and annotation difficult. Conversely, transcriptome profiling by RNA-seq has been used to decipher differentially expressed genes in grass systems (Wang et al., 2009; Gutierrez-Gonzalez et al., 2013; Wachholtz et al., 2013; Yang et al., 2013). The number of short-reads from RNA-seq data gives an indication of the level of gene expression and therefore is highly suitable for gene expression studies (Wang et al., 2009).

To identify differentially expressed defense-related genes in buffalograss, we profiled transcriptomes of two leaf spot resistant and two susceptible lines after challenging with *C. inaequalis*. For this study, we chose to use *Curvularia* over *Bipolaris* since it is more virulent and produces disease symptoms faster. *De novo* assembly of RNA-seq data from a previous Prestige buffalograss NGS study resulted in a reference assembly of 91,519 contigs (Wachholtz et al., 2013); this previously published buffalograss transcriptome was used as a reference in our study. In the previous study, basal transcriptional expression differences were compared between the two buffalograss cultivars Prestige and 378. However, identification of defense-related genes in response to a pathogen was not part of the previous study. The main

objective of our study was to identify buffalograss leaf spot resistance genes differentially expressed between resistant and susceptible buffalograss.

## MATERIALS AND METHODS

### Buffalograss Inoculation and Leaf Tissue Sampling

Two leaf spot resistant (95-55 and NE-BFG-7-3459-17) and two susceptible buffalograss lines (Prestige and NE-BFG-7-3453-50) identified previously were used in our study (Amaradasa and Amundsen, 2014a). Stolons of leaf spot resistant lines and susceptible lines were planted in 7-cm-diameter plastic pots filled with Fafard® 3B Mix potting medium. Pots were kept in a greenhouse with a 16 h day and 8 h night photoperiod. The average daytime and nighttime temperature of the greenhouse was maintained at 30 and 22°C, respectively. Plants were watered daily, fertilized biweekly with 20–20–20 to provide an approximate annual rate of 10 g N m<sup>-2</sup>, and clipped with scissors regularly to a height of 6 to 7 cm to promote prostrate growth and pot coverage. After 12 weeks of growth, plants were arranged in a randomized complete block design (RCBD) with three replications. Single-spore *C. inaequalis* strain 4L-SS01 was used to prepare a spore culture of 1 × 10<sup>6</sup> spores ml<sup>-1</sup> according to published methods (Brecht et al., 2007). Each pot was sprayed with 15 ml of the spore solution. Untreated controls were also included with three replicates and sprayed with water in place of the spore solution. After 10 days, when susceptible lines were exhibiting distinct disease symptoms, leaf tissue from both inoculated and uninoculated pots was harvested into separate freezer bags and immediately frozen in liquid nitrogen. Samples were kept at –80°C for later use.

### Total RNA Sequencing and Analysis

Approximately 100 mg of leaf tissue of each sample (95-55, NE-BFG-7-3459-17, Prestige, and NE-BFG-7-3453-50) was homogenized in liquid nitrogen using a mortar and pestle and RNA was extracted using an RNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions. RNA samples were qualitatively analyzed by agarose gel electrophoresis, and quantified using a NanoDrop 2000C spectrophotometer (Thermo Fisher Scientific Inc., Wilmington, DE, USA). Total RNA from 24 samples [4 buffalograss lines × 3 replicates × 2 treatments (inoculated/uninoculated)] was sent to the High-Throughput DNA Sequencing and Genotyping Core Facility located at the University of Nebraska Medical Center, Omaha, Nebraska for transcriptome sequencing. The cDNA libraries were prepared and then sequenced using a HiSeq 2000 sequencing platform (Illumina, San Diego, CA, USA) according to the manufacturer's RNA-seq protocol. The 24 samples were separately barcoded and run on three lanes of the HighSeq 2000 to obtain 100 bp single-end reads. Quality filtering of the reads was done by the Genotyping Core Facility. FastQC<sup>1</sup> was used to visualize the quality of the reads using

<sup>1</sup><http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

default parameters. Since FastQC showed several overrepresented reads consisting of Illumina adapter and primer sequences, Trimmomatic-0.30 (Bolger et al., 2014) was used to remove those contaminants. The reads were trimmed to a uniform length of 80 bp prior to downstream analysis. A fastq file containing the sequencing reads and quality data was used for downstream analysis. The sequencing reads were mapped with Bowtie2-2.1.0 (Langmead and Salzberg, 2012) to the *B. dactyloides* cv. Prestige transcriptome (Wachholtz et al., 2013). Reads that did not map to the reference were retained and assembled using Trinity-r2013-02-25 (Grabherr et al., 2011). The Trinity assembled contigs and the Prestige reference transcriptome were merged and cd-hit-est version 4.5.4 (Weizhong and Godzik, 2006) was used to remove redundancy with a 100% identity threshold to create the buffalograss transcriptome. Single-end raw sequencing reads of each individual were mapped with Bowtie2 to the buffalograss transcriptome to allow for the estimation of transcript abundance per individual relative to the buffalograss transcriptome. A read count table was produced using SAMtools (Li et al., 2009) and Perl<sup>2</sup>. To account for the variability of total initial Illumina sequencing results among samples, mapped read counts were subjected to normalization and then analyzed for differential expression using the DESeq2 Bioconductor package (Love et al., 2014) in R program (version 3.0.2).

Read counts of the two inoculated resistant (R) lines were compared separately to each inoculated susceptible (S) line. We used a final adjusted *P*-value of < 0.01 to select transcripts that showed a difference in expression between inoculated resistant and susceptible lines. The differentially up-regulated transcripts of 95-55 (R) vs. NE-BFG-7-3453-50 (S) were compared with 95-55 (R) vs. Prestige (S) and common transcripts were identified. Similarly, common up-regulated transcripts between inoculated resistant line NE-BFG-7-3459-17 (R) vs. inoculated susceptible lines were selected. Then the two sets of up-regulated genes (i.e., 95-55 vs. susceptible lines and NE-BFG-7-3459-17 vs. susceptible lines) were compared to each other and common transcripts were identified for annotation. By this filtering procedure, we identified transcripts that are common and differentially up-regulated in both inoculated resistant lines compared to the susceptible lines (up-regulated in resistant inoculated; URI). In the same way, we compared each uninoculated resistant line with each uninoculated susceptible line and used the same filtering procedure to identify genes in common that have different levels of expression between both resistant and susceptible lines (basal up-regulated expressions; BUE). Then we identified down-regulated transcripts in resistant inoculated and uninoculated cultivars compared to inoculated and uninoculated susceptible lines, respectively (down-regulated in resistant inoculated: DRI; basal down-regulated expressions: BDE).

The four sets of transcripts were pooled and annotated with Blast2GO using default settings (Conesa et al., 2005). The annotated transcripts were analyzed separately to identify genes responsible for induced resistance (transcripts of URI and DRI) and innate immunity (transcripts of BUE and BDE) in buffalograss. Blast2GO was used to prepare graphs of biological,

cellular and molecular processes at level two gene ontologies. Gene ontology IDs resulting from the analysis were mined for disease resistance related terms.

## Validation of BUE and URI Gene Expression

Ten differentially expressed transcripts were selected for validation by reverse transcription PCR (RT-PCR). These transcripts showed up-regulation either in inoculated or uninoculated resistant buffalograss lines compared to inoculated or uninoculated susceptible lines, respectively. Some transcripts were chosen because they did not have any read counts for either inoculated or uninoculated susceptible lines (Supplementary Table S1). The primers (Table 1) were synthesized for each transcript using Primer3web<sup>3</sup> version 4.0.0. The primers for ubiquitin conjugating enzyme (*UCE*) were used as a positive control (Table 1). RNA was extracted from new plants of inoculated and uninoculated 95-55, NE-BFG-7-3459-17, Prestige, and NE-BFG-7-3453-50 as described previously. cDNA was synthesized using an Invitrogen<sup>TM</sup> SuperScript<sup>®</sup> III First-Strand Synthesis System (Life Technologies, Grand Island, NY, USA) according to the manufacturer's instructions. Using a standard PCR protocol (each 25 µl reaction mixture contained 1x Taq DNA polymerase buffer, 0.2 µM forward and reverse primers, 0.2 mM each dNTPs, 1 U Taq polymerase) each primer pair was used to amplify 100 ng of cDNA template in a Mastercycler<sup>®</sup>Pro thermocycler (Eppendorf, Hamburg, Germany) with the following conditions: Initial denaturation at 94°C for 3 min, was followed by 35 cycles at 94°C for 30 s, 55°C for 1 min, and 72°C for 1 min, and a final extension at 72°C for 10 min. Thereafter the reaction was stopped by reducing the temperature to 4°C and PCR products were stored at -20°C. Aliquots (5 µl) of amplified products were separated by electrophoresis on a gel containing 1.7% (w/v) agarose and 1x TAE, at 100 V for 1 h. The presence and size of the DNA fragments were verified by staining the gel with ethidium bromide and observing under UV light.

## RESULTS

### Filtering and De Novo Assembly of Raw Reads

Sequencing of twenty four cDNA libraries constructed from *C. inaequalis* infected as well as non-infected buffalograss lines produced approximated 655.2 million 100 bp single-end reads (Table 2). On average, 30.5 million reads were produced per sample from inoculated leaf tissue and 24.0 million reads were produced per non-inoculated leaf tissue sample. Trimmomatic removed 9 to 17% of overrepresented sequences (Table 2). The existing *B. dactyloides* cv. Prestige reference transcriptome (Wachholtz et al., 2013) had 91,519 contigs. When raw reads of each of the buffalograss samples were mapped to the Prestige reference, 2 to 14% of the reads did not map (Table 2).

<sup>2</sup><http://www.perl.org>

<sup>3</sup><http://bioinfo.ut.ee/primer3/>

The unmapped raw reads were assembled by Trinity into 156,721 contigs. These contigs were merged with the existing Prestige reference resulting in the creation of the buffalograss transcriptome which consisted of 248,221 transcripts. The final buffalograss reference used for read mapping was prepared by removing all isoforms and keeping only the longest transcript for each gene. This final buffalograss reference had 196,168 transcripts with the longest transcript having 13,148 bp, average

transcript length of 565 bp, and median transcript length of 340 bp. We submitted the sequences of final reference to NCBI BioProject repository<sup>4</sup> under the project ID PRJNA297834. The distribution of transcript lengths is depicted in **Figure 1**. When the sequencing reads from the 24 buffalograss samples were

<sup>4</sup><http://www.ncbi.nlm.nih.gov/bioproject/>

**TABLE 1 | Primers used for validating expression of selected buffalograss transcripts.**

Transcript name	Transcript length	Forward primer (5'–3')	Reverse primer (3'–5')	PCR product size
BG_UCE <sup>a</sup>		GACCCGCCTACATCATGCA	TTGCTTGCCAGTGAACATGTC	61
Bodac2015c080259	258	CCAGCAAGTTCGAGAGATCC	GGCACTGCCGGAATGTATAA	203
Bodac2015c126512	223	AAAATCAGGGGATCCCAAC	GGAAGTGCTTGGCCTAGGTA	176
Bodac2015c127986	226	AGCCTCTGCAAACCAAGAAA	CCGCAAGCTTAACCTTCATC	158
Bodac2015c166783	1483	AGCACTACTCTGCGTCAGCA	CGTCATTGCTATGGATGGTG	194
Bodac2015c163663	2499	AAACCACGGAAGCCTTTTCT	GCATGTGCATTTTGTGTTGG	489
Bodac2015c100000	563	GATAGGCGTGTGAAACAGGAA	CAGTGTGGACACCTTGTTG	423
Bodac2015c000129	1874	GCACAGGCCGTACAATACCT	CGACTGACGACCCGAATAAT	210
Bodac2015c064306	317	GACGCAGTGAGGATCAGTTG	GCCATGCGTTTCTACTCTCC	164
Bodac2015c084994	414	TTTGATGATCGGTGCATCTC	GCCTCCATCTCATCCTCTTG	249
Bodac2015c087283	501	AAAGGCAACTCATCGACCTC	CAACCTTGCGTTCATTGTC	216

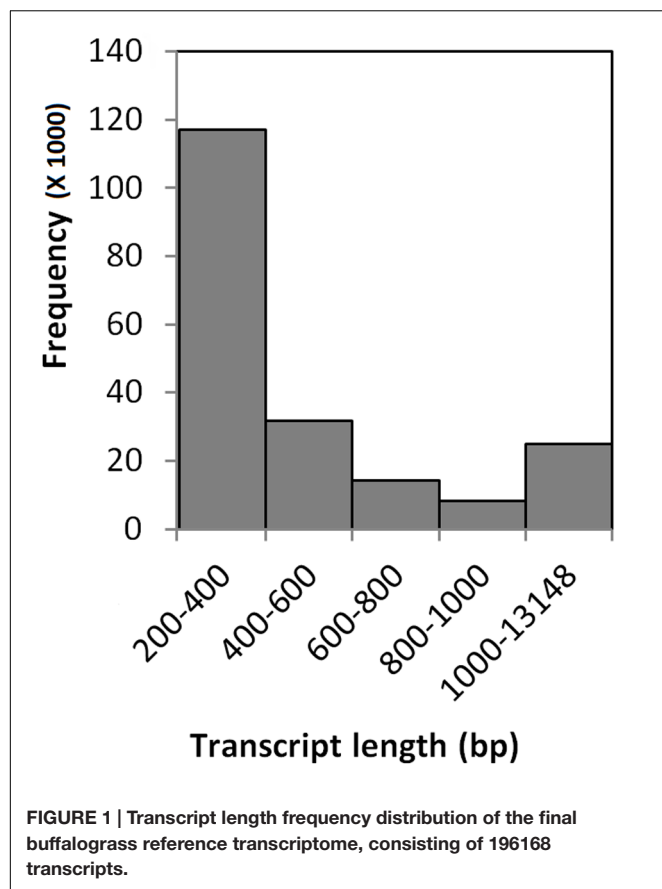
<sup>a</sup>Endogenous control ubiquitin conjugating enzyme gene.

**TABLE 2 | Summary of raw reads per sample, quality filtering, and mapping percentages of surviving reads.**

Buffalograss samples <sup>a</sup>	Total raw reads (millions)	Total reads after removing contaminants (millions)	Surviving reads mapped to Prestige reference <sup>b</sup> (%)	Surviving reads mapped to final buffalograss reference (%)
17C1	24.4	21.9	97.01	99.10
17C2	26.5	24.1	97.25	99.23
17C3	31.5	28.3	97.6	99.38
17T1	27.6	24.3	94.44	99.06
17T2	33.3	29.8	94	98.97
17T3	32.3	28.6	92.91	98.93
50C1	25.8	23.1	96.83	99.13
50C2	19.6	17.6	97.05	99.23
50C3	26.8	23.7	97.14	99.27
50T1	20.3	17.1	92.81	98.58
50T2	58.1	49.0	94.47	99.11
50T3	20.9	18.0	94.99	99.14
95C1	21.9	19.4	97.56	99.31
95C2	21.6	19.5	97.47	99.34
95C3	24.0	21.7	98.08	99.39
95T1	33.0	27.9	94.92	99.16
95T2	28.7	24.3	95.11	99.15
95T3	30.9	27.4	94.42	99.06
PC1	18.7	16.6	90.71	99.06
PC2	23.5	20.4	86.57	98.78
PC3	24.6	22.0	92.37	99.07
PT1	27.2	23.3	90.18	98.57
PT2	25.0	20.8	94.8	99.11
PT3	29.0	24.5	95.89	99.13

<sup>a</sup>17, 50, 95 and P refers to NE BG 7-3459-17, NE BG 7-3453-50, 95-55 and Prestige, respectively. C and T indicate untreated and treated samples. <sup>b</sup>Wachholtz et al. (2013).





mapped to the final buffalograss reference, more than 98% of the reads mapped at least once (Table 2).

## Identification of Differentially Expressed Transcripts

The DESeq2 analysis of inoculated lines resulted in 355 up-regulated transcripts (URI) and 106 down-regulated transcripts (DRI) in both resistant lines when compared to the susceptible lines (Figures 2A and 3A). Similarly, the uninoculated resistant lines had 1,076 transcripts with higher expression (BUE) and 476 transcripts with lower expression (BDE) in common when compared with the uninoculated susceptible lines (Figures 2B and 3B). There were 75 transcripts in common among the two up-regulated transcript sets (URI and BUE) and 14 common transcripts in the two down-regulated transcript sets (DRI and BDE). Eight URI transcripts had more than 10 mapped reads for each of the resistant samples and no mapped reads for the susceptible samples. Of these eight transcripts, five were in common with the BUE transcripts. Similarly, nine BUE transcripts had at least 10 mapped reads for each of the resistant samples and no mapped reads for the susceptible samples. Of these nine transcripts, two were in common with the URI transcripts. Two transcripts (Bodac2015c153835 and Bodac2015c154561) of DRI group had no mapped reads for any of the resistant samples while all of

the inoculated susceptible samples had more than 10 mapped reads.

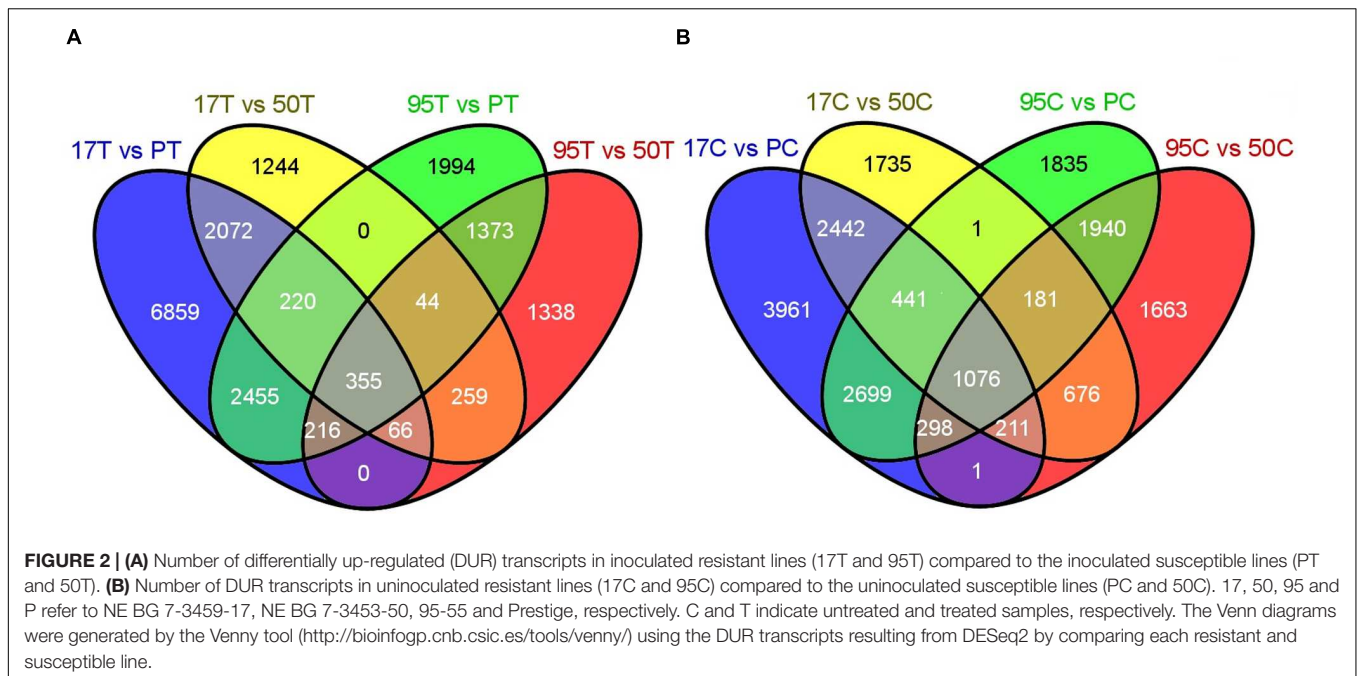
## Annotation of Differentially Expressed Transcripts

In total, there were 1,356 unique differentially expressed up-regulated transcripts (BUE and URI). These 1,356 transcripts were subjected to Blast2GO and 678 had blast hits resulting in the annotation of 528 transcripts while the other 828 were not annotated by Blast2GO. However, some of the 150 (678 blast hits - 528 annotated sequences) sequences that had blast hits but were not annotated did have assigned protein descriptions that were associated with disease resistance. Among the 528 annotated sequences, 381 sequences had biological process associated gene ontology (GO) terms, 425 sequences had molecular function associated GO terms, and 341 sequences had cellular component associated GO terms. The level 2 GO terms for biological process, molecular function, and cellular component are summarized in Figure 4. The most prevalent biological process GOs were cellular process (289 sequences), metabolic process (283 sequences), single-organism process (153 sequences), and response to stimulus (130 sequences). The largest molecular function gene ontologies are catalytic activity (293 sequences) and binding (278 sequences). Cell (301 sequences), organelle (257 sequences), and membrane (166 sequences) are the three largest cellular component gene ontologies.

Additionally, there were a total of 568 unique down-regulated transcripts common to resistant lines (DRI and BDE sets) and they had 209 Blast2GO annotations. Annotated sequences had 152 sequences associated with biological process GO terms, 164 sequences with molecular function GO terms, and 146 sequences with cellular component GO terms. The level 2 GO terms for biological process, molecular function, and cellular component are summarized in Supplementary Figure S1. The most common level 2 GO terms were similar to the ones mentioned in the up-regulated transcripts (BUE and URI) with 49 sequences representing plant defense related GO term response to stimulus.

Sequences annotated with plant defense related GO terms were searched and selected ontologies are depicted in Figure 5. Level 4 and above defense related GO terms with the highest number of associated transcripts included defense response (43 transcripts), response to other organism (33 transcripts), response to external biotic stimulus (33 transcripts), and innate immune response (14 transcripts). These defense related terms are in the biological process GO and were represented by 57 unique transcripts and each transcript was associated with multiple GO terms. These multiple GO terms and the description of proteins encoded by the above 57 sequences are provided in Supplementary Table S2.

Transcripts were identified that shared similarity to known defense response genes. For example, five transcripts (Bodac2015c170619, Bodac2015c160020, Bodac2015c000447, Bodac2015c139533, and Bodac2015c153835) encode ABC transporter-like proteins and have potential to confer non-host resistance (Shimizu et al., 2010). These transcripts were expressed more in leaf spot resistant lines compared (URI, BUE, DRI, and



BDE) to the susceptible lines. The transcript Bodac2015c185871 is similar to a bacterial blight resistance gene. Transcripts Bodac2015c098349, Bodac2015c142490, Bodac2015c170497, Bodac2015c106585, and Bodac2015c139347 were up-regulated in resistant lines and similar to genes encoding *Verticillium* wilt disease resistance proteins. Transcript Bodac2015c141715 found in BUE group was homologous to the gene encoding the immediate-early fungal elicitor protein CMPG1. This had been reported to confer a hypersensitive response (HR) in many plants (González-Lamothe et al., 2006).

We identified a LysM domain receptor-like kinase (Bodac2015c159804) along with several (Bodac2015c146460, Bodac2015c164725, Bodac2015c161960, Bodac2015c142835, and Bodac2015c195842) mitogen-activated protein (MAP) kinase and MAP kinase family sequences. Similar genes were reported to cause pattern triggered immunity (PTI) in rice (Shimizu et al., 2010; Balmer et al., 2013). The transcript Bodac2015c130002 was identified among the BUE transcripts and is similar to *Xa21* which mediates resistance against *Xanthomonas* bacteria in rice (Lee et al., 2009).

Nucleotide-binding site leucine rich repeat (NBS-LRR) like transcripts Bodac2015c176239, Bodac2015c100000, and Bodac2015c134520, which encode nb-arc domain proteins, had higher expression in resistant lines. Eighteen leucine-rich repeat receptor-like protein kinase family transcripts were also expressed more in resistant lines. In addition, 16 RPM1-like disease resistance transcripts and one RPS2-like disease resistance transcript (Bodac2015c147648) were identified. The RPM1-like and RPS2-like disease resistance genes have been reported to confer effector-triggered immunity (ETI) in other plants (Mackey et al., 2002; Jones and Dangl, 2006; Balmer et al., 2013). One transcript (Bodac2015c098081) that was similar to the wheat stripe rust resistance gene *Yr10* and

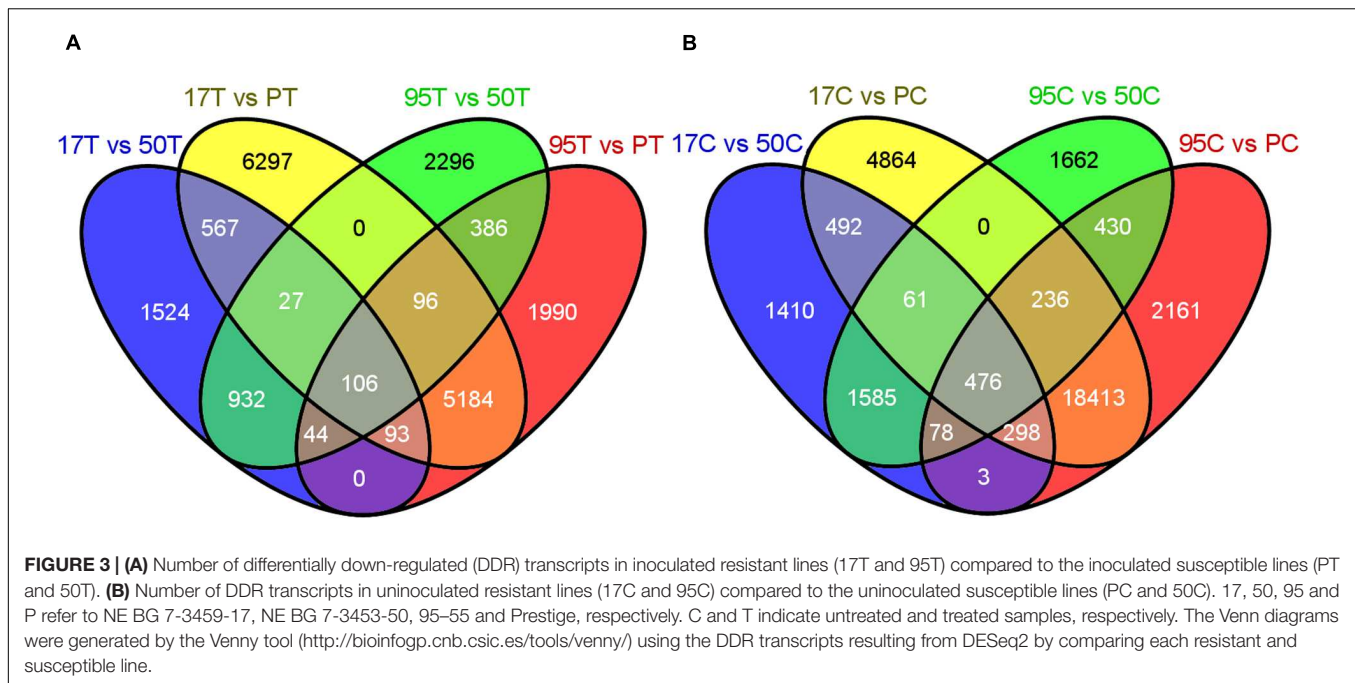
the transcripts Bodac2015c145339 and Bodac2015c146056 were similar to the barley stem rust resistant gene *Rpg1* had higher expression in buffalograss leaf spot resistant lines. The *Yr10* and *Rpg1* genes are also responsible for ETI (Brueggeman et al., 2002; Balmer et al., 2013; Zhang et al., 2013).

The transcripts Bodac2015c139945 and Bodac2015c165851 had higher expression ( $P < 0.006$ ) in leaf spot resistant buffalograss lines and encode a heat shock transcription factor-like protein. Heat shock proteins can fold NBS-LRR proteins and make them active against pathogens (Balmer et al., 2013). Nine transcripts with response to salicylic acid GO term were also identified in leaf spot resistant buffalograss lines. All these were found in up-regulated transcript sets URI and BUE. We also found three transcripts (Bodac2015c146262, Bodac2015c159389, and Bodac2015c130171) encoding pathogenesis-related (PR) proteins with higher expression in the resistant lines.

Although this study was not designed to identify pathogen related genes, we searched for genes expressed by *C. inaequalis* since pathogen encoded RNA may have been included in our sequencing reads. The transcript Bodac2015c163958 showed homology to gene encoding NEP1 effector which is a phytotoxic protein identified in *Botrytis cinerea* (Oliver and Solomon, 2010).

## Gene Expression Validation by RT-PCR

Reverse transcription PCR could distinguish most of the differentially up-regulated genes based on host susceptibility (Figure 6). When the difference of gene expression from the transcriptional profiling study was not high, the intensity of PCR bands was similar across all samples (e.g., J and L genes of untreated resistant vs. untreated susceptible samples in Figure 6 and Supplementary Table S1). The primers used to amplify the endogenous gene *UCE* resulted in a PCR product of the



expected size (61 bp) that was uniformly expressed across all samples.

## DISCUSSION

This study was designed to identify buffalograss defense related genes contributing to host resistance against leaf spot disease. We compared inoculated leaf spot resistant and susceptible buffalograss lines to identify genes with either higher or lower expression in the resistant lines. We also identified differentially expressed transcripts in uninoculated resistant lines when compared to uninoculated susceptible lines. Previous research has demonstrated higher basal expression of defense genes in buffalograss lines resistant to chinch bugs (*Blissus occiduus* Barber) compared to susceptible lines (Ramm et al., 2013). Higher basal gene expression may also play a role in buffalograss defense against leaf spot disease.

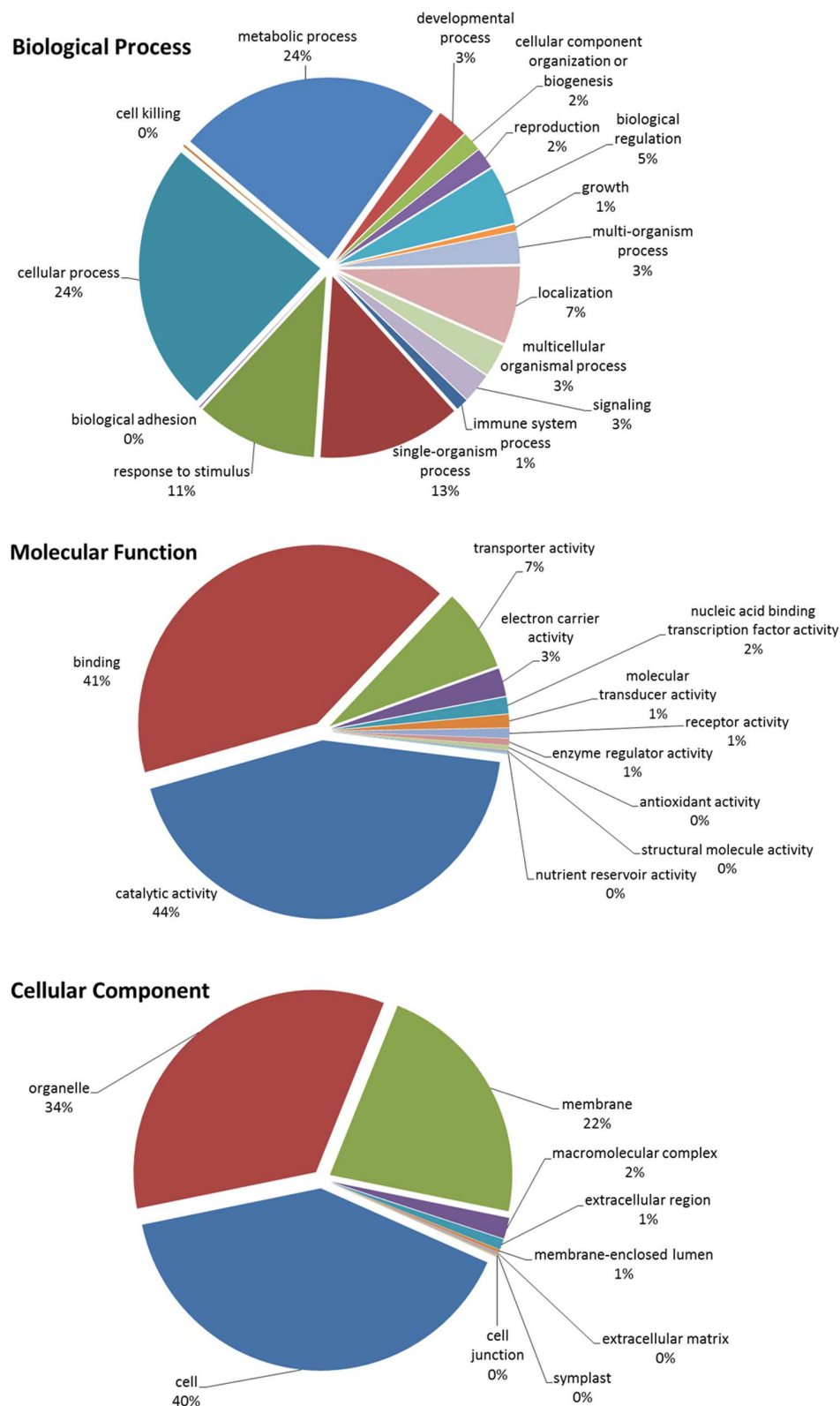
The ABC transporter like proteins have been reported to confer non-host resistance in *Arabidopsis* against the non-adapted pathogen *Blumeria graminis* f. sp. *hordei* (Stein et al., 2006). Genes conferring non-host resistance don't produce HR but are normally involved in rapid production of cell wall appositions (physical barriers) and antimicrobial metabolites at the site of pathogen entry (Jones and Dangl, 2006). Infection by *C. inaequalis* has also induced expression of the ABC transporter-like genes Bodac2015c170619, Bodac2015c160020, Bodac2015c000447, Bodac2015c139533, and Bodac2015c153835 in leaf spot resistance lines compared to the susceptible lines.

During the early stages of a pathogen attack the innate immunity of the host allows the plant to overcome the pathogen. In the first stage of this type of immunity, the microbe-associated molecular patterns (MAMP) such as chitin or flagellin are

recognized by membrane-localized pattern-recognition receptors (PRR) in plants (Balmer et al., 2013) which results in PTI. In the model monocot rice, LysM PRRs CEBiP and CERK1 have been identified for sensing MAMP chitin (Shimizu et al., 2010). MAMP-signaling activates mitogen-activated protein kinase (MAPK) cascades, which regulate transcription factors (TFs) driving the expression of defense genes (Balmer et al., 2013). In these buffalograss samples we also found genes that encode both LysM domain receptor-like kinase and MAP kinase family proteins. These genes may be involved in PTI in buffalograss against *C. inaequalis*.

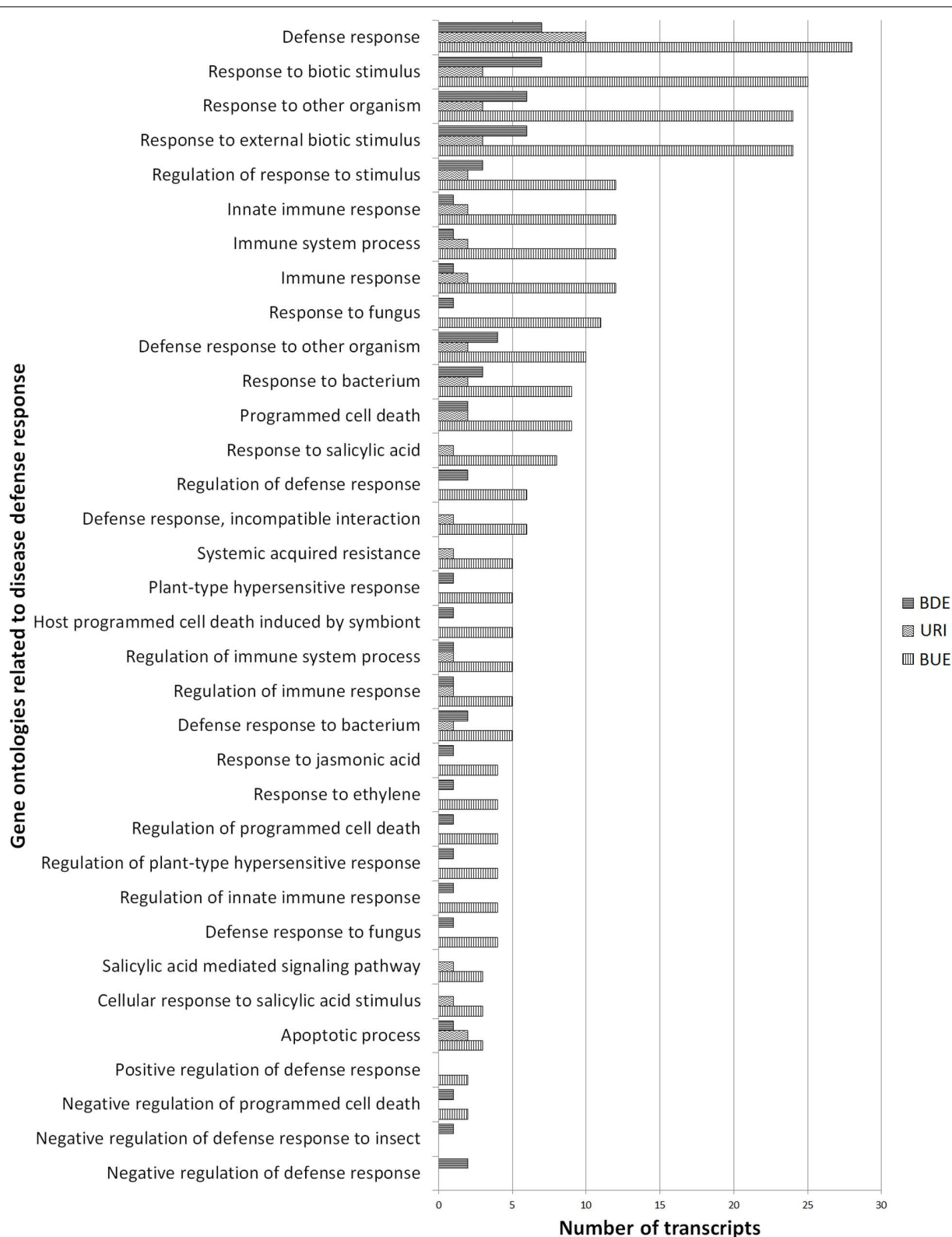
The *Xa21* homolog found in buffalograss was originally reported in rice mediating resistance to *Xanthomonas* bacteria. This gene encodes extracellular and intracellular receptors which can sense the 194-amino acid bacterial protein Ax21 and is conserved across all known *Xanthomonas* strains (Lee et al., 2009). During an attack, XA21 induces downstream defense mechanisms which lead to the expression of pathogenesis-related (PR) genes and the development of HR (Tena et al., 2011). *Xa21* homologs have been found in other grasses such as *Brachypodium*, sorghum, and maize (Tan et al., 2012). Another example of PTI is the B-lectin receptor kinase Pi-d2 which confers resistance against *Magnaporthe grisea* in rice (Chen et al., 2006). We found 15 transcripts similar to lectin-domain containing receptor kinases in leaf spot resistant buffalograss.

To overcome PTI related defense signaling, pathogens release avirulence (Avr) proteins. In the host, a second line of plant immunity takes place mostly in the cytoplasm and is mediated by NBS-LRR proteins encoded by plant resistance (*R*) genes. These NBS-LRR proteins can recognize and neutralize Avr proteins/effectors which results in ETI (Elmore et al., 2011), and is usually manifest in a HR (Balmer et al., 2013). The NBS-LRR family represents one of the largest and widely

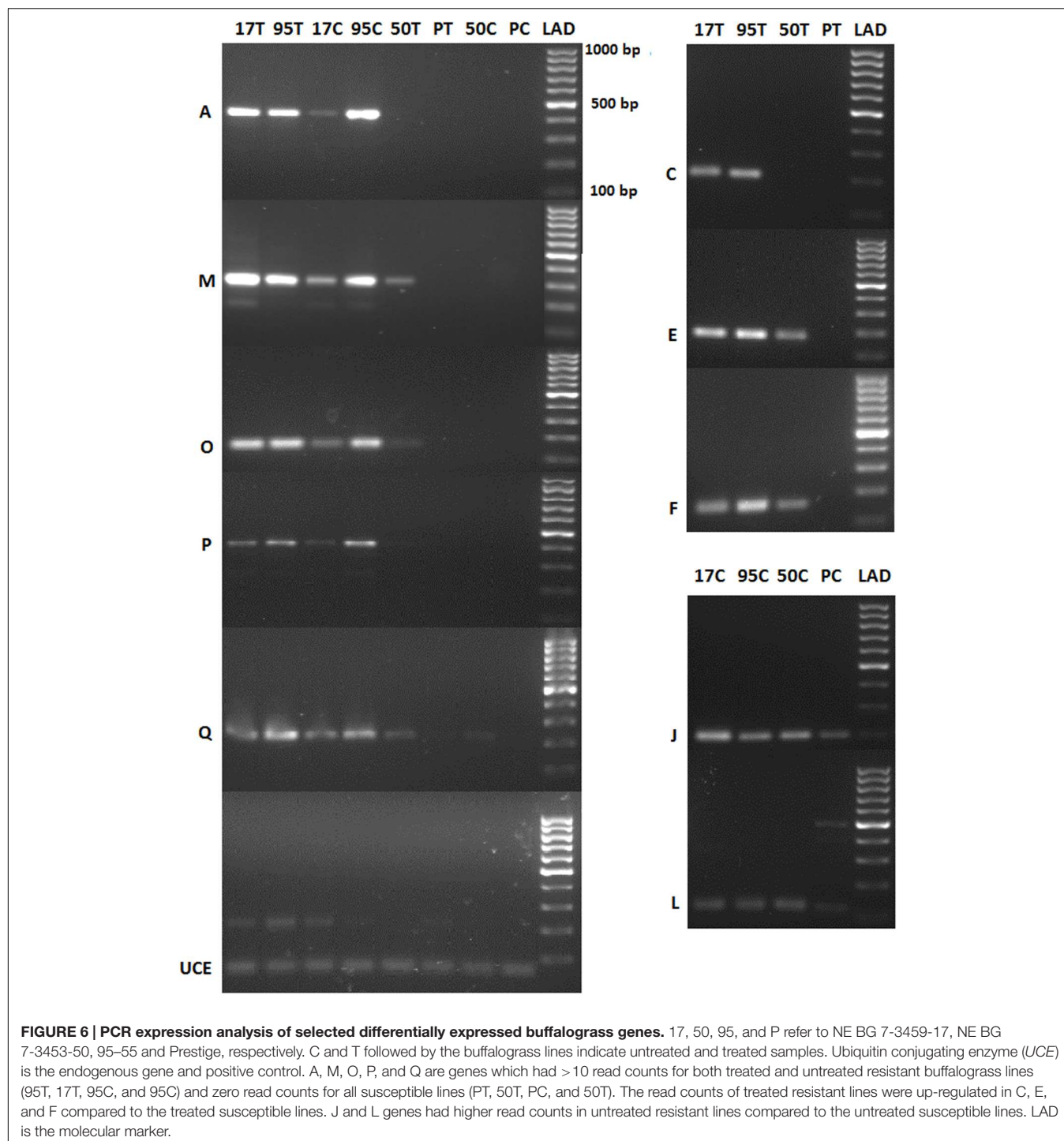


**FIGURE 4 | Blast2GO Level 2 gene ontologies of biological process, molecular function, and cellular component for the total up-regulated transcripts in inoculated and uninoculated resistant buffalograss lines compared with inoculated and uninoculated susceptible lines, respectively.** Of a total of 1356 differentially up-regulated transcripts, 528 were annotated by Blast2GO.





**FIGURE 5 | Number of differentially regulated transcripts associated with host defense response.** The URI group represents the up-regulated transcripts in inoculated resistant buffalograss lines compared to inoculated susceptible lines. The BUE group refers to the transcripts expressed more in uninoculated resistant lines in comparison to uninoculated susceptible lines. BDE represents the down-regulated transcripts in uninoculated resistant lines compared to uninoculated susceptible lines.



conserved gene families in plants. More than one hundred NBS-LRR genes have been identified in the majority of sequenced plants (Balmer et al., 2013). NBS-LRR proteins usually consist of an N-terminal domain with Toll/Interleukin-1 Receptor (TIR-NBS-LRR, or TNL) or an N-terminal coiled-coil (CC-NBS-LRR, or CNL) motif (Meyers et al., 2003). In *Arabidopsis*, RPM1 is a peripheral plasma membrane NBS-LRR protein and in the absence of RPM1, AvrRpm1 effector protein of some

*Pseudomonas syringae* strains can cause virulence (Mackey et al., 2002; Jones and Dangl, 2006). RPS2 has been reported to confer ETI to *Arabidopsis* against *P. syringae* (Jones and Dangl, 2006). We identified several NBS-LRR type genes along with genes that encode an RPM1 and RPS2-like disease resistance proteins.

The *Yr10* and *Rpg1* homologs found in leaf spot resistant buffalograss lines confer disease resistance in other graminaceous plants. Presently, 53 different *Yr* genes (*Yr1*–*Yr53*) have been

identified to cause stripe rust (*Puccinia striiformis* f. sp. *tritici*) resistance in wheat (Zhang et al., 2013). The barley *Rpg1* gene regulates resistance against the stem rust pathogen *P. graminis* f. sp. *tritici* (Brueggeman et al., 2002). Both of these gene products interact with pathogen effector proteins and confer resistance. The stripe rust resistance protein *yr10* and disease resistance protein *rpg1* were identified by the Blast2GO analysis. Similarly, many transcripts encoding NBS-LRR type defense related proteins were also identified. However, a search for transcripts that encode proteins similar to TNL and CNL using a hidden Markov model (HMM; Eddy, 1998; Meyers et al., 2003) may reveal more NBS-LRR genes and this analysis is underway.

We found two heat shock like genes (Bodac2015c139945 and Bodac2015c165851) in leaf spot resistant lines. The NBS-LRR proteins activated by heat shock proteins interact with pathogen effectors and regulate WRKY TFs to confer plant resistance (Balmer et al., 2013). WRKY TFs regulate many plant processes including response to biotic stresses (Zhang and Wang, 2005). Plants under pathogen attack can also show enhanced defense activity in tissues not yet attacked through systemic acquired resistance (SAR). When leaf pathogens show localized infection, defense signals are mobilized to distal plant tissues and induce systemic resistance against a broad range of pathogens including fungi, bacteria, viruses, nematodes, and even insects (Shah, 2009). Prior activation of defense genes in distal tissues renders them more resistance against future attacks. In dicots, salicylic acid has been found to play a major role in regulating SAR followed by the up-regulation of pathogenesis-related (PR) genes (Balmer et al., 2013). Compared with dicots, the knowledge of SAR in monocots is scarce (Balmer et al., 2013). It is interesting that we identified nine transcripts with “response to salicylic acid” GO term in leaf spot resistant buffalograss lines.

We validated the expression of 10 differentially expressed genes (Supplementary Table S1) by RT-PCR. The majority of the tested genes did produce PCR bands with different intensities and could distinguish resistant and susceptible samples confirming the accuracy of our analysis (Figure 6). Quantitative real-time RT-PCR may reveal expression differences of these potential molecular markers more accurately. Interestingly, we found 15 transcripts that have more than 10 read counts in treated and untreated resistant buffalograss and no read counts in the susceptible lines. They may be useful for identifying leaf spot resistant buffalograss lines by molecular methods.

We identified several differentially expressed transcripts that may regulate leaf spot resistance in buffalograss. In particular, the above mentioned 15 sequences that were uniquely expressed in the resistant lines are a new resource for identifying leaf

spot resistant buffalograss lines. We also found 21 transcripts in resistant lines that are similar to genes regulating PTI (e.g., sequences encoding LySM, MAPK, lectin receptor kinase-like proteins) and 20 sequences predicted to encode NBS-LRR proteins RMP1, RPS2, RPG1, and YR10 regulating ETI. There were also nine up-regulated transcripts in resistant lines that have potential to initiate SAR and three transcripts encoding PR proteins. This is the first study characterizing changes in the buffalograss transcriptome when challenged with leaf spot pathogen *C. inaequalis*. The NBS-LRR type defense related genes identified here are useful for screening large numbers of buffalograss germplasm for *C. inaequalis* resistance by molecular techniques thus eliminating laborious and time consuming traditional greenhouse and field testing. In addition, buffalograss is more susceptible to *Curvularia* and *Bipolaris* patch diseases when grown in humid regions. These new molecular resources would improve the efficiency for breeding for leaf spot resistant buffalograss cultivars and may help expand the buffalograss into areas prone of leaf spot outbreaks.

## AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

We thank the United States Golf Association for funding this research.

## ACKNOWLEDGMENTS

We also thank the University of Nebraska DNA Sequencing Core that receives partial support from the NCRR (1S10RR027754-01, 5P20RR016469, RR018788-08) and the National Institute for General Medical Science (NIGMS; 8P20GM103427, GM103471-09). This publication's contents are the sole responsibility of the authors and do not necessarily represent the official views of the NIH or NIGMS.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.00715>

## REFERENCES

- Amaradasa, B. S., and Amundsen, K. (2014a). Expression of leaf spot induced resistance gene-like sequences in *Buffalograss*. *Am. Int. J. Biol.* 2, 1–19.
- Amaradasa, B. S., and Amundsen, K. (2014b). First report of *Curvularia inaequalis* and *Bipolaris spicifera* causing leaf blight of buffalograss in Nebraska. *Plant Dis.* 98, 279. doi: 10.1094/PDIS-05-13-0487-PDN
- Balmer, D., Planchamp, C., and Mauch-Mani, B. (2013). On the move: induced resistance in monocots. *J. Exp. Bot.* 64, 1249–1261. doi: 10.1093/jxb/ers248
- Beetle, A. A. (1950). *Buffalograss*-native of the shortgrass plains. *Univ. Wyoming Agric. Exp. Stata. Bull.* 293, 1–32.
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Brecht, M. O., Stiles, C. M., and Datnoff, L. E. (2007). Evaluation of pathogenicity of *Bipolaris* and *Curvularia* spp. on dwarf and ultradwarf bermudagrasses in Florida. *Plant Health Progr.* doi: 10.1094/PHP-2007-0119-02-RS

- Brueggeman, R., Rostoks, N., Kudrna, D., Kilian, A., Han, F., Chen, J., et al. (2002). The barley stem rust-resistance gene Rpg1 is a novel disease-resistance gene with homology to receptor kinases. *Proc. Natl. Acad. Sci. U.S.A.* 99, 9328–9333. doi: 10.1073/pnas.142284999
- Chen, X., Shang, J., Chen, D., Lei, C., Zou, Y., Zhai, W., et al. (2006). A B-lectin receptor kinase gene conferring rice blast resistance. *Plant J.* 46, 794–804. doi: 10.1111/j.1365-3113X.2006.02739.x
- Conesa, A., Götz, S., Garcia-Gomez, J. M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi: 10.1093/bioinformatics/bti610
- Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics* 14, 755–763. doi: 10.1093/bioinformatics/14.9.755
- Elmore, J. M., Lin, Z. D., and Coaker, G. (2011). Plant NB-LRR signaling: upstreams and downstreams. *Curr. Opin. Plant Biol.* 14, 365–371. doi: 10.1016/j.pbi.2011.03.011
- González-Lamothe, R., Tsitsigiannis, D. I., Ludwig, A. A., Panicot, M., Shirasu, K., and Jones, J. D. G. (2006). The U-box protein CMPG1 is required for efficient activation of defense mechanisms triggered by multiple resistance genes in tobacco and tomato. *Plant Cell* 18, 1067–1083. doi: 10.1105/tpc.106.040998
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883
- Gutierrez-Gonzalez, J. J., Tu, Z. J., and Garvin, D. F. (2013). Analysis and annotation of the hexaploid oat seed transcriptome. *BMC Genomics* 14:471. doi: 10.1186/1471-2164-14-471
- Johnson, P. G., Riordan, T. P., and Arumuganatha, K. (1998). Ploidy level determinations in *buffalograss* clones and populations. *Crop Sci.* 38, 478–482. doi: 10.2135/cropsci1998.0011183X003800020034x
- Jones, J. D. G., and Dangl, J. L. (2006). The plant immune system. *Nature* 444, 323–329. doi: 10.1038/nature05286
- Langmead, B., and Salzberg, S. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Lee, S. W., Han, S. W., Sriyanum, M., Park, C. J., Seo, Y. S., and Ronald, P. C. (2009). A type I-secreted, sulfated peptide triggers XA21-mediated innate immunity. *Science* 326, 850–853. doi: 10.1126/science.1173438
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8
- Mackey, D., Holt, B. F. I., Wiig, A., and Dangl, J. L. (2002). RIN4 interacts with *Pseudomonas syringae* type III effector molecules and is required for RPM1-mediated resistance in *Arabidopsis*. *Cell* 108, 743–754. doi: 10.1016/S0092-8674(02)00661-X
- Meyers, B. C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R. W. (2003). Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* 15, 809–834. doi: 10.1105/tpc.009308
- Oliver, R. P., and Solomon, P. S. (2010). New developments in pathogenicity and virulence of necrotrophs. *Curr. Opin. Plant Biol.* 4, 415–419. doi: 10.1016/j.pbi.2010.05.003
- Ramm, C., Saathoff, A., Donze, T., Heng-Moss, T., Baxendale, F., Twigg, P., et al. (2013). Expression profiling of four defense-related buffalograss transcripts in response to chinch bug (Hemiptera: Blissidae) feeding. *J. Econ. Entomol.* 106, 2267–2576. doi: 10.1603/EC13267
- Reeder, J. R. (1971). Notes on Mexican grasses. IX. Miscellaneous chromosome numbers. *Brittonia* 23, 105–117. doi: 10.2307/2805426
- Riordan, T. P., de Shazer, S. A., Johnson-Cicalese, J. M., and Shearman, R. C. (1993). An overview of breeding and development of *buffalograss*. *Int. Turf. Soc. Res. J.* 7, 816–822.
- Shah, J. (2009). Plants under attack: systemic signals in defence. *Curr. Opin. Plant Biol.* 12, 459–464. doi: 10.1016/j.pbi.2009.05.011
- Shimizu, T., Nakano, T., Takamizawa, D., Desaki, Y., Ishii-Minami, N., Nishizawa, Y., et al. (2010). Two LysM receptor molecules, CEBiP and OsCERK1, cooperatively regulate chitin elicitor signaling in rice. *Plant J.* 64, 204–214. doi: 10.1111/j.1365-3113X.2010.04324.x
- Smiley, R. W., Dernoe, P. H., and Clarke, B. B. (2005). *Compendium of Turfgrass Diseases*, 3rd Edn. St. Paul, MN: American Phytopathological Society.
- Smith, J. D., Jackson, N., and Woolhouse, A. R. (1989). *Fungal Diseases of Amenity Turf Grasses*, 3rd Edn. London: E. & F.N. Spon.
- Stein, M., Dittgen, J., Sánchez-Rodríguez, C., Hou, B. H., Molina, A., Schulze-Lefert, P., et al. (2006). *Arabidopsis* PEN3/PDR8, an ATP binding cassette transporter, contributes to nonhost resistance to inappropriate pathogens that enter by direct penetration. *Plant Cell* 18, 731–746. doi: 10.1105/tpc.105.038372
- Tan, S., Wang, D., Ding, J., Tian, D., Zhang, X., and Yang, S. (2012). Adaptive evolution of Xa21 homologs in Gramineae. *Genetica* 139, 1465–1475. doi: 10.1007/s10709-012-9645-x
- Tena, G., Boudsocq, M., and Sheen, J. (2011). Protein kinase signaling networks in plant innate immunity. *Curr. Opin. Plant Biol.* 14, 519–529. doi: 10.1016/j.pbi.2011.05.006
- Wachholtz, M., Heng-Moss, T., Twigg, P., Baird, L., Lu, G., and Amundsen, K. (2013). Transcriptome analysis of two *buffalograss* cultivars. *BMC Genomics* 14:613. doi: 10.1186/1471-2164-14-613
- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63. doi: 10.1038/nrg2484
- Weizhong, L., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158
- Yang, X., Yu, X. Y., and Li, Y. F. (2013). *De novo* assembly and characterization of the *barnyardgrass* (*Echinochloa crus-galli*) transcriptome using next-generation pyrosequencing. *PLoS ONE* 8:e69168. doi: 10.1371/journal.pone.0069168
- Zhang, X., Han, D., Zeng, Q., Duan, Y., Yuan, F., Shi, J., et al. (2013). Fine mapping of wheat stripe rust resistance gene Yr26 based on Collinearity of wheat with *Brachypodium istachyon* and Rice. *PLoS ONE* 8:e57885. doi: 10.1371/journal.pone.0057885
- Zhang, Y., and Wang, L. (2005). The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants. *BMC Evol. Biol.* 5:1. doi: 10.1186/1471-2148-5-1

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Amaradasa and Amundsen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Validating DNA Polymorphisms Using KASP Assay in Prairie Cordgrass (*Spartina pectinata* Link) Populations in the U.S.

Hannah Graves<sup>1</sup>, A. L. Rayburn<sup>1</sup>, Jose L. Gonzalez-Hernandez<sup>2</sup>, Gyoungju Nah<sup>3</sup>, Do-Soon Kim<sup>3</sup> and D. K. Lee<sup>1\*</sup>

<sup>1</sup> Department of Crop Science, University of Illinois at Urbana-Champaign, Urbana, IL, USA, <sup>2</sup> Plant Science Department, South Dakota State University, Brookings, SD, USA, <sup>3</sup> Department of Plant Science, Research Institute of Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University, Seoul, Korea

## OPEN ACCESS

### Edited by:

Keenan Amundsen,  
University of Nebraska-Lincoln, USA

### Reviewed by:

Elisa Bellucci,  
Università Politecnica delle Marche,  
Italy

Serge J. Edme,  
United States Department of  
Agriculture-Agricultural Research  
Service, USA

### \*Correspondence:

D. K. Lee  
leedk@illinois.edu

### Specialty section:

This article was submitted to  
Crop Science and Horticulture,  
a section of the journal  
Frontiers in Plant Science

**Received:** 27 August 2015

**Accepted:** 28 December 2015

**Published:** 22 January 2016

### Citation:

Graves H, Rayburn AL,  
Gonzalez-Hernandez JL, Nah G,  
Kim D-S and Lee DK (2016) Validating  
DNA Polymorphisms Using KASP  
Assay in Prairie Cordgrass (*Spartina  
pectinata* Link) Populations in the U.S.  
Front. Plant Sci. 6:1271.  
doi: 10.3389/fpls.2015.01271

Single nucleotide polymorphisms (SNPs) are one of the most abundant DNA variants found in plant genomes and are highly efficient when comparing genome and transcriptome sequences. SNP marker analysis can be used to analyze genetic diversity, create genetic maps, and utilize marker-assisted selection breeding in many crop species. In order to utilize these technologies, one must first identify and validate putative SNPs. In this study, 121 putative SNPs, developed from a nuclear transcriptome of prairie cordgrass (*Spartina pectinata* Link), were analyzed using KASP technology in order to validate the SNPs. Fifty-nine SNPs were validated using a core collection of 38 natural populations and a phylogenetic tree was created with one main clade. Samples from the same population tended to cluster in the same location on the tree. Polymorphisms were identified within 52.6% of the populations, split evenly between the tetraploid and octoploid cytotypes. Twelve selected SNP markers were used to assess the fidelity of tetraploid crosses of prairie cordgrass and their resulting F<sub>2</sub> population. These markers were able to distinguish true crosses and selfs. This study provides insight into the genomic structure of prairie cordgrass, but further analysis must be done on other cytotypes to fully understand the structure of this species. This study validates putative SNPs and confirms the potential usefulness of SNP marker technology in future breeding programs of this species.

**Keywords:** Prairie cordgrass, SNP, marker, *Spartina*, polymorphism, transcriptome

## INTRODUCTION

Prairie cordgrass (*Spartina pectinata* Link) is a native grass species of the North American Prairie that has a geographic distribution, ranging from the southern U.S. (Texas, Arkansas, and New Mexico) to northern Canada, and from the east coast through the Midwest to the western coast of the U.S. (Hitchcock, 1950; Voight and Mohlenbrock, 1979; Barkworth et al., 2007; Gedye et al., 2010). This species is adapted to a wide range of environmental conditions and, in addition, responds well to abiotic stresses, such as moderate salinity, water logged soils, drought, and cold tolerance (Montemayor et al., 2008; Boe et al., 2009; Gonzalez-Hernandez et al., 2009; Kim et al., 2011; Zilverberg et al., 2014; Anderson et al., 2015). Because of its wide

adaptability, this warm season, C4, perennial grass is highly valued for conservation practices, wetland revegetation, streambank stabilization, wildlife habitat, forage production, and recently bioenergy feedstock production (Hitchcock, 1950; Barkworth et al., 2007; Montemayor et al., 2008; Gonzalez-Hernandez et al., 2009; Kim et al., 2011; Boe et al., 2013; Zilverberg et al., 2014; Guo et al., 2015). This ability to adapt to such a wide diversity of conditions results in populations becoming adapted to specific environments, ultimately leading to genetically diverse populations. Adding to the potential genetic diversity of prairie cordgrass is polyploidy.

Prairie cordgrass is a polyploid species, composed of three cytotypes: tetraploid ( $2n = 4x = 40$ ), hexaploid ( $2n = 6x = 60$ ), and octoploid ( $2n = 8x = 80$ ) (Church, 1940; Kim et al., 2010, 2012). Because of the reproductive and geographic isolation between the cytotypes, there is likely an increase in polymorphisms and potential genetic diversity, especially within the tetraploid and octoploids cytotypes (Soltis et al., 1992; Wendel and Doyle, 2005; Hirakawa et al., 2014). There is a large amount of phenotypic variation present in all cytotypes of prairie cordgrass (Boe and Lee, 2007; Kim et al., 2012; Guo et al., 2015), but there is a lack of knowledge about the genomic structure. A few studies have revealed diversity within highly polymorphic chloroplast DNA regions observed within and among tetra- and octoploid populations (Kim et al., 2013; Graves et al., 2015). In prairie cordgrass, EST-SSR markers (Gedye et al., 2010), SSR (Gedye et al., 2012), and AFLP markers (Moncada et al., 2007) have been developed. However, these technologies may not be as cost-effective, scalable, successful, or as flexible as using single nucleotide polymorphisms (SNPs) (Semagn et al., 2014).

SNPs provide a highly efficient way to conveniently compare genomic and transcriptome sequences. Because they are one of the most abundant DNA variants found in plant genomes, SNPs are more likely to be related to specific biological functions and phenotypes (Rafalski, 2002; Bundock et al., 2006; Salem et al., 2012). This technology has been applied in genetic diversity analysis, genetic map construction, association map analysis, and marker-assisted selection breeding in many different types of crop species (Byers et al., 2012; Saxena et al., 2012; Semagn et al., 2014; Sindhu et al., 2014; Wei et al., 2014). SNP marker technology is also utilized in high-throughput genotyping, increasing the speed of the selection process by eliminating growing plants to maturity for phenotypic selection (Paux et al., 2012). In order to use SNP markers for genetic improvement, there is a three-step process one must follow: (1) SNP discovery after aligning sequence reads generated by next-generation sequencing technologies for different genotypes of a given species; (2) validate SNPs to distinguish DNA polymorphisms of actual allelic variants from those of other biological phenomena such as gene duplication events; (3) SNP genotyping of germplasm collection or genetic/breeding populations (Saxena et al., 2012).

Step one of the process was accomplished in prairie cordgrass by using a transcriptome assembly derived from multiple genotypes and tissues (Gonzalez et al., personal communication). The second and third steps are yet to be completed for polyploid prairie cordgrass. Several parameters, such as sample size,

number of SNPs to be used for analysis, cost effectiveness, and the SNP genotyping platform, must be considered in these analyses (Semagn et al., 2014). Many technologies exist for use in SNP genotyping analysis, but one technology performs well when it comes to adaptability, efficiency, and cost-effectiveness. Kompetitive allele-specific PCR (KASP), developed by LGC Genomics (Teddington, UK; [www.lgcgenomics.com](http://www.lgcgenomics.com)), is a PCR-based homogeneous fluorescent SNP genotyping system, which determines the alleles at a specific locus within genomic DNA (Semagn et al., 2014). The KASP technology has been utilized on other polyploid plant species, including switchgrass (LGC Genomics, 2014), cotton (Byers et al., 2012), wheat (Paux et al., 2012), potato (Uitdewilligen et al., 2013), and various triploid citrus species (Cuenca et al., 2013).

In this study, SNPs, identified in the nuclear transcriptome, were converted to the KASP marker system in order to validate that these SNPs are true allelic variants. In addition, KASP markers were used in quality control analysis when making crosses, prairie cordgrass being a putative self-compatible species. The main objectives of this study were (1) to validate SNP polymorphisms identified in the nuclear transcriptome of natural populations of prairie cordgrass in the U.S. and (2) to assess the fidelity of specific tetraploid crosses and selfs, and to elucidate inheritance patterns of SNP markers.

## MATERIALS AND METHODS

### Development and Validation of KASP Genotyping Assays

In a separate study by Gonzalez et al. (personal communication) at South Dakota State University, a transcriptome of prairie cordgrass was assembled using ~1.2 billion Illumina paired-end reads from various vegetative tissues (roots, leaves, and rhizomes) under various conditions (salt stress, cold stress, and differing photoperiods) in order to obtain an abundance in diversity, with regards to the number and type of transcripts. The assembly was developed using CLC Genomics Workbench 7.0 (Arhaus, Denmark) and annotated against the sorghum genes models. About 146,549 contigs, or transcript assemblies, of 230 bp or more with an N50 of 973 bp were used to mine over 1 million SNPs, insertions, and deletions using the variant detection function in CLC Genomics Workbench. Putative SNPs were filtered based on coverage (minimum of 100 X), a window of 80–100 bp free from additional SNPs and an allele frequency of 20–80%. Initially, nine bi-allelic SNPs were selected for analysis, associated with enzymes within the lignin biosynthesis pathway. Additional SNPs were selected without regard to putative function of the transcript assembly. A total of 121 bi-allelic SNPs were identified for use in this study (Table 1). SNPs were sent for primer development to be used in KASP genotyping assays. Genotyping with KASP was performed as follows.

For all samples, each amplification reaction contained 50 ng template DNA, KASP V4.0 2x Master mix standard ROX (LGC Genomics, Beverly, MA, USA) and KASP-by-Design assay mix (LGC Genomics, Beverly, MA, USA). The PCR thermocycling

**TABLE 1 | Summary of SNP sequences, including SNP ID, SNP sequences, and SNP alleles.**

SNP ID	SNP sequence/allele	SNP ID	SNP sequence
pcg_00001	GTCTTGTAGCTCGGC[G/A]TCCACGTCCAAGCG	pcg_00032	GCCAGTATTGGCAAG[A/C]ATGCAACAATTA
*pcg_00002	CGCCAGGTACACCGG[C/G]GCCGCTGTTAGT	pcg_00033	AAAGACTACCTTCC[A/C]TATCGAATAGAGAA
pcg_00003	GTCGGCCCCGGCCTC[A/G]AACACGGGACGCC	pcg_00034	ACAGCTCCGGATGAA[A/G]TGGTACTTGATCCG
pcg_00004	ACCCGAAGGAGAAGG[G/T]CGCGATGGCGCCCG	*pcg_00035	TCTTTGACCAAGTA[A/G]CTCACCCAGTAGGC
pcg_00005	AAGAACAAATTTATA[A/G]GTTAAATACATGCA	*pcg_00036	GCTCGTGTGATGTC[G/A]CCGGCGAGGTGCT
pcg_00006	GCCAAAGGACAGATC[A/G]TGAATAACATGACT	pcg_00037	CGAGGTGTGATGCAC[T/C]AGAACGCCGCTCGT
pcg_00007	CGGAAGTGAAGAAC[A/G]TAGCATACATGCTT	*pcg_00038	GCTCACATACCCGAC[A/G]GCGAACGCCAAGTC
pcg_00008	GTTGACCCGCGCGGC[A/C]ATCGCCGAGCTCGA	*pcg_00039	TGGGCGAGGTTGCAG[T/C]CACCCATGCCTCCC
pcg_00009	GAGAAGAAGAGAGTG[A/G]TTGCATCATTGGAC	*pcg_00040	CGCTTCTTCCGTGCC[A/G]GTGATGACGAGGTC
pcg_00010	GGTGGCGCTTGACAA[T/C]GTCACAATACAAGT	*pcg_00041	GTGTCCCCGGCCTCG[C/T]CGGTACACCCGCCG
pcg_00011	CTGTTTGTAAAGTGC[A/G]CTGAATTTGAGATT	pcg_00042	GCGGTGCTTGCCGCA[A/G]CCCGTACAAGGCCT
pcg_00012	GCATTCATGTTCCCA[A/G]TACATCCTGGCAAA	*pcg_00043	CTTCTTGAGCTTGAA[T/C]ACCCACTTCAGGGT
pcg_00013	ACAATCATTGTTTTT[T/C]GTAATTGGGGA	*pcg_00044	CGGGCGGTGGCCGGC[T/C]GGCAAGTCGACGAG
pcg_00014	CCAAATGGCAAAAT[T/G]TACTCAGATTTCCA	pcg_00045	AAGTCAGTTGTTGTC[T/A]GCAACCTCATCGT
pcg_00015	ACTTGATTAGAGTC[G/A]GCAGACATCATTTT	*pcg_00046	TCTGTTGATTACCA[C/T]GGTAAGCTCACTCA
pcg_00016	AGCGCTTCACGCGAT[A/G]GAGTTCTCCGAAAT	*pcg_00047	TTACCAAATACCCAG[A/G]TGACAGAGTTCAAGC
pcg_00017	TAGCTTTAGGTGTTG[G/A]GTTTCGCATCAGTA	pcg_00048	AAGCAACAACCTACT[C/T]GAGCAAAAGTCAAG
pcg_00018	AGAACCAACTCTTTA[C/T]ATCAGACTGCGTAT	pcg_00049	TTACTTTTATATAAC[G/A]GGATGAAGCATGCA
pcg_00019	AACAAAGACAACATG[A/G]CTCAGGAGAAATTG	pcg_00050	CAGGGACATTCGTTT[C/T]GTCTCCAAAATA
pcg_00020	TTTGGATGTTGAAC[T/G/A]TCTCAGATGTCTT	pcg_00051	CCCTTGAATGGCTTC[T/C]TTTTCTTTTGTGCA
pcg_00021	TTTGGATGTTGAAC[T/G/A]TCTCAGATGTCTT	pcg_00052	CACCAACCACTTGTG[C/A/G]TGGTGACGCTTCGT
pcg_00022	ATGAATTTTGGCACG[A/G]ACTTTTTGTTTGA	pcg_00053	CGAGGTTGATGTTTA[T/C]GCTCGTCGATGACG
pcg_00023	GCATCCACAAGATG[G/C]CCATGAACAATTAA	pcg_00054	AAAGTATTTGTAGGA[G/A]ACCCCTGAGGGTTC
pcg_00024	GATCGAGAAAAAAA[A/T]TTGGATGAAGATTC	pcg_00055	CTCGCGTGGCCTTCT[C/G]TGTCATAAACCATG
pcg_00025	TTTGAGGAGGACGGT[G/A]ATGATAGCAAACTCT	*pcg_00056	GGAACGTATCCTGTG[T/C]ATAAGGGCTCTCCG
pcg_00026	GTGAGGGATAGATTG[T/G]CAAGCAATGCAAGT	pcg_00057	AACTTGGTATCAGAC[C/G]GCCAAGGTTAAACC
pcg_00027	CCATCTAAGGTCAGG[A/G]TTCTAAGTTCAATC	pcg_00058	GGCACGGTAAACCTT[T/G]GCAAAAGTCCCTTG
pcg_00028	ACATTCCTCCGATCT[C/A]GGGTTTTTAACCCA	pcg_00059	TCAACCGTCTCCCC[C/G]AGATGATTGTCTAA
pcg_00029	GGACCATTTGTTGTC[A/G]TCAAGGTTTCCAG	pcg_00060	CACCCCAAGACCA[T/A]ATGTGCGGCTTTTGC
pcg_00030	GAGAGCATTGATGTC[G/A]CTGGCTCTTGAAA	pcg_00061	AAATCTTTTTTCCA[G/T]TATCTTTTTCTTA
pcg_00031	AGTTAGACCTGAGAT[T/C]GAACATCTGAAAA	pcg_00062	GTAATTGTTTGCAGA[C/G]AACTTTTCATTTGT
pcg_00063	GGAAGATATGCAACA[C/T]TTTGGGAGGAAGC	pcg_00093	TACTGGGAAGAAACC[G/A]TCCACTTGTCTCTG
pcg_00064	GGGGATGTACCCCTT[C/T]CCCGCGCGGTGAT	pcg_00094	GCTCTCCGCACACGC[C/T]GCCACCGCTACATC
pcg_00065	CAGCGGCAGCGACGC[G/A]GCGCTCCTGAGCCC	*pcg_00095	TGGTGAAAAGGTCCT[G/C]ATCCAGTTTGAGGA
pcg_00066	CGGCTTCGACCCGCT[C/G]GGCTTGCGGAGGA	*pcg_00096	CAGGGACCGGAACCG[G/A]TTCCACCGGTTACG
pcg_00067	GAGGATGTTGTCGAG[C/T]TTGACGTGCGGGTG	*pcg_00097	TTTTGTTACAAAATA[C/T]GAGCAAGCTCTGTT
*pcg_00068	CAATCTGGAAAGGA[C/T]CCACTAATGTTTGT	pcg_00098	GTACAATGTCTGGGC[C/A]AGTACTCCTAATGG
pcg_00069	TGAAGTAACTACTAA[A/T]ATAGTACTGTTGTA	pcg_00099	AAAAAAGATGATG[A/T]CAGGTTACAAATTG
pcg_00070	AGGCTCTCACGATCA[T/G]TCCGAGTCTGCTGTC	*pcg_00100	GACTCTCTACGGCTC[C/A]TCCAGGCTCACCGC
pcg_00071	GGCAAGGCTTTTACA[A/C]AAGAAGTTGTCGAG	pcg_00101	AGTACATGCAGGAGG[G/A]GCATTCTCTTCTT
*pcg_00072	GGAGTACAATGAAA[A/G]CTTCATGTGCTGG	pcg_00102	CCATTTGAATCTCAA[G/A]GCATGACGTGAAC
pcg_00073	TTTCCCTGGATTTGG[C/T]CTGGGTCTTGTTAT	pcg_00103	GCTAGCTTTTGCGCC[C/T]CTATACATCTTTTC
*pcg_00074	CGAGCATATAATATG[G/A]CCCTAAATGATGG	pcg_00104	CGTCCTCGTCTGCTT[C/G]TTCTCTGCGTGCT
pcg_00075	CGGCCGCGAGGACTC[G/C]CCGCTCGACATCAT	pcg_00105	GCTTGTGCTCATGGA[T/C]GTGGTTCACAGCCA
pcg_00076	CATCCCCACCTACGT[C/G]GTCGGAGTCAATGC	pcg_00106	ATTGGTGCTGTTGCT[G/C]GACGTGAAGCTGAC
*pcg_00077	CTCCTGCACCAACAA[C/T]TGCCCTCGCGCCCTT	pcg_00107	AGATGACGAGTCCGG[C/A]GACGACGTGGGAGC
pcg_00078	ATGGAGGGACACAGC[C/A]GGCAAGGTGATGT	pcg_00108	CTCTTTGCGCATGTG[A/G]CTCTTTTCCAGGGC
pcg_00079	AGATTCTGATATTGA[T/C]TTGGATGACTATTC	pcg_00109	ACTCAGACCATTTTG[A/G]ACCACCTCAGATGT
pcg_00080	TGCGTATATTCTCCG[T/G]GGTGAGACCAAAAT	pcg_00110	TATGTTATCTCAATG[T/G]GATCTACACCTGCA
pcg_00081	GCTCGCCCTCGCAAC[T/A]ATCGGATCTTGCGC	pcg_00111	GCCGACGGGATGCGG[C/G]CGATTACATTTGC
*pcg_00082	CTGGCTGTAGGAATG[G/A]CCTTTTACCTGAA	pcg_00112	TGACCACATGCCATG[A/G]GTATCAAGCCTATT
pcg_00083	TGAAGTTATGTATGA[T/C]CTGAGAGCTAGTGG	pcg_1186	GACCTCGCAGAACAC[T/C]GCAGACATGACCTC

(Continued)

TABLE 1 | Continued

SNP ID	SNP sequence/allele	SNP ID	SNP sequence
pcg_00084	AAGTTCGGGATCAGC[A/T]CCGTGTATTGGGA	pcg_13880	TCAAGTACCTCACCG[G/A]CGAGGCCAAGGCTT
pcg_00085	CTTCTGAAGTCGGAA[C/A]TGCCATCAAATGG	pcg_14142	CACGCAGTTGGGGGC[C/G]AGGATGAGGACGAC
*pcg_00086	AGGAGTATCCACCTG[G/T]AATAACACTTGATAC	pcg_2412	CACATTGCGATTAGC[G/A]TATCGATCATGAAA
*pcg_00087	CAACACAATGAATCG[T/G]ATTGGAAAAGGAAG	pcg_37652	ACTTGAAGAGAGACG[C/A]ATCTGAAGGCAGAT
pcg_00088	TTTACAAATGCATAA[A/G]ATCTATGTTGGTAA	*pcg_38909	CACGCAGTTGGGGGC[C/G]AGGATGAGGACGAC
pcg_00089	CGTACCTGCAGTTCA[T/C]GTTGCGCTACATCT	pcg_77221	GAGCTCGCCAGGCAC[G/T]CTGGCTTCTGTGGC
pcg_00090	GAGGGGTAGTAAGAA[A/G]ACAAAGGAGACGTG	pcg_7965	TGACCAGCCGACGCA[G/A]CCGCTCGTGGTAGT
pcg_00091	GGTACATAGTTTGAT[C/T]CACCTCCCTTCCTC	pcg_80876	TGGCGTCGTAGGTGC[G/A]CCACGGAGGACGCG
*pcg_00092	ATGGGAAGACAGGTT[T/C]GCAGCTTCATTATT		

\*Failed primers.

Bold letters are actual SNPs (SNP alleles).

conditions for all primers, except pcg\_1186, was 15 min at 94°C followed by 10 cycles of 94°C for 20 s and 61°C for 1 min (dropping  $-0.6^{\circ}\text{C}$  per cycle to achieve a 55°C the annealing temperature) followed by 26 cycles of 94°C for 20 s and 55°C for 1 min. The PCR thermocycling conditions for primer pcg\_1186 was 15 min at 94°C followed by 10 cycles of 94°C for 20 s and 65°C for 1 min (dropping  $-0.8^{\circ}\text{C}$  per cycle to achieve a 57°C annealing temperature) followed by 26 cycles of 94°C for 20 s and 57°C for 1 min. After amplification, PCR plates were read with a Spectramax M5 FRET capable plate reader (Molecular Devices, Sunnyvale, CA, USA) using the recommended excitation and emission values. Data was then analyzed using Klustercaller software (LGC Genomics, Beverly, MA, USA) to identify SNP genotypes.

## Core Collection Analysis

In order to validate SNP polymorphisms of prairie cordgrass using KASP, seeds and rhizomes of natural populations were collected from across the continental U.S.A. (Kim et al., 2013) and grown at the Energy Biosciences Institute (EBI) Farm, Urbana, Illinois, USA. Individuals from 38 of these populations were selected as core collection based on geographic distribution; and two plants from each population were sampled, for a total of 76 plants (Table 2). Leaf tissue samples were stored at  $-80^{\circ}\text{C}$  until DNA extraction was performed. Total genomic DNA was extracted from frozen leaf tissue using the CTAB method (Mikkilineni, 1997) with slight modifications as described by Kim et al. (2013). Fifty-nine KASP genotyping assays out of 121 were selected and used to analyze the collection and five additional *Spartina* species samples, namely; *S. alterniflora*, *S. patens* (Flageo vt.), *S. patens* (Sharp vt.), *S. patens*, and *S. bakeri*. All of the KASP genotyping assay results were recorded as a two-letter code, or SNP code, i.e., AA, AG, GG. A DNA fingerprint was made using all the SNP genotypes creating a concatenated DNA-like sequence, which was then imported into MEGA 6 (Tamura et al., 2013) to make a phylogenetic tree. The maximum parsimony (MP) tree, inferred from 1000 replicates, was obtained using the Subtree-Pruning-Regrafting algorithm with a search level one in which the initial trees were obtained by the random addition of sequences (Felsenstein, 1985; Nei and Kumar, 2000). All positions with  $<95\%$  site coverage were eliminated.

## F<sub>1</sub> Cross

In order to assess the utility of the KASP marker system in confirming specific tetraploid crosses of prairie cordgrass, a reciprocal cross involving two individuals (PC17-109  $\times$  PC20-102) of two populations differing in morphological characteristics of potential agronomic importance was developed. PC17-109 is a tetraploid population from Illinois with a phalanx rhizome type and low seed mass, whereas PC20-102 is a tetraploid population from Kansas with a guerilla rhizome type and high seed mass. In a greenhouse, the female inflorescence was covered  $\sim 1$  day prior to stigma emergence, while pollen was collected from the male parent. Pollen was directly applied to the stigmas with a brush, and rebagged until anthesis was completed. A total of 83 individuals, 70 F<sub>1</sub> individuals from PC17-109 (female)  $\times$  PC20-102 (male) and 13 F<sub>1</sub> individuals from PC20-102 (female)  $\times$  PC17-109 (male) were sampled. F<sub>1</sub> seeds were planted in greenhouse setting. Leaf tissue samples of each seedling were collected and stored at  $-80^{\circ}\text{C}$  until DNA extraction was performed. Total genomic DNA was extracted from frozen leaf tissue as described previously. For the F<sub>1</sub> individuals, 12 KASP genotyping assays were selected based on the parental SNP genotypes (Table 3). All of the assay results were recorded as two-letter SNP codes. To determine if the F<sub>1</sub> progeny followed segregation of a typical monohybrid cross in relation to SNP genotype, a  $\chi^2$  analysis was performed using  $P = 0.05$ ,  $df = 2$ , and  $\chi^2$  critical value = 5.991. The observed, along with the expected genotype, was recorded for each KASP genotyping assay.

## F<sub>2</sub> Self

To assess the utility of the KASP marker system in identifying selfed individuals in the tetraploid background and gauge the segregation pattern, F<sub>2</sub> individuals were generated and genotyped. In a greenhouse, the prairie cordgrass inflorescence was covered  $\sim 1$  day prior to stigma emergence with bags constructed to view progression of inflorescence development of F<sub>1</sub> plants. When anthesis was reached, the bags were shaken to promote self-pollination. Bags remained until anthesis was complete. F<sub>2</sub> seeds were collected and planted in a greenhouse setting. A total of eight F<sub>1</sub> individuals were selfed (6 F<sub>1</sub> of



**TABLE 2 | Summary of plant materials used including, location, cytotype, and number of plants used per population.**

ID	Location	Ploidy	Number of samples
103 4X	IL	4X	2
9046803	NY	4X	2
IL102	IL	4X	2
IL99A	IL	4X	2
MBB4X	IL	4X	2
PC09-101	CT	4X	2
PC09-102	CT	4X	2
PC17-109	IL	4X	2
PC17-111 4X	IL	4X	2
PC19-101	IA	4X	2
PC19-103	IA	4X	2
PC19-105	IA	4X	2
PC20-102	KS	4X	2
PC20-105	KS	4X	2
PC22-101	LA	4X	2
PC23-101	ME	4X	2
PC23-104	ME	4X	2
PC29-101	MO	4X	2
PC29-104	MO	4X	2
PC34-101	NJ	4X	2
PC40-101	OK	4X	2
PC55-102	WI	4X	2
PC55-103	WI	4X	2
ND-2-51-4	ND	8X	2
PC17_111 8X	IL	8X	2
PC19-106	IA	8X	2
PC19-107	IA	8X	2
PC19-108	IA	8X	2
PC20-104	KS	8X	2
PC20-106	KS	8X	2
PC27-103	MN	8X	2
PC31-101	NE	8X	2
PC31-104	NE	8X	2
PC38-101	ND	8X	2
PC40-104	OK	8X	2
PC46-110	SD	8X	2
PCG109	SD	8X	2
Red River	MN, SD, ND	8X	2
Total			76

PC17-109 × PC20-102 and 2 F<sub>1</sub> of PC20-102 × PC17-109) and 8–11 individuals were sampled from the planted seeds of each of the selfed plants (total of 76). Leaf tissue samples were stored at −80°C until DNA extraction was performed. All 12 of the KASP genotyping assays selected to score the F<sub>1</sub> individuals were also tested on the F<sub>2</sub> individuals. All of the assay results were recorded as a SNP code as done in the F<sub>1</sub> analysis. All SNP codes that were not accurately identified were removed from analysis.

## RESULTS

### Development and Validation of KASP Assays

Twenty-six (21.5%) SNPs failed KASP marker development. From the remaining 95 (78.5%), 11 SNPs were found to be monomorphic when tested on the core collection DNA, resulting in 84 SNPs that were true allelic variants. Three of the eleven monomorphic markers were selected to discover if future plant samples would reveal the SNP polymorphisms previously identified in the transcriptome. From the 84 allelic variants, 56 of the most highly polymorphic SNPs were selected for further use in this study, resulting in 59 total KASP genotyping assays (Table 4).

### Core Collection

The resulting data set from the DNA fingerprint contained 118 characters. There was an average of 3.8 missing character data points (SNP codes) per population. The maximum parsimony tree identified one clade after correcting for the missing data (Figure 1). For 47.4% of the populations, plants sampled from the same populations were observed to form subclades; however, intrapopulation variation was observed.

Out of the 38 prairie cordgrass populations, 52.6% showed polymorphisms within populations. Of the 52.6% polymorphic populations, 50% were octoploid and 50% were tetraploid. Out of the 15 octoploid populations sampled, 66.7% of the populations showed polymorphisms between the two plants sampled and 43.5% of the 23 tetraploid populations showed polymorphisms. The average number of polymorphisms that occurred within each population was 16. In the octoploid populations, 16.4 was the average number of polymorphisms observed, and 15.5 polymorphisms were observed as the average for tetraploids.

### F<sub>1</sub> Analysis

Only 6 out of 59 possible KASP genotyping assays showed both parents as homozygous SNPs but for opposite alleles. Three representative assays were selected which showed one SNP heterozygous for one parent and one SNP homozygous for the other parent, and three representative assays were selected which showed both parents as heterozygous SNPs (Table 3). All SNP codes that could not be accurately identified or called, due to not appearing in one of the three genotypes, were removed from the  $\chi^2$  analysis. Four individuals did not consistently satisfy the expected heterozygous SNP genotype, with regards to KASP genotyping assays for which both parents were homozygous for opposite alleles (pcg\_00050, pcg\_00058, pcg\_00059, pcg\_000106, pcg\_1186, and pcg\_14142). These four individuals, after being analyzed across all 12 assays, were identified as being selfs, and were removed from the  $\chi^2$  analysis (Table 3). Using the resulting trimmed data, the  $\chi^2$  analysis indicated normal monohybrid 1:2:1 and 1:1 Mendelian inheritance patterns and could not be rejected for any of the primers (Table 5).

### F<sub>2</sub> Analysis

The F<sub>1</sub> parent genotype was identified in order to find SNPs that indicated the parent was homozygous (Table 6). For 3 F<sub>1</sub>

TABLE 3 | Primers selected for use on the prairie cordgrass F<sub>1</sub> progeny.

	pcg_00011	pcg_00012	pcg_00024	pcg_00050*	pcg_00058*	pcg_00059*	pcg_00106*	pcg_1186*	pcg_14142*	pcg_00061	pcg_00062	pcg_7965
PC17_109 (Parent)	GA	GA	TA	TT	GG	CC	CC	CC	GG	TG	GC	AG
PC20_102 (Parent)	GG	GG	AA	CC	TT	GG	GG	TT	CC	TG	GC	AG
13_F1001†	GG	GG	AA	CC	TT	GG	GG	TT	CC	TT	CC	AG
13_F1002	GA	GA	TA	TC	GT	CG	CG	CT	GC	GG	GG	GG
13_F1003	GA	GA	TA	TC	GT	CG	CG	CT	GC	GG	GG	AG
13_F1004	GA	GA	TA	TC	GT	CG	CG	CT	GC	TG	GC	AA
13_F1005	GA	GA	TA	TC	GT	CG	CG	CT	GC	TT	CC	AG
13_F1006	GA	GA	TA	TC	GT	CG	CG	CT	GC	TG	GC	AG
13_F1007†	GG	GG	AA	CC	TT	GG	GG	TT	CC	GG	GG	AG
13_F1008	GG	GG	AA	TC	GT	CG	CG	CT	GC	TG	GC	AA

The first three primers indicate one parent as heterozygous and one parent as homozygous, the next six primers indicate both parents as homozygous for opposite alleles, and the last three primers indicate both parents as heterozygous. Also shown are SNP assay results for eight out of the 83 F<sub>1</sub> hybrids. Indicated are samples that can be identified as true crosses and selfs.

\*Primers that can distinguish true crosses from selfed samples.

†F<sub>1</sub> individuals that are identified as selfs of the PC20\_102 parent.

parents that were selfed, there were F<sub>2</sub> progeny that did not fall into the expected homozygous parental genotype (example in Table 7). Two F<sub>2</sub> progeny were identified consistently as unexpected offspring genotype of 13-F1008, 1 progeny of 14-F1014, and 4 progeny of 14-F1071. Individuals that consistently fell into the heterozygous (unexpected) genotype category across multiple homozygous primers were considered outcrosses and not true selfs of the F<sub>1</sub> (Table 7). Most of the F<sub>2</sub> progeny were identified as expected SNP genotypes when considering the parental genotype.

DISCUSSION

In order to validate SNP polymorphisms in prairie cordgrass, 121 SNPs identified from the nuclear transcriptome were sent for KASP assay development. Among 121 SNPs, the assay success rate was 78.5% with 26 assays failing development. This is comparable with findings in the literature of success rates of 83% (Cockram et al., 2012), 88.4% (Saxena et al., 2012), and 80.9% (Semagn et al., 2014). The assays failed mainly due to paralogs within the prairie cordgrass genome. Because not all of the populations used to develop the transcriptome were in the core collection of DNA used in this study, some assays appeared as monomorphic. These selected SNPs may have been derived from the octoploid populations not present in the core collection. Three monomorphic SNPs were selected for further analysis, to see if the SNPs would be polymorphic in future studies. With the failed and monomorphic assays removed, 84 putative SNPs were validated as true allelic variants and 59 SNPs were selected for this study. The 59 highly polymorphic assays were selected based on the criteria that there were at least two of the three genotypes present in a large portion of the samples analyzed. These assays were tested on the 38 natural populations, creating a phylogenetic tree that resulted in one clade containing all of

the prairie cordgrass populations. If subclades were observed, the two plants of a single population were represented in the subclade.

Just over half of the populations showed polymorphisms within, with an equal number of octoploid and tetraploid populations. The average number of polymorphisms that occurred within each population did not vary between octoploid and tetraploid populations. This is different from a chloroplast DNA study of prairie cordgrass, in which there was little, if any, polymorphisms observed in the tetraploid cytotype (Graves et al., 2015).

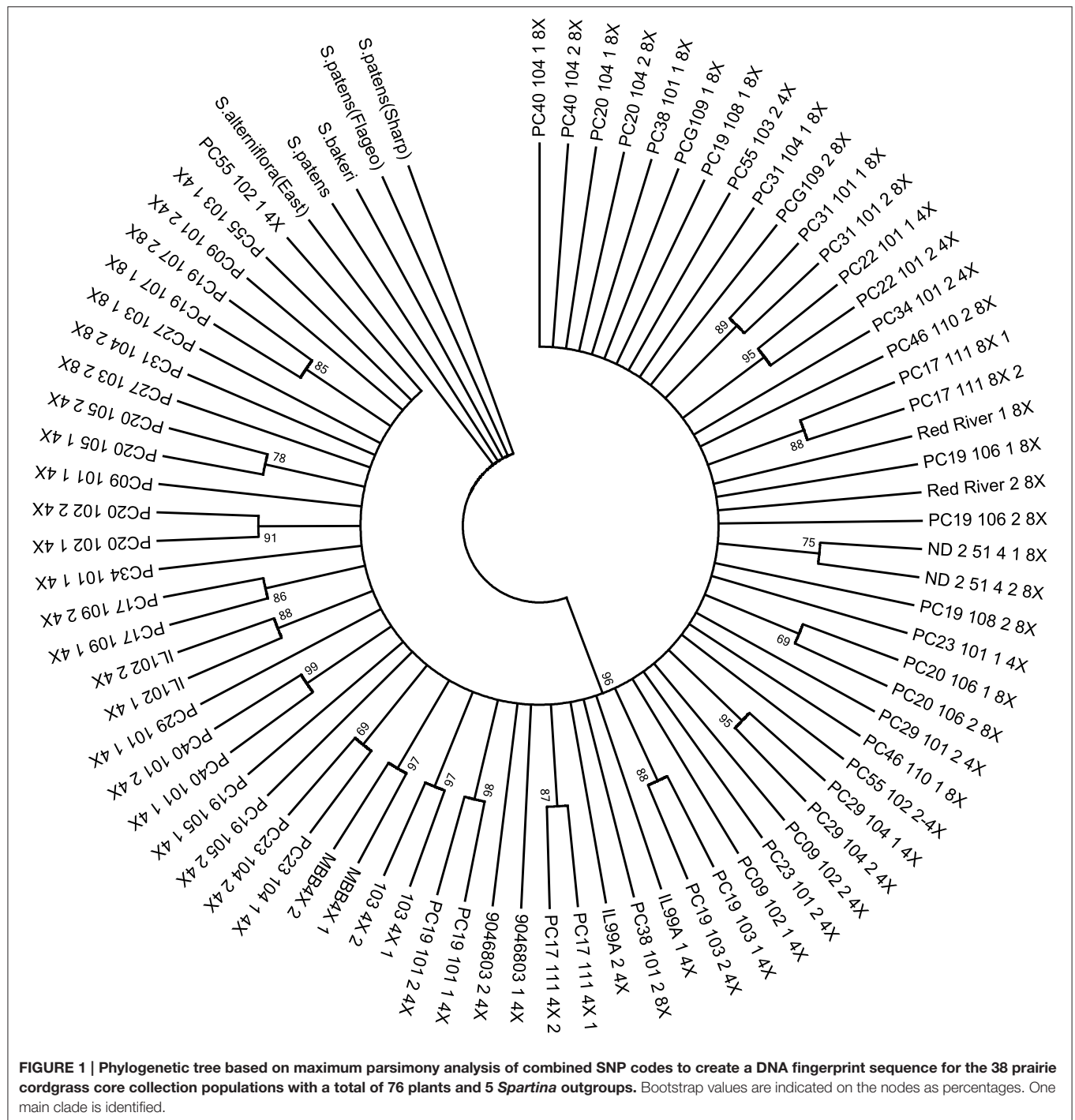
SNPs were successfully identified in nuclear transcriptomes of prairie cordgrass and validated as allelic variants that can be used in prairie cordgrass. SNP markers were used to detect significant polymorphisms in prairie cordgrass populations collected from distinct geographic regions in the U.S. These SNP polymorphisms appear to reflect genetic relationships in prairie cordgrass and, therefore, can be used to assess genetic diversity within and among populations in future studies.

The F<sub>1</sub> population, consisting of 83 plants, allows for the assessment of the fidelity of a specific tetraploid cross. Due to the lack of synchronization between the pollen and the ovaries, fewer seeds were obtained when PC20-12 was used as the female, compared with crosses involving PC17-109 as the female. Progeny that had SNP genotypes matching the female parent only were determined to be selfs. Of the F<sub>1</sub> progeny, 95.2% were identified to be hybrids. Prairie cordgrass is a protogynous outcrossing species (Gedye et al., 2012), leading to the possibility that later-maturing stigmas could have been exposed to pollen from the same female parent, resulting in 4.8% of the F<sub>1</sub> being selfs. The analysis of the 76 F<sub>2</sub> progeny obtained by selfing eight F<sub>1</sub> plants indicate that the SNPs, and the SNP markers chosen, could distinguish between a true selfed plant and an outcrossed plant. This is based on individuals

TABLE 4 | Sampling of DNA fingerprints created using SNP codes from 59 KASP genotyping assays.

Subject ID	pcg_00060																													
	pcg_00061	pcg_00062	pcg_00065	pcg_00066	pcg_00078	pcg_00081	pcg_00083	pcg_00084	pcg_00085	pcg_00088	pcg_00092	pcg_00093	pcg_00098	pcg_00101	pcg_00102	pcg_00103	pcg_00104	pcg_00106	pcg_00109	pcg_00110	pcg_00111	pcg_00112	pcg_00116	pcg_00142	pcg_00186	pcg_0013880	pcg_002412	pcg_0037652		
PC19_103.1	-	AA	AA	AA	TT	GA	GA	CT	GT	AG	GA	AG	TC	AG	AA	CG	AA	GG	GT	GA	AC	GA	CA	GG	CC	AG	TT	GG	CC	
	-	AA	AA	AA	TT	GA	GA	CT	GT	AG	GA	AG	TC	AG	AA	CG	AA	GG	GT	GA	AC	GA	CA	GG	CC	AG	TT	GG	CC	
	GA	-	-	-	TT	GA	GA	CT	GT	AG	GA	AG	TC	AG	TA	CG	TA	AG	GT	GA	AC	GA	CA	GG	CC	AG	TT	GG	CC	
	GA	-	-	-	TT	GA	GA	CT	GT	AG	GA	AG	TC	AG	-	CG	TA	AG	GT	GA	AC	GA	CA	GG	CC	AG	TT	GG	CC	
	GA	CA	-	-	CT	-	GA	CT	GT	AG	-	TC	AG	-	GG	TA	AG	TT	AA	CC	AC	AA	GA	CT	CC	GG	CC	GT	-	
	GA	AA	-	-	CT	-	GA	CT	GT	AG	-	TC	AG	-	GG	TA	AG	TT	AA	CC	AC	AA	GA	CT	CC	GG	CC	GT	-	
	GA	CA	GA	-	TT	GA	GA	-	GT	AG	-	CC	AG	GG	GA	GG	TA	AG	TT	AA	CC	AA	GA	CT	CC	GG	CC	GT	-	
	GA	-	-	-	TT	GA	GA	-	GT	AG	-	CC	AG	GG	GA	GG	TA	AG	TT	AA	CC	AA	GA	CT	CC	GG	CC	GT	-	
	Red_River_1	AA	CA	GA	CT	GA	GA	CT	GT	AG	GA	AG	TC	AG	AA	CG	TA	AG	GT	GA	AC	GA	CA	GG	CC	AG	TT	GG	CC	AT
	Red_River_2	AA	AA	AA	CT	GA	GA	CT	GT	AG	AA	AG	TC	AG	AA	CG	TA	AG	GT	GA	AC	GA	CA	GG	CC	AG	TT	GG	CC	AT
PC17_109	GA	CA	-	TT	GA	GA	CC	CC	GG	AA	AA	TT	AA	AA	CC	CC	TA	GG	GG	AA	GG	CC	GG	CC	AG	TT	GG	CC	TT	GG
	GA	AA	AA	TT	GG	GG	CC	GG	AA	AA	AA	TT	AA	AA	CC	CC	AA	GG	GG	AA	GG	CC	GG	CC	AG	TT	GG	CC	TT	GG
	AA	CC	AA	TT	AA	AA	CC	-	-	-	GG	CC	AA	AA	CC	CC	AA	GG	GG	AA	GG	CC	GG	CC	-	GG	CC	GG	AA	
	AA	CC	AA	TT	AA	AA	CC	-	-	-	GG	CC	AA	AA	CC	CC	TA	GG	GG	AA	GG	CC	GG	CC	-	GG	CC	GG	AA	
	AA	CC	AA	TT	AA	AA	CC	-	-	-	GG	CC	AA	AA	CC	CC	AA	GG	GG	AA	GG	CC	GG	CC	-	GG	CC	GG	AA	
	AA	CC	AA	TT	AA	AA	CC	-	-	-	GG	CC	AA	AA	CC	CC	AA	GG	GG	AA	GG	CC	GG	CC	-	GG	CC	GG	AA	
	AA	CC	AA	TT	AA	AA	CC	-	-	-	GG	CC	AA	AA	CC	CC	AA	GG	GG	AA	GG	CC	GG	CC	-	GG	CC	GG	AA	
	AA	CC	AA	TT	AA	AA	CC	-	-	-	GG	CC	AA	AA	CC	CC	AA	GG	GG	AA	GG	CC	GG	CC	-	GG	CC	GG	AA	
	AA	CC	AA	TT	AA	AA	CC	-	-	-	GG	CC	AA	AA	CC	CC	AA	GG	GG	AA	GG	CC	GG	CC	-	GG	CC	GG	AA	
	S. patens	AA	CC	AA	TT	AA	AA	CC	-	-	-	GG	CC	AA	AA	CC	CC	AA	GG	GG	AA	GG	CC	GG	CC	-	GG	CC	GG	AA

Subject ID	pcg_00060																													
	pcg_00061	pcg_00062	pcg_00065	pcg_00066	pcg_00078	pcg_00081	pcg_00083	pcg_00084	pcg_00085	pcg_00088	pcg_00092	pcg_00093	pcg_00098	pcg_00101	pcg_00102	pcg_00103	pcg_00104	pcg_00106	pcg_00109	pcg_00110	pcg_00111	pcg_00112	pcg_00116	pcg_00142	pcg_00186	pcg_0013880	pcg_002412	pcg_0037652		
PC19_103.1	GG	GG	GG	GG	AC	TT	CT	TA	-	AA	CC	GG	CC	GG	AG	TC	CC	CG	GG	GG	GG	CC	AG	GC	CT	GG	GG	CC	CC	
	GG	GG	GG	-	AC	TT	CT	TA	-	AA	CC	GG	CC	GG	AG	TC	CC	CG	GG	GG	GG	CC	AG	GC	CT	GG	GG	CC	CC	
	TG	GC	-	GC	AC	AT	TT	AA	AC	GA	CC	AG	-	GG	AG	CC	CC	-	GA	GT	CC	AA	AG	GC	CT	GG	GG	CC	CC	
	TG	GC	-	GC	AC	AT	TT	AA	AC	GA	CC	AG	-	GG	AG	CC	CC	-	GA	GT	CC	AA	AG	GC	CT	GG	GG	CC	CC	
	TG	GC	AG	GG	AC	TT	CT	TA	AA	-	CT	GG	CC	GG	AA	TC	CC	CC	GG	GG	GC	AA	AG	GC	CT	GG	GG	CC	CC	
	TG	-	AG	GC	-	TT	CT	TA	CC	-	CT	GG	AC	CC	AG	CC	CC	CC	GG	GG	-	AA	AG	-	TT	GG	GG	CC	CC	
	TG	GG	GG	GC	AC	TT	CT	TT	AC	GA	CT	AG	CC	CC	AG	TC	GG	GG	GA	-	CC	AA	AG	GC	CT	GG	GG	CC	CC	
	PC38_101.1	GG	GG	GG	GC	CC	TT	CC	TT	CC	GG	TT	GG	AA	-	AG	CC	CC	CC	GG	GG	GC	AA	AG	GC	CT	GG	GG	CC	CC
	Red_River_1	TG	GC	AG	GC	AC	AT	CT	TA	CC	GA	CT	AG	AC	AG	GG	-	GC	-	GG	GT	GC	AA	AG	GC	CT	GG	GG	CC	CC
	Red_River_2	TG	GC	-	GC	CC	AT	CT	TA	CC	GA	CT	AG	AC	AG	GG	CC	CC	-	GG	GT	GC	AA	AG	GC	CT	GG	GG	CC	CC
PC17_109	TG	GC	AG	GG	CC	TT	CC	TT	AC	GG	TT	GG	CC	AG	AG	TC	CC	CC	GG	GG	GC	AA	AG	GC	CT	GG	GG	CC	CC	
	TG	GC	GG	GC	CC	AT	CT	TA	CC	GA	CT	AG	CC	AG	AG	TC	CC	CC	GG	GG	GC	AA	AG	GC	CT	GG	GG	CC	CC	
	TG	GC	GG	GC	CC	TT	CC	TT	CC	GA	CT	AG	CC	AG	AG	TC	CC	CC	GG	GG	GC	AA	AG	GC	CT	GG	GG	CC	CC	
	TG	GC	GG	GC	CC	TT	CC	TT	CC	GA	CT	AG	CC	AG	AG	TC	CC	CC	GG	GG	GC	AA	AG	GC	CT	GG	GG	CC	CC	
	TG	GC	GG	GC	CC	TT	CC	TT	CC	GA	CT	AG	CC	AG	AG	TC	CC	CC	GG	GG	GC	AA	AG	GC	CT	GG	GG	CC	CC	
	TG	GC	GG	GC	CC	TT	CC	TT	CC	GA	CT	AG	CC	AG	AG	TC	CC	CC	GG	GG	GC	AA	AG	GC	CT	GG	GG	CC	CC	
	TG	GC	GG	GC	CC	TT	CC	TT	CC	GA	CT	AG	CC	AG	AG	TC	CC	CC	GG	GG	GC	AA	AG	GC	CT	GG	GG	CC	CC	
	TG	GC	GG	GC	CC	TT	CC	TT	CC	GA	CT	AG	CC	AG	AG	TC	CC	CC	GG	GG	GC	AA	AG	GC	CT	GG	GG	CC	CC	
	TG	GC	GG	GC	CC	TT	CC	TT	CC	GA	CT	AG	CC	AG	AG	TC	CC	CC	GG	GG	GC	AA	AG	GC	CT	GG	GG	CC	CC	
	S. alterniflora (East)	-	CC	GG	GC	CC	TT	TT	AA	CC	GG	TT	GG	AA	GG	GG	CC	-	CG	GG	TT	-	AA	AG	GC	CT	GG	GG	CC	CC
S. bakeri	TT	CC	GG	GC	CC	TT	TT	AA	CC	GG	TT	GG	AA	GG	GG	CC	-	CG	GG	TT	-	AA	AG	GC	CT	GG	GG	CC	CC	
	TT	CC	GG	GC	CC	TT	TT	AA	CC	GG	TT	GG	AA	GG	GG	CC	-	CG	GG	TT	CC	AA	AG	GC	CT	GG	GG	CC	CC	
	TT	CC	GG	GC	CC	TT	TT	AA	CC	GG	TT	GG	AA	GG	GG	CC	-	CG	GG	TT	CC	AA	AG	GC	CT	GG	GG	CC	CC	
	TT	CC	GG	GC	CC	TT	TT	AA	CC	GG	TT	GG	AA	GG	GG	CC	-	CG	GG	TT	CC	AA	AG	GC	CT	GG	GG	CC	CC	
	TT	CC	GG	GC	CC	TT	TT	AA	CC	GG	TT	GG	AA	GG	GG	CC	-	CG	GG	TT	CC	AA	AG	GC	CT	GG	GG	CC	CC	
	TT	CC	GG	GC	CC	TT	TT	AA	CC	GG	TT	GG	AA	GG	GG	CC	-	CG	GG	TT	CC	AA	AG	GC	CT	GG	GG	CC	CC	
	TT	CC	GG	GC	CC	TT	TT	AA	CC	GG	TT	GG	AA	GG	GG	CC	-	CG	GG	TT	CC	AA	AG	GC	CT	GG	GG	CC	CC	
	TT	CC	GG	GC	CC	TT	TT	AA	CC	GG	TT	GG	AA	GG	GG	CC	-	CG	GG	TT	CC	AA	AG	GC	CT	GG	GG	CC	CC	
	TT	CC	GG	GC	CC	TT	TT	AA	CC	GG	TT	GG	AA	GG	GG	CC	-	CG	GG	TT	CC	AA	AG	GC	CT	GG	GG	CC	CC	
	S. patens	TT	CC	GG	GC	CC	TT	TT	AA	CC	GG	TT	GG	AA	GG	GG	CC	-	CG	GG	TT	CC	AA	AG	GC	CT	GG	GG	CC	CC



consistently being genotyped as heterozygous (outcrossed) rather than being homozygous (selfed) as expected. Ninety-one percent of the  $F_2$  progeny were identified as successful selfs. Because of the protogynous nature of this species, there is already a natural element working against selfing. This could explain why outcrossed individuals were identified. There is also a possibility that some of the early-maturing stigmas were exposed to pollen in the greenhouse before bagging. This could explain

why more  $F_2$  progeny were identified as unexpected genotypes (outcrosses) than the expected genotype (selfs) of the  $F_1$  progeny.

There is evidence that the tetraploid cytotype is an allotetraploid that may follow a disomic inheritance pattern. Two divergent copies in the *Waxy* lineages of *Spartina* genus support the allotetraploid origin of *S. pectinata* (Fortune et al., 2007). The bivalent pairing that occurs during meiosis (Church,



**TABLE 5 | Summary of  $\chi^2$  analysis on F<sub>1</sub> progeny of a specific tetraploid prairie cordgrass cross with selfed data removed.**

Primer ID	Observed X Allele	Expected X Allele	Observed Y Allele	Expected Y Allele	Observed XY Allele	Expected XY Allele	$\chi^2$
<b>1:1 TEST</b>							
pcg_00011	0	0	44	39.5	35	39.5	1.025
pcg_00012	0	0	43	39	35	39	0.821
pcg_00024	43	39	0	0	35	39	0.821
pcg_00050	0	0	0	0	77	77	0
pcg_00058	0	0	0	0	77	77	0
pcg_00059	0	0	0	0	78	78	0
pcg_00106	0	0	0	0	75	75	0
pcg_1186	0	0	0	0	78	78	0
pcg_14142	0	0	0	0	79	79	0
<b>1:2:1 TEST</b>							
pcg_00061	27	19.8	15	19.8	37	39.5	3.962
pcg_00062	15	19.8	27	19.8	37	39.5	3.962
pcg_7965	17	19.8	20	19.8	42	39.5	0.544

Analysis indicates that all primers produce expected results from a monohybrid Mendelian cross.  $df = 2$ ,  $p = 0.05$ , critical  $\chi^2 = 5.991$ .

**TABLE 6 | SNP assay results for the F<sub>1</sub> progeny used to determine SNP codes that could indicate true selfs in F<sub>2</sub> progeny.**

Parent	pcg_00011	pcg_00012	pcg_00024	pcg_00050	pcg_00058	pcg_00059	pcg_00106	pcg_1186	pcg_14142	pcg_00061	pcg_00062	pcg_7965
13-F1008	GG*	GG*	AA*	TC	GT	CG	CG	CT	GC	TG	GC	AA*
13-F1011	GG*	GG*	AA*	TC	GT	CG	CG	CT	GC	GG*	GG*	AG
14-F1008	GA	GA	TA	TC	GT	CG	CG	CT	GC	TT*	CC*	AG
14-F1014	GG*	GG*	AA*	TC	GT	CG	CG	CT	GC	TG	GC	AG
14-F1015	GG*	GG*	AA*	TC	GT	CG	CG	CT	GC	TT*	CC*	AG
14-F1042	GG*	GG*	AA*	TC	GT	CG	CG	CT	GC	GG*	GG*	AG
14-F1067	GG*	GG*	AA*	TC	GT	CG	CG	CT	GC	TG	GC	GG*
14-F1071	GA	GA	TA	TC	GT	CG	CG	CT	GC	TG	GC	AA*

\*Indicates homozygous SNPs.

1940; Marchant, 1968a,b; Bishop, 2015) and the observation of disomic inheritance using genotyping-by-sequencing (Crawford, 2015) both suggest a disomic inheritance pattern in *S. pectinata*. This hypothesis was tested in a cross between two prairie cordgrass populations, exploiting the bi-allelic nature of the KASP technology to suggest Mendelian segregation ratios in a monohybrid type cross. The analysis of the F<sub>1</sub> hybrids and F<sub>2</sub> selfs conclude that disomic inheritance of SNPs in tetraploid prairie cordgrass is in agreement with the chromosomal and genomic evidence, and a possibility in this cross (Marchant, 1968a,b; Fortune et al., 2007; Bishop, 2015; Crawford, 2015).

The primary requirement of any breeding program is to ensure that accurate crosses are made (Glaszmann et al., 2010). The small flower size of prairie cordgrass and the large number of flowers per head make it hard to perform physical emasculation. Possibilities of self-pollination always exist and, therefore, developing a molecular way to confirm true crosses

from selfs is warranted (Fang et al., 2004; Gedye et al., 2012). In prairie cordgrass, SSR markers have been developed that identified successful crosses in this protogynous species without the need for emasculation. This study also confirms that hybrids of prairie cordgrass can be created and verified with molecular markers. However, utilizing SSRs can be time-consuming, limited in number, and more expensive than SNP markers, making a way for the introduction of these newly developed and validated KASP assays.

## CONCLUSION

This study reports the first research of SNP marker development for use in prairie cordgrass. SNP markers developed from the nuclear transcriptome were tested on a core collection of DNA and found to be polymorphic among and within populations. The amount of variation differs from previous findings based on chloroplast DNA, which identified the octoploid cytotype as the

**TABLE 7 | SNP assay results for F<sub>2</sub> individuals of two out of the eight selfed F<sub>1</sub> samples.**

	pcg_00011*	pcg_00012*	pcg_00024*	pcg_00050	pcg_00058	pcg_00059	pcg_00106	pcg_1186	pcg_14142	pcg_00061	pcg_00062	pcg_7965*
13-F1008 (Parent)	GG	GG	AA	TC	GT	CG	CG	CT	GC	TG	GC	AA
F2:2012_13_F1_008_1	GG	GG	AA	TC	GT	CG	GG	TT	GG	GG	GG	AA
F2:2012_13_F1_008_2	GG	GG	AA	TT	GT	CG	GG	TT	GG	TG	GC	AA
F2:2012_13_F1_008_3	GG	GG	AA	TC	GT	CG	CG	CT	GC	TG	GC	AA
F2:2012_13_F1_008_4	GG	GG	AA	TC	GG	CC	CG	CT	GC	TG	GC	AA
F2:2012_13_F1_008_5†	GA	GA	TA	TT	TT	GG	CG	CC	GG	TT	CC	AG
F2:2012_13_F1_008_6†	GA	GA	TA	TC	GT	CG	CC	CC	GC	TG	CC	AG
F2:2012_13_F1_008_7	GG	GG	AA	TC	TT	GG	CG	CC	GC	TT	CC	AA
F2:2012_13_F1_008_8	GG	GG	AA	CC	TT	GG	CG	TT	GC	GG	GG	AA
F2:2012_13_F1_008_9	GG	GG	AA	TC	GG	CC	GG	CT	GC	TG	GC	AA
	pcg_00011*	pcg_00012*	pcg_00024*	pcg_00050	pcg_00058	pcg_00059	pcg_00106	pcg_1186	pcg_14142	pcg_00061*	pcg_00062*	pcg_7965
13-F1011 (Parent)	GG	GG	AA	TC	GT	CG	CG	CT	GC	GG	GG	AG
F2:2012_13_F1_011_1	GG	GG	AA	TC	TT	GG	GG	CT	GC	GG	GG	AG
F2:2012_13_F1_011_3	GG	GG	AA	TT	GT	CG	CG	CT	GC	GG	GG	AG
F2:2012_13_F1_011_4	GG	GG	AA	CC	GT	CG	GG	CT	GC	GG	GG	AG
F2:2012_13_F1_011_5	GG	GG	AA	TT	GT	CG	CG	CT	GC	GG	GG	GG
F2:2012_13_F1_011_6	GG	GG	AA	CC	GT	CG	CG	CC	GC	GG	GG	AA
F2:2012_13_F1_011_7	GG	GG	AA	CC	GG	CC	GG	CC	GC	GG	GG	AG
F2:2012_13_F1_011_8	GG	GG	AA	TC	GT	CG	CG	CT	GC	GG	GG	AA
F2:2012_13_F1_011_9	GG	GG	AA	TC	GT	CG	CG	CT	GC	GG	GG	AG
F2:2012_13_F1_011_10	GG	GG	AA	CC	TT	GG	CG	CC	GC	GG	GG	AG
F2:2012_13_F1_011_2	GG	GG	AA	TC	GT	CG	CG	CT	GC	GG	GG	AA

Indicated are samples that can be identified as true selfs and as outcrossed.

\*Primers that can distinguish true selfs from outcrossed samples.

†F<sub>2</sub> individuals that are identified as outcrossed samples.

most variable. However, one must recognize these SNP markers cover a wide range of expressed genomic DNA vs. two non-coding chloroplast DNA regions, giving nucleic SNP markers an advantage in identifying random genetic variation. These markers were used to assess the validity of true crosses that were made between two different populations using F<sub>1</sub> and F<sub>2</sub> (selfs of F<sub>1</sub>) progeny. Utilizing the biallelic nature of the KASP system,  $\chi^2$  analysis of the F<sub>1</sub> samples suggests that tetraploid prairie cordgrass may follow Mendelian disomic inheritance although other modes of inheritance were not ruled out. This analysis provides insight into the genomic structure of this species, supporting the hypothesis that tetraploid prairie cordgrass is an allotetraploid. However, further analysis must be done on other cytotypes to completely understand the genome structure of this

species and to evaluate genetic diversity. In addition, this study underlines the usefulness of using SNP marker technology in future breeding programs of prairie cordgrass, and opens up the ability for the final step using SNP markers in genotyping germplasm collections or genetic/breeding populations of prairie cordgrass.

## ACKNOWLEDGMENTS

This work was funded by the Department of Crop Sciences and the Energy Biosciences Institute, the University of Illinois and supported by Brain Pool Program through the Korean Federation of Science and Technology Societies (KOFST) funded by the Ministry of Science, ICT and Future Planning (151S-4-3-1269).

## REFERENCES

- Anderson, E. K., Voigt, T. B., Kim, S. M., and Lee, D. K. (2015). Determining effects of sodicity and salinity on switchgrass and prairie cordgrass germination and plant growth. *Ind. Crops Prod.* 64, 79–87. doi: 10.1016/j.indcrop.2014.11.016
- Barkworth, M. E., Anderton, L. K., Capels, K. M., Long, S., and Piep, M. B. (eds.). (2007). *Manual of Grasses for North America*. Logan, UT: Utah State University Press.
- Bishop, J. (2015). *Chromosome Pairing and its Relationship to Fertility in Cytotypes of Prairie Cordgrass (Spartina pectinata Link)*. MS Thesis, University of Illinois. Available online at: <http://hdl.handle.net/2142/88316>
- Boe, A., and Lee, D. K. (2007). Genetic variation for biomass production in prairie cordgrass and switchgrass. *Crop Sci.* 47, 929–934. doi: 10.2135/cropsci2006.05.0323
- Boe, A., Owens, V., Gonzalez–Hernandez, J., Stein, J., Lee, D. K., and Koo, B. C. (2009). Morphology and biomass production of prairie cordgrass on marginal lands. *GCB Bioenerg.* 1, 240–250. doi: 10.1111/j.1757-1707.2009.01018.x
- Boe, A., Springer, T., Lee, D. K., Rayburn, A. L., and Gonzalez-Hernandez, J. (2013). “Underutilized grasses,” in *Bioenergy Feedstocks: Breeding and Genetics*, eds M. C. Saha, H. S. Bhandari, and J. H. Bouton (Oxford, UK: John Wiley & Sons, Inc.), 173–205.
- Bundock, P. C., Cross, M. J., Shapter, F. M., and Henry, R. J. (2006). Robust allele-specific polymerase chain reaction markers developed for single nucleotide polymorphisms in expressed barley sequences. *Theor. Appl. Genet.* 112, 358–365. doi: 10.1007/s00122-005-0137-6
- Byers, R. L., Harker, D. B., Yourstone, S. M., Maughan, P. J., and Udall, J. A. (2012). Development and mapping of SNP assays in allotetraploid cotton. *Theor. Appl. Genet.* 124, 1201–1214. doi: 10.1007/s00122-011-1780-8
- Church, G. L. (1940). Cytotaxonomic studies in the gramineae *Spartina*, *Andropogon* and *Panicum*. *Am. J. Bot.* 27, 263–271. doi: 10.2307/2436892
- Cockram, J., Jones, H., Norris, C., and O’Sullivan, D. M. (2012). Evaluation of diagnostic molecular markers for DUS phenotypic assessment in the cereal crop, barley (*Hordeum vulgare* ssp. *vulgare* L.). *Theor. Appl. Genet.* 125, 1735–1749. doi: 10.1007/s00122-012-1950-3
- Crawford, J. (2015). *Linkage Mapping in Prairie Cordgrass (Spartina pectinata Link) Using Genotyping-by-Sequencing Markers*. MS thesis, University of Illinois. Available online at: <http://hdl.handle.net/2142/88128>
- Cuenca, J., Aleza, P., Navarro, L., and Ollitrault, P. (2013). Assignment of SNP allelic configuration in polyploids using competitive allelespecific PCR: application to citrus triploid progeny. *Ann. Bot.* 111, 731–742. doi: 10.1093/aob/mct032
- Fang, X., Subudhi, P. K., Venuto, B. C., and Harrison, S. A. (2004). Mode of pollination, pollen germination, and seed set in smooth cordgrass (*Spartina alterniflora*, Poaceae). *Int. J. Plant Sci.* 165, 395–401. doi: 10.1086/382810
- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791. doi: 10.2307/2408678
- Fortune, P. M., Schierenbeck, K. A., Ainouche, A. K., Jacquemin, J., Wendel, J. F., and Ainouche, M. L. (2007). Evolutionary dynamics of Waxy and the origin of hexaploid *Spartina* species (Poaceae). *Mol. Phylogenet. Evol.* 43, 1040–1055. doi: 10.1016/j.ympev.2006.11.018
- Gedye, K., Gonzalez-Hernandez, J., Ban, Y., Ge, X., Thimmapuram, J., Sun, F., et al. (2010). Investigation of the transcriptome of prairie cord grass, a new cellulosic biomass crop. *Plant Genome* 3, 69–80. doi: 10.3835/plantgenome2010.06.0012
- Gedye, K. R., Gonzalez-Hernandez, J. L., Owens, V., and Boe, A. (2012). Advances towards a marker-assisted selection breeding program in prairie cordgrass, a biomass crop. *Int. J. Plant Genomics* 2012:313545. doi: 10.1155/2012/313545
- Glaszmann, J. C., Kilian, B., Upadhyaya, H. D., and Varshney, R. K. (2010). Accessing genetic diversity for crop improvement. *Curr. Opin. Plant Bio.* 13, 167–173. doi: 10.1016/j.pbi.2010.01.004
- Gonzalez-Hernandez, J. L., Sarath, G., Stein, J. M., Owens, V., Gedye, K., and Boe, A. (2009). A multiple species approach to biomass production from native herbaceous perennial feedstocks. *In Vitro Cell. Dev. Biol.-Plant* 45, 267–281. doi: 10.1007/s11627-009-9215-9
- Graves, H., Rayburn, A. L., Kim, S., and Lee, D. K. (2015). Chloroplast DNA variation within prairie cordgrass (*Spartina pectinata*) populations in the U.S. *J. Syst. Evol.* doi: 10.1111/jse.12168. [Epub ahead of print].
- Guo, J., Thapa, S., Voigt, T., Rayburn, A. L., Boe, A., and Lee, D. K. (2015). Phenotypic and biomass yield variations in natural populations of prairie cordgrass (*Spartina pectinata* Link) in the USA. *BioEnergy. Res.* 8, 1371–1383. doi: 10.1007/s12155-015-9604-3
- Hirakawa, H., Shirasawa, K., Kosugi, S., Tashiro, K., Nakayama, S., Yamada, M., et al. (2014). Dissection of the octoploid strawberry genome by deep sequencing of the genomes of *Fragaria* species. *DNA Res.* 21, 169–181. doi: 10.1093/dnares/dst049
- Hitchcock, A. S. (1950). “*Spartina* Scherb. Cordgrass” in *Manual of the Grasses of the United States, 2nd Edn. Rev.*, ed A. Chase (Mineola: Dover Publications, Inc.), 508–514.
- Kim, S., Rayburn, A. L., Boe, A., and Lee, D. K. (2012). Neopolyploidy in *Spartina pectinata* Link: 1. morphological analysis of tetraploid and hexaploid plants in a mixed natural population. *Plant Syst. Evol.* 298, 1073–1083. doi: 10.1007/s00606-012-0617-5
- Kim, S., Rayburn, A. L., and Lee, D. K. (2010). Genome size and chromosome analyses in prairie cordgrass. *Crop Sci.* 50, 2277–2282. doi: 10.2135/cropsci2010.03.0140
- Kim, S., Rayburn, A. L., Voigt, T. B., Ainouche, M. L., Ainouche, A. K., and Lee, D. K. (2013). Chloroplast DNA intraspecific phylogeography of prairie cordgrass (*Spartina pectinata* bosc ex link). *Plant Mol. Bio. Rep.* 31, 1376–1383. doi: 10.1007/s11105-013-0619-7
- Kim, S., Rayburn, A. L., Voigt, T. B., Parrish, A., and Lee, D. K. (2011). Salinity effects on germination and plant growth of prairie cordgrass and switchgrass. *Bioenerg. Res.* 5, 225–235. doi: 10.1007/s12155-011-9145-3
- LGC Genomics. (2014). *Organisms Genotyped Using KASP* [online]. Available online at: <http://www.lgcgroup.com/LGCGroup/media/PDFs/Products/Genotyping/Species-tested-with-KASP-technology.pdf?ext=.pdf> (Accessed April 28, 2015).
- Marchant, C. J. (1968a). Evolution in *Spartina* (Gramineae). II. chromosomes, basic to relationships and the problem of *S. x townsendii* agg. *J. Linn. Soc. Bot.* 60, 381–409. doi: 10.1111/j.1095-8339.1968.tb00096.x
- Marchant, C. J. (1968b). Evolution in *Spartina* (Gramineae). III. species chromosome numbers and their taxonomic significance. *J. Linn. Soc. Bot.* 60, 411–417. doi: 10.1111/j.1095-8339.1968.tb00097.x
- Mikkilineni, V. (1997). *Restriction Fragment Length Polymorphism Analysis of the Illinois Long-Term Selection Chemical Strains*. *Crop Sci.* MS thesis, University of Illinois, Urbana, 108–113.
- Moncada, K. M., Ehlike, N. J., Muehlbauer, G. J., Sheaffer, C. C., Wyse, D. L., and DeHaan, L. R. (2007). Genetic variation in three native plant species across the state of Minnesota. *Crop Sci.* 47, 2379–2389. doi: 10.2135/cropsci2007.02.0082
- Montemayor, M. B., Price, J. S., Rochefort, L., and Boudreau, S. (2008). Temporal variations and spatial patterns in saline and waterlogged peat fields: 1. Survival and growth of salt marsh graminoids. *Environ. Exp. Bot.* 62, 333–342. doi: 10.1016/j.envexpbot.2007.10.004
- Nei, M., and Kumar, S. (2000). *Molecular Evolution and Phylogenetics*. New York, NY: Oxford University Press.
- Paux, E., Sourdil, P., Mackay, I., and Feuillet, C. (2012). Sequence-based marker development in wheat: advances and applications to breeding. *Biotechnol. Adv.* 30, 1071–1088. doi: 10.1016/j.biotechadv.2011.09.015
- Rafalski, A. (2002). Applications of single nucleotide polymorphisms in crop genetics. *Curr. Opin. Plant Biol.* 5, 94–100. doi: 10.1016/S1369-5266(02)00240-6
- Salem, M., Vallejo, R. L., Leeds, T. D., Palti, Y., Liu, S., Sabbagh, A., et al. (2012). RNA-Seq identifies SNP markers for growth traits in rainbow trout. *PLoS ONE* 7:e36264. doi: 10.1371/journal.pone.0036264
- Saxena, R. K., Penmetsa, R. V., Upadhyaya, H. D., Kumar, A., Carrasquilla-Garcia, N., Schlueter, J. A., et al. (2012). Large-scale development of cost-effective single-nucleotide polymorphism marker assays for genetic mapping in pigeonpea and comparative mapping in legumes. *DNA Res.* 19, 449–461. doi: 10.1093/dnares/dss025
- Semagn, K., Babu, R., Hearne, S., and Olsen, M. (2014). Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. *Mol. Breed.* 33, 1–14. doi: 10.1007/s11032-013-9917-x
- Sindhu, A., Ramsay, L., Sanderson, L. A., Stonehouse, R., Li, R., Condie, J., et al. (2014). Gene-based SNP discovery and genetic mapping in pea. *Theor. Appl. Genet.* 127, 2225–2241. doi: 10.1007/s00122-014-2375-y

- Soltis, D. E., Soltis, P. S., and Milligan, B. G. (1992). "Intraspecific chloroplast DNA variation: systematic and phylogenetic implications," in *Molecular Systematics of Plants*, ed P. S. Soltis (New York, NY: Chapman & Hall), 117–150.
- Tamura, K., Stecher, G., Peterson, D., Filipowski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Bio. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Uitdewilligen, J. L., Wolters, A. A., D'Hoop, B. B., Borm, T. A., Visser, R. F., and van Eck, H. J. (2013). A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. *PLoS ONE* 8:e62355. doi: 10.1371/journal.pone.0062355
- Voight, J., and Mohlenbrock, R. (1979). *Prairie Plants of Illinois*. Springfield, IL: Illinois Department of Conservation Division of Forestry.
- Wei, L., Miao, H., Li, C., Duan, Y., Niu, J., Zhang, T., et al. (2014). Development of SNP and InDel markers via *de novo* transcriptome assembly in *Sesamum indicum* L. *Mol. Breed.* 34, 2205–2217. doi: 10.1007/s11032-014-0174-4
- Wendel, J., and Doyle, J. (2005). "Polyploidy and evolution in plants," in *Plant Diversity and Evolution Genotypic and Phenotypic Variation in Higher Plants*, ed R. J. Henry (Cambridge, MA: CABI Publishing), 97–118.
- Zilverberg, C., Johnson, W. C., Archer, D., Kronberg, S., Schumacher, T., Boe, A., et al. (2014). Profitable prairie restoration: the EcoSun Prairie Farm experiment. *J. Soil Water Conserv.* 69, 22A–25A. doi: 10.2489/jswc.69.1.22a

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Graves, Rayburn, Gonzalez-Hernandez, Nah, Kim and Lee. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Implementation of Genomic Prediction in *Lolium perenne* (L.) Breeding Populations

Nastasiya F. Grinberg<sup>1</sup>, Alan Lovatt<sup>2</sup>, Matt Hegarty<sup>2</sup>, Andi Lovatt<sup>2</sup>, Kirsten P. Skot<sup>2</sup>, Rhys Kelly<sup>2</sup>, Tina Blackmore<sup>2</sup>, Danny Thorogood<sup>2</sup>, Ross D. King<sup>1</sup>, Ian Armstead<sup>2</sup>, Wayne Powell<sup>2,3</sup> and Leif Skot<sup>2\*</sup>

<sup>1</sup> Manchester Institute of Biotechnology, University of Manchester, Manchester, UK, <sup>2</sup> Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Aberystwyth, UK, <sup>3</sup> CGIAR Consortium, CGIAR Consortium Office, Montpellier, France

## OPEN ACCESS

### Edited by:

Keenan Amundsen,  
University of Nebraska-Lincoln, USA

### Reviewed by:

Hao Peng,  
Washington State University, USA  
Scott Eric Warnke,  
United States Department  
of Agriculture–Agricultural Research  
Service, USA

### \*Correspondence:

Leif Skot  
lfs@aber.ac.uk

### Specialty section:

This article was submitted to  
Crop Science and Horticulture,  
a section of the journal  
Frontiers in Plant Science

**Received:** 11 September 2015

**Accepted:** 25 January 2016

**Published:** 12 February 2016

### Citation:

Grinberg NF, Lovatt A, Hegarty M,  
Lovatt A, Skot KP, Kelly R,  
Blackmore T, Thorogood D, King RD,  
Armstead I, Powell W and Skot L  
(2016) Implementation of Genomic  
Prediction in *Lolium perenne* (L.)  
Breeding Populations.  
Front. Plant Sci. 7:133.  
doi: 10.3389/fpls.2016.00133

Perennial ryegrass (*Lolium perenne* L.) is one of the most widely grown forage grasses in temperate agriculture. In order to maintain and increase its usage as forage in livestock agriculture, there is a continued need for improvement in biomass yield, quality, disease resistance, and seed yield. Genetic gain for traits such as biomass yield has been relatively modest. This has been attributed to its long breeding cycle, and the necessity to use population based breeding methods. Thanks to recent advances in genotyping techniques there is increasing interest in genomic selection from which genomically estimated breeding values are derived. In this paper we compare the classical RRBLUP model with state-of-the-art machine learning techniques that should yield themselves easily to use in GS and demonstrate their application to predicting quantitative traits in a breeding population of *L. perenne*. Prediction accuracies varied from 0 to 0.59 depending on trait, prediction model and composition of the training population. The BLUP model produced the highest prediction accuracies for most traits and training populations. Forage quality traits had the highest accuracies compared to yield related traits. There appeared to be no clear pattern to the effect of the training population composition on the prediction accuracies. The heritability of the forage quality traits was generally higher than for the yield related traits, and could partly explain the difference in accuracy. Some population structure was evident in the breeding populations, and probably contributed to the varying effects of training population on the predictions. The average linkage disequilibrium between adjacent markers ranged from 0.121 to 0.215. Higher marker density and larger training population closely related with the test population are likely to improve the prediction accuracy.

**Keywords:** perennial ryegrass, genomic selection, BLUP, machine learning, forage crop

## INTRODUCTION

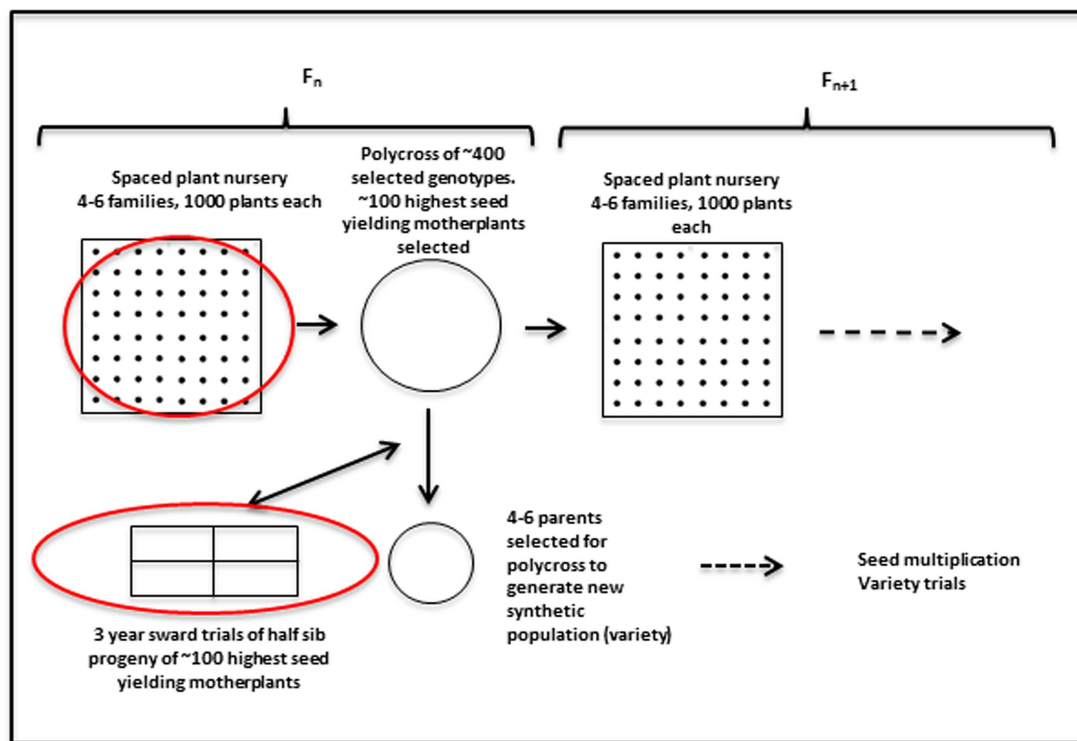
Genetic improvement of crops involves the selection of plants with superior characteristics in terms of traits that are considered important. This could be yield (biomass or seed), resistance to diseases and pests and better tolerance to abiotic stress. The selection criteria have been and still are based largely on phenotypic performance. Phenotypic assessment can be time consuming and laborious, particularly for perennial crops. There is pressure to increase

agricultural output at a faster rate to keep up with population growth and reduced area available for agricultural production. Molecular marker assisted selection (MAS) represents a way of potentially reducing the time and effort needed for phenotypic testing (Lande and Thompson, 1990; Dekkers and Hospital, 2002; Xu and Crouch, 2008). The success of MAS is dependent upon sufficient linkage disequilibrium (LD) between a marker and the phenotypic QTL (quantitative trait locus), and the QTL explaining a substantial proportion of the variation for the trait. Often, this is not the case, and the association between marker and QTL is not significant, and thus discarded. Therefore, the use of MAS in plant breeding has not been widespread. Recent improvements in genotyping techniques have made it possible to cover the genome with densely populated molecular markers, and this has paved the way for genome wide association studies (GWASs) (Rafalski, 2002; Flint-Garcia et al., 2003) in which marker-trait associations can be identified in breeder relevant and more diverse populations, rather than bi-parental mapping populations. The disadvantages of this approach includes low statistical power from small population sizes, confounding population structure of the germplasm used, and overestimation of the effect of few significant marker associations with QTL (Heffner et al., 2009).

Genomic selection (GS) represents a way of dealing with many of the problems of current MAS methodology. The term was first used by Meuwissen et al. (2001) to describe the use of genome wide molecular markers to simultaneously estimate the effect of all markers across the genome, irrespective of whether they are significant, in order to calculate a genomically estimated breeding value (GEBV) of selection candidates. GS depends upon the establishment of a training population, for which both phenotypic and genotypic data are available. The marker effects calculated from these data can be used to estimate the breeding values in populations with only genotypic data available (Meuwissen et al., 2001; Heffner et al., 2009). In terms of prediction methods the most widely used are the genomic or ridge regression. BLUP (best linear unbiased prediction) and other penalized regression methods (Gianola et al., 2006; de los Campos et al., 2009; Li and Sillanpää, 2012) and various Bayesian techniques (Meuwissen et al., 2001; de los Campos et al., 2009; Habier et al., 2011). However these techniques do not explicitly account for interactions. There is currently considerable interest in applying machine learning (ML) to science, and reviews have recently appeared (Ghahramani, 2015; Jordan and Mitchell, 2015). These methods are increasingly being applied in GWASs and GS (Dudoit et al., 2002; Long et al., 2007; Ziegler et al., 2007; Szymczak et al., 2009; Ogutu et al., 2011, 2012; Ornella et al., 2014; Spindel et al., 2015). ML algorithms are well suited to application in plant-breeding datasets. Most are easy to use and are easily available in a variety of implementations. Many methods perform attribute selection (e.g., lasso, regression trees) or assign importance scores to variables (e.g., random forest, boosted trees). Some methods, such as tree based approaches, do not require any assumptions about the underlying trait (e.g., additivity of effects, the numbers and size of interactions, depth of interactions etc.) and are able to capture complex non-linear relationships between response and regressors.

Genomic selection is an attractive alternative to classic selection methods, and it has been adopted in animal breeding, particularly dairy cattle (Schaeffer, 2006; Pryce and Daetwyler, 2012; Hayes et al., 2013b). The uptake of GS has been slower in plant breeding, but is now gathering pace. Many papers have assessed the potential use of GS in simulation and empirical studies of some of the major crops (Bernardo and Yu, 2007; Heffner et al., 2009, 2010, 2011; Piepho, 2009; Zhong et al., 2009; Jannink et al., 2010; Albrecht et al., 2011; Poland et al., 2012; Zhao et al., 2012, 2013; Xu, 2013; Bentley et al., 2014; Jarquin et al., 2014; Wang et al., 2014; Xu et al., 2014). The applicability of GS in perennial crops such as trees and forages is even more appealing, due to the possibility of significantly reducing the length of the breeding cycle (Grattapaglia and Resende, 2011). Some empirical studies in trees suggest reasonable prediction accuracies can be obtained (Resende et al., 2012a,b; Zapata-Valenzuela et al., 2012; Beaulieu et al., 2014). Two factors need to be taken into consideration when dealing with breeding in many forage crops such as perennial ryegrass. Firstly, the performance of individual spaced plants generally does not correlate well with the phenotype in sward for many economically important traits (Casler and Brummer, 2008). Secondly, most of the important forage crops are outbreeding, so variety development is usually based on population improvement via recurrent selection schemes (Posselt, 2010; Conaghan and Casler, 2011). These factors probably contribute to the low genetic gains achieved in forages, but other factors have been suggested, including a lack of a harvest index trait to breed for, inability to exploit heterosis and a large number of target traits with no or negative correlation between them (Casler and Brummer, 2008). Two recent reviews have assessed the prospects for GS in perennial forage crops such as grasses and legumes (Hayes et al., 2013a; Resende et al., 2014). The latter concluded that GS is likely to be most beneficial when phenotypic values of spaced plants do not correlate with those in sward, when within-family selection is difficult or impossible, and for traits that can be assessed only after several years of plot trials. Hayes et al. (2013a) also suggested that significant modifications to most current mass selection breeding schemes in, e.g., perennial ryegrass would be desirable/necessary to implement GS effectively.

However, there is very little empirical data available from forage crops with evaluation of GS performance. Lipka et al. (2014) described the use of GS in predicting breeding values in switchgrass (*Panicum virgatum* L.), a perennial grass which is being developed as an energy crop. They obtained cross validation accuracies of up to 0.52. Slavov et al. (2014) reported prediction accuracies varying between 0.05 (dry matter) and 0.95 (moisture) with an average of 0.57 for 17 traits in the energy grass, *Miscanthus sinensis*. Both used association panels as the training and validation population. Recently, an empirical study of genomic prediction of biomass yield in tetraploid alfalfa reported prediction accuracies between 0.21 and 0.60 depending on the breeding cycle (Li et al., 2015). The authors concluded that the selection efficiencies per unit time based on GS were better than for phenotypic selection. To our knowledge, no empirical data have been published of GS performance in perennial ryegrass, the most important forage crops in temperate



**FIGURE 1 | Outline of the recurrent selection breeding program in *L. perenne*.** The red circles indicate stages where genomic selection (GS) can be implemented, given availability of genotypic, and phenotypic data. The work described here is aimed at utilizing the phenotypic data from the sward trials to facilitate prediction of the best mother-plants for new variety production.

grassland agriculture. Here we report our first results of an evaluation of GS in the populations from a long standing and successful recurrent selection breeding program at the Institute of Biological, Environmental, and Rural Sciences (IBERS). The current populations were established in the late 1980's from a relatively small founder population, and have now been through up to 14 generations of selection and recombination. We have used current and some historical phenotypic data from plot trials of half sib progeny of mother-plants in combination with genotypic data from the mother-plants. Higher prediction accuracies were obtained for traits related to forage quality, particularly water soluble carbohydrates (WSCs) and digestibility (DMD) than for biomass yield. For most trait-training population combinations the ridge regression BLUP prediction method outperformed the three ML methods employed here. We discuss possible explanations for the results as well as potential ways of improving prediction accuracies particularly for biomass yield.

## MATERIALS AND METHODS

### Plant Material and Breeding Cycle

Plant material from the perennial ryegrass breeding populations was used to obtain genotypic and phenotypic data. In order to put the data collection into context, a brief description of the breeding cycle is given. It is also illustrated in **Figure 1**. Any

given cycle starts with a polycross of about 400–600 plants from four to six families. Those parents have been collected from spaced plant field plots. Approximately 100 of the highest seed yielding mother-plants are selected to provide half sib progeny for evaluation in sward plot trials. Four replicate plots of the half sib progeny are evaluated over three growing seasons. Biomass yield was recorded for seven cuts each year for the first 2 years, and material from cuts 4 and 5 in the 1st year was used to obtain estimates of dry matter digestibility (DMD), WSCs and nitrogen, with near infrared reflectance spectroscopy (NIRS) (Lister and Dhanoa, 1998). The mean of results from those two cuts were used in the present analysis. At several stages during all three growing seasons persistency was assessed by scoring ground cover visually on a scale of 0–9. In the breeding program the phenotypic data are used to select three–five parents from the mother-plants for poly crossing to obtain a synthetic population for variety trials. The results are also used to inform the selection of 3–6 half-sib families for each new generation. Around 400–600 genotypes from the spaced plant trials of 1000 plants from each family are selected for poly crossing. However, other factors, such as plant stature, disease resistance, and winter survival are also taken into consideration in this selection.

Broad sense heritabilities were calculated as follows:

$$H_B^2 = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_E^2} \quad (1)$$

where  $\sigma_G^2$  is the genetic variance, and  $\sigma_E^2$  is the residual error variance. The variance components were obtained from a one-way analysis of variance of each of the traits separately. The standard deviation was obtained via leave-one-out Jackknife analysis.

## Genotyping and Linkage Disequilibrium

A 3K Illumina Infinium iSelect Array was used for genotyping of the mother-plants. The SNPs in the array were identified on the basis of polymorphisms in transcriptome libraries from perennial ryegrass plants representing six diverse populations. The development and validation of this array was described in detail previously (Blackmore et al., 2015). The DNA was extracted from leaf material of the mother-plants from each generation as described (Skøt et al., 2011), except for the F12 generation. None of the mother-plants from that generation are in existence, so the DNA was obtained from the husks of the seed derived from the respective mother-plants. In total, DNA samples of sufficient quality were obtained from 86 mother-plants of the F12 generation. After allele calling in the Illumina GenomeStudio software, the genotypic scores were converted to  $-1$ ,  $0$ , and  $1$  for input into the various prediction models.

Linkage disequilibrium data ( $r^2$ ) were obtained using a consensus genetic map containing 1670 markers from the 3 K Infinium Array as described in Blackmore et al. (2015). The LD landscape plots were generated based on an R script described earlier (Wang et al., 2013), but modified and improved for *L. perenne*.

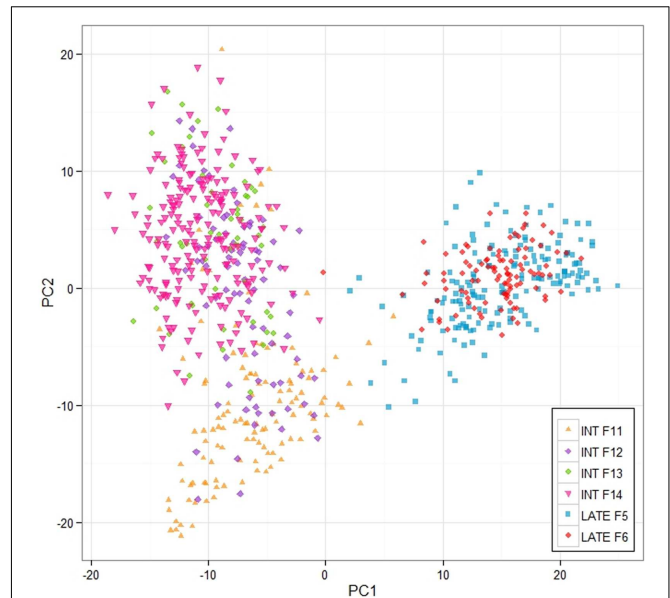
## Training and Test Populations

This work was aimed at making genomic predictions of the breeding values of the 100 mother plants of the F14 generation based on training populations consisting of various parts of the previous generations of both the intermediate- and late flowering breeding populations. We wanted to assess the effect of training population size and relatedness to F14 on prediction ability, and also to compare a number of different prediction models in terms of their performance. Three training populations were used. The first was based on the F13 generation, which is closest genetically to F14 (see Figure 2). It consisted of 54 mother plants. The second included data from all the intermediate-flowering generations for which we have genotypic and phenotypic data, namely F11, F12, and F13 (this training population is referred to as 'INT'). The size of that training population was 259. Finally, we also included the late-flowering population F5. This brought the training population size up to 364 (we refer to this training population as 'ALL').

All phenotypic data were normalized with respect to each sub-population's mean and scaled to have variance 1. Thus, hybrid phenotypes F11 + F12 + F13 and F5 + F11 + F12 + F13 do not have variance of exactly 1.

## Prediction Models

We investigated predictive abilities of four methods: GBLUP from statistical genomics and three ML methods. The advantage of GBLUP compared to standard multivariate regression is the ability to cope with the  $p \gg n$  situation and prevent overfitting



**FIGURE 2 | Principle components analysis of the breeding populations.** The analysis and the plot was based on the 3 K Infinium Array SNP data, and the analysis was implemented in R.

via the penalty mechanism. We use GBLUP as the benchmark method against which we compared the three ML models.

We used two tree-based methods: random forests (RF) (Breiman, 2001) and boosted trees GBM (Friedman, 2001). Both methods are non-parametric and make no assumptions about the distribution or any other properties of the data they are applied to, which is an advantage.

For RF we have used the standard values for the number of variables considered at each split ( $1/3$  of the total number), a minimum of five observations per terminal node; trees were grown to their maximal depth and were not pruned and we have grown 500 trees per forest.

For GBM we have used a shrinkage parameter (which discounts each successive tree to avoid overfitting) of 0.01, subsampling rate (proportion of data used to construct each tree) of 0.5 and trees of depth 5, of which have grown 1500 per model.

Thirdly, we used k-nearest neighbors algorithm (KNN) – a model that predicts each new sample point based on the values of its nearest (according to some metric) neighbors in the training set. In KNN regression this prediction is just the average over the values in the neighborhood. This is an example of a *lazy learning* method – generalization beyond training data only occurs when test data is introduced. The advantage of the method is its simplicity and ease of use (one effectively only has one tuning parameter,  $k$ , the number of neighbors to consider for each new instance) and in the context of GS – the fact that genetic relatedness of plants in the training and test populations is exploited as only plants genetically close to the target are used to calculate each GEBV. For each trait we used the optimal number of neighbors chosen via cross-validation on the corresponding training population (between 1 and 10 for the F13 training



**TABLE 1 | Broad-sense heritability for different generations and traits of the breeding populations.**

Trait/Population	F5	F11	F12	F13	F14
totaly7c_yr1	0.06 (0.04)	0.22 (0.05)	0.16 (0.06)	0.40 (0.08)	0.19 (0.05)
totaly7c_yr2	0.23 (0.05)	0.02 (0.04)	0.03 (0.04)	0.34 (0.08)	0.08 (0.04)
conscuty_yr1	0.12 (0.05)	0.19 (0.05)	0.16 (0.06)	0.17 (0.06)	0.20 (0.05)
conscuty_yr2	0.27 (0.05)	0.21 (0.05)	0.08 (0.05)	0.17 (0.07)	0.11 (0.05)
vegylid_yr1	0.11 (0.05)	0.46 (0.05)	0.07 (0.05)	0.44 (0.08)	0.18 (0.05)
vegylid_yr2	0.20 (0.06)	0.03 (0.04)	0.05 (0.05)	0.34 (0.07)	0.11 (0.05)
gcscore_yr1	0	0.22 (0.05)	0.30 (0.06)	0.19 (0.06)	0.19 (0.04)
gcscore_yr2	0.27 (0.04)	0.33 (0.05)	0.34 (0.06)	0.12 (0.07)	0.26 (0.05)
dmd	0.17 (0.05)	0.40 (0.05)	0.08 (0.05)	0.59 (0.06)	0.52 (0.05)
n	0.24 (0.04)	0.37 (0.05)	0.03 (0.04)	0.35 (0.08)	0.23 (0.06)
wsc	0.35 (0.06)	0.37 (0.05)	0.22 (0.06)	0.41 (0.08)	0.42 (0.05)

Standard deviations are in brackets. Trait identification: totaly7c, total biomass yield over all 7 cuts; conscuty, Yield of conservation cut, i.e., second cut; vegylid, Total biomass yield minus conservation cut; gcscore, Ground cover score; dmd, Dry matter digestibility (%); n, nitrogen (%); wsc, Water soluble carbohydrates (%).

population, between 3 and 20 for INT and between 4 and 26 for ALL).

Performance of each model was assessed by calculating Spearman's rank correlation ( $r(y, \text{GEBV})$ ) between the corresponding predicted values and the observed F14 phenotypic values.

All analysis was done in R (R Core Development Team, 2014); we used the `gbm` package for GBM, `randomForest` for RF, `FNN` for KNN (Hastie et al., 2009) and `rrBLUP` (Endelman, 2011) for BLUP.

## RESULTS

### Phenotypic Data and Heritabilities

The phenotypic data were obtained from sward trials derived from half-sib progeny of the 100 or so mother-plants of each generation. The quality traits, such as digestibility, WSCs and nitrogen tended to have higher heritability than the yield-related traits (Table 1). There is also variation between years and cuts, highlighting the effects of time. The heritabilities for the biomass yields in the 2nd year tended to be lower than for the 1st year, particularly for F14, but also for the other Intermediate generations.

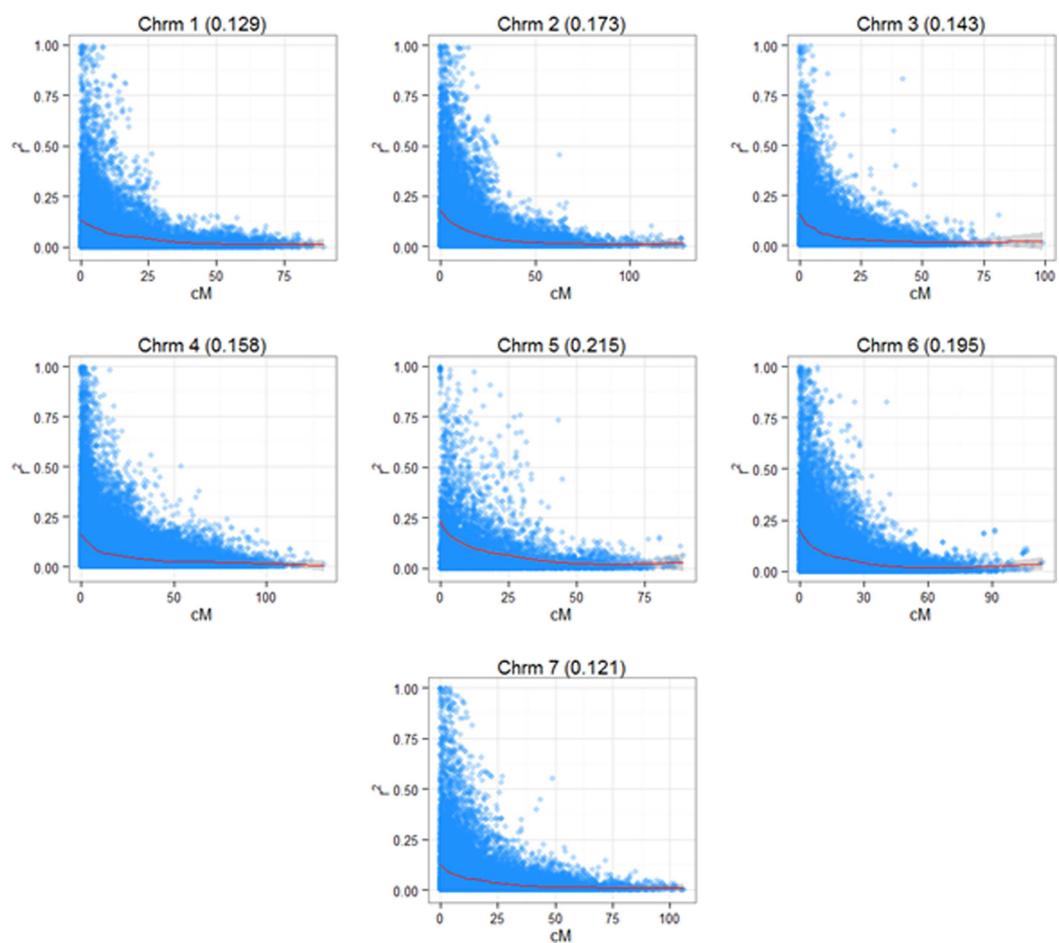
### Structure of the Breeding Populations

A 3K SNP Infinium array was used as a platform for genotyping the ryegrass breeding populations (Blackmore et al., 2015). Figure 2 shows the first two principal components from a PCA analysis on the full genotypic data set, Intermediate F11–F14, Late F5–F6 (note that F6 was not used in the analysis elsewhere, since no phenotypic information for it was available at the time of writing, but was included in the PCA analysis, since genotype data were available). The first principle component clearly separates the genotypes in two clusters, one containing the Intermediate population and one containing the Late. The two generations of the Late breeding populations, F5 and F6 form one single cluster, while the Intermediate generations separate along the second principle component. While F13 and F14 form

one cluster, F12 and in particular F11 are partially separated from the F13–F14 cluster. LD in the total breeding population is illustrated in two ways. Figure 3 shows  $r^2$  between pairs of markers against the corresponding pairwise distances for each of the seven chromosomes. The average distance between consecutive markers is given in brackets above each plot. The average LD for each pairwise marker distance ranged from 0.121 to 0.215. Supplementary Figure S1 shows landscape and heat-map plots, and they demonstrate that the average pairwise LD ignores some local variations in LD along the chromosomes. The landscape plots and heat-maps show the presence of some hotspots of LD particularly on chromosomes 2 and 6, while the overall level of LD fluctuates between 0.1 and 0.2.

### Genomic Predictions

There are two phases where GS can potentially accelerate the breeding program (Figure 1). One is at the spaced plant nursery stage where genotypic information of all the mother plants could assist in the selection of the families being taken forward to the next generation. We do not yet have that information. The other stage is the selection of parents for a new variety or synthetic population, and this is the focus of this first experiment. This is based on genotypic information from the 100 or so mother-plants selected for sward trials of its half sib progeny. We compared four prediction models for the three training sets. The results, recorded as correlations between genomically predicted values and phenotypic data, are summarized in Tables 2–4. All four methods were poor at predicting the conservation cut yield, while the predictions of total yield were slightly better overall and for vegetative yield even better. For most traits the BLUP method outperformed the other methods (see Tables 2–4). RF was the second best method with KNN and GBM trailing behind. The highest correlation between observed and predicted values was observed for the forage quality traits, particularly WSCs. This was especially pronounced for the BLUP method, where the correlation approached 0.6 when the INT and INT + F5 = ALL was used as a training population. There was, however, no consistent pattern to the effect of the training population. For BLUP and RF a trend toward better performance was discernible



**FIGURE 3 |** Linkage disequilibrium (LD) in the seven chromosomes of *L. perenne*. The diagrams show the pairwise LD ( $r^2$ ) based on a consensus map derived from three mapping families (see Materials and Methods) with a superimposed cubic smoothing spline.

with increasing size of the training population, particularly for the quality traits. However, even that was not entirely consistent. For example DMD had the highest accuracy with F13 as the training population (Tables 2–4). For the yield based traits, the best prediction accuracies were generally found in the 1st year harvests for the BLUP method (Tables 2–4). For the two largest training populations (INT and ALL), the prediction accuracy for ground cover (gcscore) was higher in year 2 than in year 1. Data for ground cover in year 3 is not yet available for F14, so prediction accuracies could not be calculated. Of the three biomass yield related traits the highest prediction accuracies were obtained with the BLUP method. The prediction accuracies for these traits were all higher for year 1 data with the BLUP model.

DISCUSSION

Accuracy and Prediction Model

This work represents the first empirical evaluation of GS in perennial ryegrass, the most important temperate forage grass

crop. We tested four prediction models and three training populations in order to assess the effect of the method and the size and composition of the training populations. The

**TABLE 2 |** Correlation between observed phenotyped and GEBV predicted by the four methods trained on F13.

F14/F13	BLUP	KNN	RF	GBM
totaly7c_yr1	0.095	<b>0.234</b>	−0.025	0.03
totaly7c_yr2	<b>0.139</b>	0.078	0.025	0.059
conscuty_yr1	0.013	0.048	−0.013	<b>0.145</b>
conscuty_yr2	−0.009	−0.046	−0.084	0.053
vegylid_yr1	<b>0.328</b>	0.188	0.12	0.191
vegylid_yr2	<b>0.167</b>	0.103	0.13	0.116
gcscore_yr1	0.345	0.25	<b>0.392</b>	0.354
gcscore_yr2	<b>0.268</b>	0.143	0.242	0.232
dmd	0.441	0.281	<b>0.45</b>	0.188
n	<b>0.319</b>	0.104	0.172	0.241
wsc	<b>0.454</b>	0.334	0.267	0.303

*The highest prediction accuracy for each trait is highlighted in bold. Trait identification is as described in Table 1.*

**TABLE 3 | Correlation between observed phenotyped and GEBV predicted by the four methods trained on INT (Intermediate F11 + F12 + F13).**

F14/INT	BLUP	KNN	RF	GBM
totaly7c_yr1	<b>0.275</b>	−0.015	0.09	0.223
totaly7c_yr2	−0.013	0.067	0.057	−0.005
conscuty_yr1	<b>0.114</b>	0.005	0.14	0.054
conscuty_yr2	−0.048	0.062	−0.105	−0.071
vegylid_yr1	<b>0.315</b>	0.093	0.295	0.134
vegylid_yr2	0.044	0.055	<b>0.071</b>	0.01
gcscore_yr1	<b>0.28</b>	−0.038	0.124	−0.206
gcscore_yr2	<b>0.339</b>	0.168	0.266	0.182
dmd	<b>0.396</b>	0.148	0.347	0.132
n	0.290	0.166	<b>0.357</b>	0.041
wsc	<b>0.590</b>	0.353	0.365	0.292

The highest prediction accuracy for each trait is highlighted in bold.

comparison between the different prediction models was most straightforward, since this can be done for each training population. Overall BLUP was the best performing method but ML techniques were reasonably successful on the F13 training population, where they outperformed BLUP for 4 out of 11 traits (Table 2). Traits with higher heritability consistently gave better prediction accuracy. This was particularly evident for DMD and WSCs (Tables 2–4), which both have the highest heritability (Table 1) and the highest prediction accuracy. One of the characteristics of the quality traits is that the frequency distribution in terms of percentage of dry matter was unimodal even after combining the data for different generations, while yield-related traits differ markedly between years, location and generation, and so have bi- or tri-modal frequency distributions. We have tried to mitigate these environmental effects here by scaling the trait values separately for each generation, and also normalizing them against phenotypic values of control varieties. Furthermore, we considered yield-related traits in different years as different traits. The effectiveness of this is very much dependent on the presence or absence of genotype by environment interaction ( $G \times E$ ). If there is considerable  $G \times E$  the predictions will be different for different years. Furthermore, variation in heritability between years is also likely to have an effect on accuracy. Tables 2–4 show that there are differences in prediction accuracies between years, and in particular differences between the effects of composition of the training population. The generally lower prediction accuracies for the yield-related traits are consistent with their lower heritability (Table 1).

## Accuracy and Training Population

The relationship between size and composition of training population on the one hand and prediction accuracy on the other was complex, and more difficult to interpret. This is because the change in training population size is compounded by the population structure (Figure 2). For 16 of the 44 trait/prediction method combinations, the prediction accuracies increased when replacing F13 with all intermediate generations, i.e., F11 + F12 + F13. This increased the training population size from 54 to 259, so everything else being equal, an increase

**TABLE 4 | Correlation between observed phenotyped and GEBV predicted by the four methods trained on ALL (INT + Late F5).**

F14/ALL	BLUP	KNN	RF	GBM
totaly7c_yr1	<b>0.224</b>	−0.066	0.135	0.012
totaly7c_yr2	<b>0.082</b>	−0.026	0.044	−0.087
conscuty_yr1	<b>0.078</b>	0.041	0.012	0.048
conscuty_yr2	0.001	−0.016	−0.036	<b>0.067</b>
vegylid_yr1	<b>0.281</b>	0.157	0.264	0.109
vegylid_yr2	0.095	0.078	<b>0.155</b>	0.064
gcscore_yr1	<b>0.234</b>	−0.005	0.01	−0.205
gcscore_yr2	<b>0.369</b>	0.136	0.252	0.175
dmd	<b>0.414</b>	0.239	0.353	0.076
n	0.314	<b>0.337</b>	0.314	0.231
wsc	<b>0.598</b>	0.366	0.402	0.277

The highest prediction accuracy for each trait is highlighted in bold.

in accuracy would be expected. However, for more than a half of the combinations this was not the case. A further increase in the training population with 105 individuals from the Late F5 generation did not improve accuracy appreciably for most of the traits/methods combinations. Population structure could partially explain this result. While the most obvious difference was between the Intermediate and the Late groups, F12 and particularly F11 diverged from the F13/F14 cluster (Figure 2). The genetic distance between the generations could possibly explain why we do not see a consistent increase in prediction accuracy with an increase in training population size. This may be equivalent to the situation in animal breeding where there are examples of loss in prediction accuracy across breed predictions as compared to within breed (Daetwyler et al., 2012; Erbe et al., 2012). Due to the limited extent of LD across breeds, it is estimated that large cross-breed reference populations are needed (Goddard and Hayes, 2009). In the ryegrass breeding populations, the genetic separation is most likely driven by a combination of deliberate selection and genetic drift, the latter of which is more important in a population with a small effective population size. A small effective population size limits the number of genes causing an effect on a trait. The original number of founders of the Intermediate population was low (10), but polycrossing in subsequent generations included approximately 400 plants, and thus helped generate a great many more haplotypes than the original 20. A combination of a larger effective population size and genetic separation requires a higher coverage of SNP markers. An estimate of the effective population size of the breeding population can be obtained as described from the empirical estimates of LD we have obtained. The expectation of LD is given by  $r^2 = 1/(4N_e c + 1)$ , where  $N_e$  is the effective population size, and  $c$  is the distance between adjacent marker in Morgans (Sved, 1971). Assuming an LD estimate of 0.1 (Figure 3, Supplementary Figure S1) and an average distance of 0.003 Morgans between adjacent markers, the effective population size is 281. This is somewhere between the original number of founders (10) and the number of parents in the polycrosses of selected spaced plants at each generation (400). Given a prediction accuracy of  $r = \sim 0.5$  and heritability

of 0.4 (approximate values for WSC), one would expect to require a training population size of 2983 unrelated individuals. The appropriate values have been substituted in the following equation:  $r^2 = Nh^2/(Nh^2 + M_e)$ , where  $M_e = 2N_eL/\ln(4N_eL)$ .  $L$  is the genome size in Morgans (eight for *L. perenne*),  $h^2$  is heritability, and  $N$  is the size of the training population (Meuwissen, 2009). The prediction accuracies we have obtained here, at least for the quality traits, with a much smaller training population is likely due to the strong relatedness of the training population to the test population. Relatedness would thus appear to be a very important factor determining the success of GS.

## Genomic Prediction in Future Ryegrass Breeding

The breeding program described here is similar to the suggested generalized scheme for implementation of GS in forage crops (Hayes et al., 2013a). It thus represents a suitable template for this initial evaluation of prediction accuracies. The particular methodology of the ryegrass breeding program, however, presents a challenge. The need to use sward trials to obtain realistic phenotypic data, especially for biomass-related traits, means that the prediction accuracies in our implementation are based on genotypic data from mother-plants and phenotypic data from sward derived from seed of half-sib progeny of the mother-plants. Given the mixture of genotypes in such a sward this is likely to lower the obtainable prediction accuracies. If genotypic data were available from all the potential pollen donors in the poly crosses, it would enable us to predict allele frequencies in the progeny, but this was not economically feasible. In a white spruce population it was also found that prediction accuracies decreased markedly when the validation population was unrelated (or had unknown relationship) to the training population (Beaulieu et al., 2014). Prediction accuracies between 0.327 and 0.435 were found where the relationship between training and validation population was closest. The larger training population and number of markers (1694 and 6358, respectively) could explain the more consistent results across traits compared to our results. Nevertheless, the prediction accuracies for the forage quality traits are comparable to those in white spruce. In switchgrass prediction accuracies for a range of morphological and quality traits varied between 0 and 0.55, and are thus also within the same range as ryegrass. In alfalfa it was recently reported that genomic prediction accuracies of biomass yield were highest within the same breeding cycle compared to prediction across cycles (Li et al., 2015). This is consistent with the situation in the ryegrass breeding program. The higher and more consistent accuracies reported in alfalfa is most likely due to higher heritabilities for the biomass traits, and that the phenotypic and genotypic data were obtained from the same spaced plants, and not half sib progeny.

As has been pointed out previously (Daetwyler et al., 2012; Liu et al., 2015) prediction accuracies are determined to a large extent by genomic relationships (population structure) and LD. Given the limited number of markers used in this study and the extent of LD in the breeding populations, it would seem likely that the accuracies obtained here are attributable to the capture

of the relatedness between genotypes. In other words, the closer the relationship between training population and test population, the fewer markers are required to obtain a given accuracy (Liu et al., 2015).

Other factors that influence the accuracy are the environmental factors affecting plants grown in different years. This is highlighted by the variable prediction accuracies between years for the yield related traits (Tables 2–4). These factors make combining populations into homogenous training sets a non-trivial, and often difficult, task. This also makes tuning hyperparameters of ML models on the training set difficult; for instance, often parameters deemed optimal by tuning on (any of the three) training populations were suboptimal choices when tested on the F14 population. This significantly reduced accuracy results produced by the three, usually very powerful, ML models on the F14 test set. Another reason for the comparatively good performance of GBLUP is the fact that biomass-related and forage quality traits are controlled by many QTLs with small effects, a situation which is optimal for GBLUP. However, for some of the combinations RF performed better than GBLUP (e.g., Table 2, DMD). If a ML prediction method was consistently outperforming other methods, it would be easy to “mix and match” prediction methods to traits. At present the results are not sufficiently consistent to consider this. Obtaining more biomass yield data from different sites (environments) should improve prediction accuracies.

In this work we considered the phenotypic performance in sward, and the GEBV values obtained from this can be used to inform which parents to select for generating a potential new synthetic population or variety (Figure 1). For this purpose prediction accuracies would need to be as high as the predictions based on phenotypic evaluation. While this is not the case, the GEBVs can also be used to assist in the selection of families (seed of a mother plant) to select for the next generation of the spaced plant nursery. The long running IBERS breeding scheme outlined in Figure 1 is in fact very similar to the one proposed in a recent review (Hayes et al., 2013a). As we obtain more and more complete information of the pedigree of the breeding populations from the genotypic data, we can begin to make informed decisions to maximize the genetic variation in the breeding population, and perhaps even reduce the size of it, while maintaining variation. The improvement of GEBVs over generations will eventually lead to a situation where they can compete with the phenotypic evaluation, and thus begin to save time (Hayes et al., 2013a).

We demonstrated the use of a GS approach, in which one standard statistical method and three ML methods were compared for predicting GEBVs in *L. perenne*. The results are most encouraging for forage quality traits, such as WSCs and DMD, and highlight several important points. Improved prediction accuracies are desirable for the yield related traits, particularly in the second year. A larger training population closely related to the validation population and a larger number of markers would probably improve accuracy. However, low heritability of a trait makes such improvements more difficult to achieve. Future work might involve devising more efficient ways of combining different sub-populations, since small



training population size together with genome wide LD (Figure 3, Supplementary Figure S1) limit the prediction ability in GS. It would also be very interesting to incorporate meteorological data into ML models thus not only accounting for some of the environmental effects, but also uncovering  $G \times E$  interactions.

## AUTHOR CONTRIBUTIONS

NG conceived the work, analyzed the data and wrote the paper, AL provided phenotypic data from the breeding programme and analysed some of the data, MH conceived the work, and developed the SNP CHIP, Andi Lovatt maintained and propagated the plants and provided technical assistance, KPS ran the SNP CHIP analysis, RK provided technical assistance with DNA extraction and marker analysis, TB developed the SNP CHIP and provided the marker data, DT developed the genetic

map used as the basis for the LD analyses, RDK conceived the work and supervised the data analysis, IA conceived the work, WP conceived the work, LS conceived and supervised the work, analyzed some of the data and wrote the paper.

## ACKNOWLEDGMENT

This work was funded by a responsive mode grant under the Industrial Partnership Award scheme from the BBSRC (BB/J006955/1) and Germinal Holdings LTD.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.00133>

## REFERENCES

- Albrecht, T., Wimmer, V., Auinger, H.-J., Erbe, M., Knaak, C., Ouzunova, M., et al. (2011). Genome-based prediction of testcross values in maize. *Theor. Appl. Genet.* 123, 339–350. doi: 10.1007/s00122-011-1587-7
- Beaulieu, J., Doerksen, T., Clement, S., Mackay, J., and Bousquet, J. (2014). Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. *Heredity* 113, 343–352. doi: 10.1038/hdy.2014.36
- Bentley, A., Scutari, M., Gosman, N., Faure, S., Bedford, F., Howell, P., et al. (2014). Applying association mapping and genomic selection to the dissection of key traits in elite European wheat. *Theor. Appl. Genet.* 127, 2619–2633. doi: 10.1007/s00122-014-2403-y
- Bernardo, R., and Yu, J. (2007). Prospects for genomewide selection for quantitative traits in maize. *Crop Sci.* 47, 1082–1090. doi: 10.2135/cropsci2006.11.0690
- Blackmore, T., Thomas, I., McMahon, R., Powell, W., and Hegarty, M. (2015). Genetic-geographic correlation revealed across a broad European ecotypic sample of perennial ryegrass (*Lolium perenne*) using array-based SNP genotyping. *Theor. Appl. Genet.* 128, 1917–1932. doi: 10.1007/s00122-015-2556-3
- Breiman, L. (2001). Random forests. *Machine Learn.* 45, 5–32. doi: 10.1023/a:1010933404324
- Casler, M. D., and Brummer, E. C. (2008). Theoretical expected genetic gains for among-and-within-family selection methods in perennial forage crops. *Crop Sci.* 48, 890–902. doi: 10.2135/cropsci2007.09.0499
- Conaghan, P., and Casler, M. D. (2011). A theoretical and practical analysis of the optimum breeding system for perennial ryegrass. *Irish J. Agric. Food Res.* 50, 47–63.
- Daetwyler, H. D., Kemper, K. E., Van Der Werf, J. H. J., and Hayes, B. J. (2012). Components of the accuracy of genomic prediction in a multi-breed sheep population. *J. Anim. Sci.* 90, 3375–3384. doi: 10.2527/jas.2011-4557
- de los Campos, G., Naya, H., Gianola, D., Crossa, J., Legarra, A., Manfredi, E., et al. (2009). Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics* 182, 375–385. doi: 10.1534/genetics.109.101501
- Dekkers, J. C. M., and Hospital, F. (2002). The use of molecular genetics in the improvement of agricultural populations. *Nat. Rev. Genet.* 3, 22–32. doi: 10.1038/nrg701
- Dudoit, S., Fridlyand, J., and Speed, T. P. (2002). Comparison of discrimination methods for the classification of tumors using gene expression data. *J. Am. Stat. Assoc.* 97, 77–87. doi: 10.1198/016214502753479248
- Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with *r* package rrBLUP. *Plant Gen.* 4, 250–255. doi: 10.3835/plantgenome2011.08.0024
- Erbe, M., Hayes, B. J., Matukumalli, L. K., Goswami, S., Bowman, P. J., Reich, C. M., et al. (2012). Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *J. Dairy Sci.* 95, 4114–4129. doi: 10.3168/jds.2011-5019
- Flint-Garcia, S. A., Thornsberry, J. M., and Buckler, E. S. (2003). Structure of linkage disequilibrium in plants. *Ann. Rev. Plant Biol.* 54, 357–374. doi: 10.1146/annurev.arplant.54.031902.134907
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Ann. Stat.* 29, 1189–1232. doi: 10.2307/2699986
- Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence. *Nature* 521, 452–459. doi: 10.1038/nature14541
- Gianola, D., Fernando, R. L., and Stella, A. (2006). Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics* 173, 1761–1776. doi: 10.1534/genetics.105.049510
- Goddard, M. E., and Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat. Rev. Genet.* 10, 381–391. doi: 10.1038/nrg2575
- Grattapaglia, D., and Resende, M. V. (2011). Genomic selection in forest tree breeding. *Tree Genet. Genom.* 7, 241–255. doi: 10.1007/s11295-010-0328-4
- Habier, D., Fernando, R. L., Kizilkaya, K., and Garrick, D. J. (2011). Extension of the bayesian alphabet for genomic selection. *BMC Bioinform.* 12:186. doi: 10.1186/1471-2105-12-186
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning*. Berlin: Springer-Verlag.
- Hayes, B. J., Cogan, N. O. I., Pembleton, L. W., Goddard, M. E., Wang, J., Spangenberg, G. C., et al. (2013a). Prospects for genomic selection in forage plant species. *Plant Breed.* 132, 133–143. doi: 10.1111/pbr.12037
- Hayes, B. J., Lewin, H. A., and Goddard, M. E. (2013b). The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation. *Trends Genet.* 29, 206–214. doi: 10.1016/j.tig.2012.11.009
- Heffner, E. L., Jannink, J.-L., and Sorrells, M. E. (2011). Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *Plant Genome* 4, 65–75. doi: 10.3835/plantgenome2010.12.0029
- Heffner, E. L., Lorenz, A. J., Jannink, J.-L., and Sorrells, M. E. (2010). Plant breeding with genomic selection: gain per unit time and cost. *Crop Sci.* 50, 1681–1690. doi: 10.2135/cropsci2009.11.0662
- Heffner, E. L., Sorrells, M. E., and Jannink, J. L. (2009). Genomic selection for crop improvement. *Crop Sci.* 49, 1–12. doi: 10.2135/cropsci2008.08.0512
- Jannink, J. L., Lorenz, A. J., and Iwata, H. (2010). Genomic selection in plant breeding: from theory to practice. *Brief. Funct. Genom.* 9, 166–177. doi: 10.1093/bfpg/elq001
- Jarquín, D., Kocak, K., Posadas, L., Hyma, K., Jedlicka, J., Graef, G., et al. (2014). Genotyping by sequencing for genomic prediction in a soybean breeding population. *BMC Genomics* 15:740. doi: 10.1186/1471-2164-15-740
- Jordan, M. I., and Mitchell, T. M. (2015). Machine learning: trends, perspectives, and prospects. *Science* 349, 255–260. doi: 10.1126/science.aaa8415

- Lande, R., and Thompson, R. (1990). Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124, 743–756.
- Li, X., Wei, Y., Acharya, A., Hansen, J. L., Crawford, J. L., Viands, D. R., et al. (2015). Genomic prediction of biomass yield in two selection cycles of a tetraploid alfalfa breeding population. *Plant Genome* 8:90. doi: 10.3835/plantgenome2014.12.0090
- Li, Z., and Sillanpää, M. J. (2012). Overview of LASSO-related penalized regression methods for quantitative trait mapping and genomic selection. *Theor. Appl. Genet.* 125, 419–435. doi: 10.1007/s00122-012-1892-9
- Lipka, A. E., Lu, F., Cherney, J. H., Buckler, E. S., Casler, M. D., and Costich, D. E. (2014). Accelerating the switchgrass (*Panicum virgatum* L.) breeding cycle using genomic selection approaches. *PLoS ONE* 9:e112227. doi: 10.1371/journal.pone.0112227
- Lister, S. J., and Dhanoa, M. S. (1998). Comparison of calibration models for the prediction of forage quality traits using near infrared spectroscopy. *J. Agric. Sci. Camb.* 131, 241–242.
- Liu, H., Zhou, H., Wu, Y., Li, X., Zhao, J., Zuo, T., et al. (2015). The impact of genetic relationship and linkage disequilibrium on genomic selection. *PLoS ONE* 10:e0132379. doi: 10.1371/journal.pone.0132379
- Long, N., Gianola, D., Rosa, G. J. M., Weigel, K. A., and Avendaño, S. (2007). Machine learning classification procedure for selecting SNPs in genomic selection: application to early mortality in broilers. *J. Anim. Breed. Genet.* 124, 377–389. doi: 10.1111/j.1439-0388.2007.00694.x
- Meuwissen, T. H. E. (2009). Accuracy of breeding values of ‘unrelated’ individuals predicted by dense SNP genotyping. *Genet. Sel. Evol.* 41, 35–35. doi: 10.1186/1297-9686-41-35
- Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.
- Ogut, J. O., Piepho, H.-P., and Schulz-Streeck, T. (2011). A comparison of random forests, boosting and support vector machines for genomic selection. *BMC Proc.* 5(Suppl. 3):S11. doi: 10.1186/1753-6561-5-s3-s11
- Ogut, J. O., Schulz-Streeck, T., and Piepho, H.-P. (2012). Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. *BMC Proc.* 6:S10. doi: 10.1186/1753-6561-6-S2-S10
- Ornella, L., Perez, P., Tapia, E., Gonzalez-Camacho, J. M., Burgueno, J., Zhang, X., et al. (2014). Genomic-enabled prediction with classification algorithms. *Heredity* 112, 616–626. doi: 10.1038/hdy.2013.144
- Piepho, H. P. (2009). Ridge regression and extensions for genomewide selection in maize. *Crop Sci.* 49, 1165–1176. doi: 10.2135/cropsci2008.10.0595
- Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., et al. (2012). Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome* 5, 103–113. doi: 10.3835/plantgenome2012.06.0006
- Posselt, U. K. (2010). “Breeding methods in cross-pollinated species,” in *Handbook of Plant Breeding, Fodder Crops and Amenity Grasses*, eds B. Boller, U. K. Posselt, and F. Veronesi (New York, NY: Springer), 39–87. doi: 10.1007/978-1-4419-0760-8
- Pryce, J. E., and Daetwyler, H. D. (2012). Designing dairy cattle breeding schemes under genomic selection: a review of international research. *Anim. Prod. Sci.* 52, 107–114. doi: 10.1071/AN11098
- R Core Development Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Rafalski, A. (2002). Applications of single nucleotide polymorphisms in crop genetics. *Curr. Opin. Plant Biol.* 5, 94–100. doi: 10.1016/S1369-5266(02)00240-6
- Resende, M. D. V., Resende, M. F. R., Sansaloni, C. P., Petrolini, C. D., Missiaggia, A. A., Aguiar, A. M., et al. (2012a). Genomic selection for growth and wood quality in Eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. *New Phytol.* 194, 116–128. doi: 10.1111/j.1469-8137.2011.04038.x
- Resende, M. F. R., Muñoz, P., Acosta, J. J., Peter, G. F., Davis, J. M., Grattapaglia, D., et al. (2012b). Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. *New Phytol.* 193, 617–624. doi: 10.1111/j.1469-8137.2011.03895.x
- Resende, R. M., Casler, M. D., and De Resende, M. D. (2014). Genomic selection in forage breeding: accuracy and methods. *Crop Sci.* 54, 143–156. doi: 10.2135/cropsci2013.05.0353
- Schaeffer, L. R. (2006). Strategy for applying genome-wide selection in dairy cattle. *J. Anim. Breed. Genet.* 123, 218–223. doi: 10.1111/j.1439-0388.2006.00595.x
- Sköt, L., Sanderson, R., Thomas, A., Sköt, K., Thorogood, D., Latypova, G., et al. (2011). Allelic variation in the perennial ryegrass flowering locus T gene is associated with changes in flowering time across a range of populations. *Plant Physiol.* 155, 1013–1022. doi: 10.1104/pp.110.169870
- Slavov, G. T., Nipper, R., Robson, P., Farrar, K., Allison, G. G., Bosch, M., et al. (2014). Genome-wide association studies and prediction of 17 traits related to phenology, biomass and cell wall composition in the energy grass *Miscanthus sinensis*. *New Phytol.* 201, 1227–1239. doi: 10.1111/nph.12621
- Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redoña, E., et al. (2015). Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLoS Genet.* 11:e1004982. doi: 10.1371/journal.pgen.1004982
- Sved, J. A. (1971). Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theor. Pop. Biol.* 2, 125–141. doi: 10.1016/0040-5809(71)90011-6
- Szymczak, S., Biernacka, J. M., Cordell, H. J., Gonzalez-Recio, O., König, I. R., Zhang, H., et al. (2009). Machine learning in genome-wide association studies. *Genet. Epidemiol.* 33, S51–S57. doi: 10.1002/gepi.20473
- Wang, L., Sorensen, P., Janss, L., Ostensen, T., and Edwards, D. (2013). Genome-wide and local pattern of linkage disequilibrium and persistence of phase for 3 Danish pig breeds. *BMC Genet.* 14:115. doi: 10.1186/1471-2156-14-115
- Wang, Y., Mette, M., Miedaner, T., Gottwald, M., Wilde, P., Reif, J., et al. (2014). The accuracy of prediction of genomic selection in elite hybrid rye populations surpasses the accuracy of marker-assisted selection and is equally augmented by multiple field evaluation locations and test years. *BMC Genomics* 15:556. doi: 10.1186/1471-2164-15-556
- Xu, S. (2013). Genetic mapping and genomic selection using recombination breakpoint data. *Genetics* 195, 1103–1115. doi: 10.1534/genetics.113.155309
- Xu, S., Zhu, D., and Zhang, Q. (2014). Predicting hybrid performance in rice using genomic best linear unbiased prediction. *Proc. Nat. Acad. Sci. U.S.A.* 111, 12456–12461. doi: 10.1073/pnas.1413750111
- Xu, Y., and Crouch, J. H. (2008). Marker-assisted selection in plant breeding: from publications to practice. *Crop Sci.* 48, 391–407. doi: 10.2135/cropsci2007.04.0191
- Zapata-Valenzuela, J., Isik, F., Maltecca, C., Wegrzyn, J., Neale, D., Mckeand, S., et al. (2012). SNP markers trace familial linkages in a cloned population of Pinus taeda—prospects for genomic selection. *Tree Genet. Genom.* 8, 1307–1318. doi: 10.1007/s11295-012-0516-5
- Zhao, Y., Gowda, M., Liu, W., Würschum, T., Maurer, H., Longin, F., et al. (2012). Accuracy of genomic selection in European maize elite breeding populations. *Theor. Appl. Genet.* 124, 769–776. doi: 10.1007/s00122-011-1745-y
- Zhao, Y., Zeng, J., Fernando, R., and Reif, J. C. (2013). Genomic prediction of hybrid wheat performance. *Crop Sci.* 53, 802–810. doi: 10.2135/cropsci2012.08.0463
- Zhong, S., Dekkers, J. C. M., Fernando, R. L., and Jannink, J. L. (2009). Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. *Genetics* 182, 355–364. doi: 10.1534/genetics.108.098277
- Ziegler, A., Destefano, A. L., König, I. R., Bardel, C., Brinza, D., Bull, S., et al. (2007). Data mining, neural nets, trees—problems 2 and 3 of genetic analysis workshop 15. *Genet. Epidemiol.* 31, S51–S60. doi: 10.1002/gepi.20280

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Grinberg, Lovatt, Hegarty, Lovatt, Sköt, Kelly, Blackmore, Thorogood, King, Armstead, Powell and Sköt. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Identification and Characterization of Switchgrass Histone *H3* and *CENH3* Genes

Jiamin Miao<sup>1,2</sup>, Taylor Frazier<sup>1</sup>, Linkai Huang<sup>2</sup>, Xinquan Zhang<sup>2\*</sup> and Bingyu Zhao<sup>1\*</sup>

<sup>1</sup> Department of Horticulture, Virginia Tech, Blacksburg, VA, USA, <sup>2</sup> Department of Grassland Science, Sichuan Agricultural University, Ya'an, China

## OPEN ACCESS

### Edited by:

Keenan Amundsen,  
University of Nebraska-Lincoln, USA

### Reviewed by:

Alan Rose,  
University of California, Davis, USA  
Hao Peng,  
Washington State University, USA

### \*Correspondence:

Xinquan Zhang  
zhangxq@sicau.edu.cn;  
Bingyu Zhao  
bzhao07@vt.edu

### Specialty section:

This article was submitted to  
Crop Science and Horticulture,  
a section of the journal  
Frontiers in Plant Science

**Received:** 22 January 2016

**Accepted:** 21 June 2016

**Published:** 12 July 2016

### Citation:

Miao J, Frazier T, Huang L, Zhang X  
and Zhao B (2016) Identification and  
Characterization of Switchgrass  
Histone *H3* and *CENH3* Genes.  
Front. Plant Sci. 7:979.  
doi: 10.3389/fpls.2016.00979

Switchgrass is one of the most promising energy crops and only recently has been employed for biofuel production. The draft genome of switchgrass was recently released; however, relatively few switchgrass genes have been functionally characterized. *CENH3*, the major histone protein found in centromeres, along with canonical *H3* and other histones, plays an important role in maintaining genome stability and integrity. Despite their importance, the histone *H3* genes of switchgrass have remained largely uninvestigated. In this study, we identified 17 putative switchgrass histone *H3* genes *in silico*. Of these genes, 15 showed strong homology to histone *H3* genes including six *H3.1* genes, three *H3.3* genes, four *H3.3-like* genes and two *H3.1-like* genes. The remaining two genes were found to be homologous to *CENH3*. RNA-seq data derived from lowland cultivar Alamo and upland cultivar Dacotah allowed us to identify SNPs in the histone *H3* genes and compare their differential gene expression. Interestingly, we also found that overexpression of switchgrass histone *H3* and *CENH3* genes in *N. benthamiana* could trigger cell death of the transformed plant cells. Localization and deletion analyses of the histone *H3* and *CENH3* genes revealed that nuclear localization of the N-terminal tail is essential and sufficient for triggering the cell death phenotype. Our results deliver insight into the mechanisms underlying the histone-triggered cell death phenotype and provide a foundation for further studying the variations of the histone *H3* and *CENH3* genes in switchgrass.

**Keywords:** *Panicum virgatum* L., histone *H3*, *CENH3*, *Agrobacterium*-mediated transient assay, cell death

## INTRODUCTION

Plant nucleosomes are composed of a protein octamer that contains two molecules of each of the core histone proteins: H2A, H2B, H3, and H4. The canonical histone H3 protein is one of the most important protein components of this complex. The histone H3 protein has a main globular domain (including  $\alpha$ N-helix,  $\alpha$ 1-helix, Loop1,  $\alpha$ 2-helix, Loop2, and  $\alpha$ 3-helix motifs), a long N-terminal tail and is post-translationally modified significantly more than the other histone proteins (Ingouff and Berger, 2010). The histone H3 family consists of four main members. Three of these members, including H3.1, H3.2, and H3.3, share more than 95% identity. The fourth member consists of the histone H3 variant *CENH3*, which is highly divergent in sequence structure in comparison to canonical histone H3. *CENH3* is specifically present at centromeres and is essential for centromere and kinetochore formation in all organisms (Allshire and Karpen, 2008).

The histone H3 proteins have been extensively characterized in many plant species. Recent studies have demonstrated that histone H3 variants have evolved to play an important role in specialized plant functions, including heterochromatin replication (Jacob et al., 2014), gene silencing (Mozzetta et al., 2014), flowering time regulation (Shafiq et al., 2014), stem elongation (Chen et al., 2013), gene activation (Ahmad and Henikoff, 2002a,b) and so on. More recently, some studies have reported that the aberrant expression of *CENH3* may cause errors during mitosis, such as triggering the formation of micronuclei and the subsequent loss of chromosomes, thus resulting in aneuploidy and possibly reduced fertility (Ravi and Chan, 2010; Lermontova et al., 2011; Sanei et al., 2011).

Switchgrass (*Panicum virgatum* L.), a warm-season perennial grass species native to North America, has a growth habitat that ranges from northern Mexico into southern Canada. Switchgrass has recently been targeted as a model herbaceous species for biofuel feedstock development (McLaughlin and Adams Kszos, 2005). With a base chromosome number of 9, ploidy levels within naturally occurring populations of the species vary from diploid ( $2n = 2x = 18$ ) to dodecaploid ( $2n = 12x = 108$ ) (Nielsen, 1944; Hultquist et al., 1996). Switchgrass has two distinct ecotypes, lowland and upland, that are characterized based on growth habitat. The lowland ecotypes, which are commonly tetraploid ( $2n = 4x = 36$ ), are frequently found in warm and humid environments in the southern regions of North America. The upland ecotypes, however, can vary in ploidy (4X, 6X, and 8X) and are adapted to colder and more arid environments in central to northern North America (Porter, 1966; Hopkins et al., 1996). Under natural conditions, inter-ecotype hybridizations between switchgrass species with different ploidy levels are rare. Even so, aneuploidy is common in switchgrass, indicating that its genome is unstable in nature and making it difficult for switchgrass genetic studies and breeding selection (Costich et al., 2010). The cause of switchgrass genome instability could be related to abnormal centromeres that interact with the mitotic spindle, which can lead to selective chromosome loss (Bennett et al., 1976). It is possible that preferential or differential expression of switchgrass *H3* and *CENH3* genes may contribute to switchgrass genome instability. Despite their importance, both switchgrass histone *H3* and *CENH3* genes have remained largely uninvestigated.

In this study, we used a homology based approach to identify histone *H3* and *CENH3* genes in switchgrass. We identified 15 putative switchgrass histone *H3* genes and two *CENH3* genes. RNA-seq analysis allowed us to detect genetic polymorphisms in these genes, as well as differences in their expression levels, between upland cultivar Alamo and lowland cultivar Dacotah. We also characterized these switchgrass histone *H3* and *CENH3* genes with respect to their subcellular localization and phenotypic responses in *Nicotiana benthamiana*. Interestingly, we report for the first time that transient overexpression of histone *H3* or *CENH3* genes can trigger programmed cell death in *Nicotiana benthamiana*. The results of this study provide a better understanding of the function of switchgrass histone *H3* genes and may aid in improving future genetic studies of this important biofuel crop.

## MATERIALS AND METHODS

### Plant Material

Switchgrass cv. Alamo, cv. Dacotah and *N. benthamiana* (PI 555478) seeds were obtained from the USDA (United States Department of Agriculture) germplasm center. Italian ryegrass (*Lolium multiflorum* L.) cv. Changjiang No.2 seeds were obtained from Sichuan Agricultural University of China. Two independent *NtSGT1*-RNAi *N. tabacum* transgenic lines (Traore et al., under revision) were also used for histone *H3* transient expression analyses. Switchgrass, tobacco and Italian ryegrass were planted in pots and grown in a growth chamber programmed for 16 h light at 28°C and 8 h dark at 24°C.

### Database Search and Sequence Analysis

The Arabidopsis histone *H3* gene (At1g09200) was used as a query to BLASTX search against the switchgrass draft genome (*Panicum virgatum* v1.1, DOE-JGI, <http://www.phytozome.net/pvirgatum>). Multiple sequence alignment of histone H3 genes/proteins was performed using DNAMAN 7.0 (Lynnon Biosoft, San Ramon, USA), and visualized using BOXSHADE 3.21 ([http://www.ch.embnet.org/software/BOX\\_form.html](http://www.ch.embnet.org/software/BOX_form.html)). The neighbor-joining phylogenetic tree was generated using MEGA6 (Tamura et al., 2013).

### RNA-seq Analysis of the Putative Switchgrass Histone *H3* and *CENH3* Genes

RNA-sequencing reads from leaf tissue of three biological replicates of Alamo and Dacotah were obtained (Hupalo et al., in preparation) and imported into CLC Genomics Workbench version 7.5 (CLCBio/Qiagen, Boston, MA). The reads were quality trimmed and filtered and mapped to the *Panicum virgatum* reference genome (v1.1) using the default parameters with the following adjustments: maximum number of hits for a read = 1, similarity fraction = 0.9, length fraction = 1.0, mismatch cost = 2, insertion cost = 3 and deletion cost = 3. After mapping, the reads were locally realigned and variants were called using the Basic Variant Detection tool with the following settings: ploidy = 4, minimum coverage = 3, minimum count = 2, minimum frequency (%) = 25.0 and base quality filter = yes. The RPKM and fold difference values for the switchgrass histone *H3* and *CENH3* genes of Alamo and Dacotah were calculated using an unpaired two-group comparison experiment feature that is part of the CLC Genomics Workbench program.

### RT-PCR Analysis of the Histone *H3* Genes

Total RNAs from fresh leaves of Alamo, tobacco, Arabidopsis and Italian ryegrass were isolated using a RNeasy® Plant Mini Kit (Qiagen, Valencia, CA) and were treated with the RNase-Free DNase Set (Qiagen) to remove DNA contamination. Reverse transcription (RT) of the first strand cDNA synthesis was performed using a DyNAmo™ cDNA Synthesis Kit (Fisher Scientific Inc, Pittsburgh, PA). PCR was performed using either gene-specific primers or conserved primers, as listed in Table S1. Gene-specific primers were designed based on the nucleotide sequences of the 5' and 3' UTRs of specific switchgrass histone



*H3* genes. Switchgrass histone *H3* conserved primers were designed based on the sequence alignment of the 15 switchgrass histone *H3* genes. Tobacco and Italian ryegrass histone *H3* gene-specific primers were designed based on known ESTs (EF051133.1, AB366152.1, and AB205017.1) located in GenBank. Primers for *Arabidopsis thaliana* *H3.3* (At1g13370) and *CENH3* (At1g01370) were designed based on the nucleotide sequences identified in the TAIR (The Arabidopsis Information Resource) database.

The iProof<sup>TM</sup> high fidelity DNA polymerase (Bio-Rad, Hercules, CA) was used for all PCR amplifications and the PCR program was run as follows: 98°C for 3 min; 98°C for 30 s, 58°C for 45 s, 72°C for 50 s (30 cycles); and a final extension at 72°C for 7 min. The PCR products were gel purified using an E.Z.N.A.<sup>®</sup> MicroElute Gel Extraction Kit (Omega Bio-Tek Inc., Norcross, GA) and were subsequently cloned into the pENTR<sup>TM</sup>/D-TOPO<sup>®</sup> vector (Invitrogen, Carlsbad, CA). The vector was then transformed into *E. coli* (*DH5α*) (Life Technologies) and the bacteria were grown at 37°C. Plasmid DNAs were isolated using a Qiagen plasmid miniprep kit (Qiagen, Valencia, CA) and sequenced using a M13 forward primer (Table S1) at the core facility of Virginia Bioinformatics Institute (Blacksburg, VA, USA).

For deletion analyses of switchgrass histone *H3* genes, a series of primers (Table S1, Figure 3) were designed to amplify different fragments based on the gene sequence of *PvH3.3* (Pavir.Ib01857.1) and *PvCENH3* (Pavir.J05674.2). Chimeric gene fusions of the *PvH3.3* N-terminal tail and the *PvCENH3* fold-domain were constructed by overlapping PCR (Figure 3C).

## Cloning of the Histone *H3* Genes into a Plant Expression Vector

Using the Gateway<sup>®</sup> LR Clonase<sup>®</sup> II Enzyme mix kit (Invitrogen, Carlsbad, CA), the different histone *H3* genes were sub-cloned into the binary vectors pEarleyGate101, pEarleyGate104 (Earley et al., 2006) and pEAQ-HT-DEST3 (Sainsbury et al., 2009). These vectors were then transformed into *E. coli* (C2110) (Wu and Zhao, 2013) by electroporation. The derived histone *H3*-YFP, YFP-*H3*, and *H3*-Hisx6 fusion genes were cloned behind the CaMV35S promoter. The plasmid constructs were sequenced using either the 35S forward primer, the YFP forward primer or gene-specific primers (Table S1).

## Agrobacterium-Mediated Transient Assays in Tobacco Plants

The pEarleyGate101-PvH3s-YFP, pEarleyGate104-YFP-H3s, and pEAQ-HT-PvH3.3-Hisx6 plasmid DNAs were transformed into *Agrobacterium tumefaciens* strain GV2260 by electroporation. Transformed *Agrobacterium* cells were grown on LB agar medium supplemented with kanamycin 50 µg/ml and rifampicin 100 µg/ml and incubated at 28°C for 2 days. *Agrobacterium*-mediated transient assays in tobacco leaves were performed as described previously (Traore and Zhao, 2011). The fully expanded leaves of 3–4 week old *N. benthamiana* or *N. tabacum* plants were chosen for infiltration. The fluorescence signal of the histone *H3*-YFP protein was

monitored 2 days post inoculation using a confocal microscope (Zeiss Axio Observer.A1, Carl Zeiss MicroImaging, Inc., Thornwood, NY).

## Analysis of Histone *H3*-YFP Fusion Proteins by Western Blot

*Agrobacterium tumefaciens* GV2260 cultures carrying the pEG101-PvH3s-YFP constructs were infiltrated into young but fully expanded tobacco leaves at a concentration of OD<sub>600nm</sub> = 0.5. Leaf disks (1.9 cm diameter) were collected 3 days post inoculation using a cork borer, ground in liquid nitrogen and re-suspended in 100 µl 3 × Laemmli buffer containing 16% β-mercaptoethanol. The tissue was then boiled for 10 min and pelleted at a high speed for 10 min. Twenty micro liters of protein extract was applied to and separated on a 10% SDS-PAGE gel. The proteins were blotted to a PVDF membrane using a Bio-Rad Trans-Blot<sup>®</sup>Turbo<sup>TM</sup> Transfer System. The membrane was blocked with 5% nonfat skim milk in 1 × Tris-saline buffer supplemented with 0.5% Tween 20 (1 × TBST). Next, the membrane was probed with anti-HA-HRP (Abcam, 1:2000) and the signal was detected with using SuperSignal<sup>®</sup> West Pico Chemiluminescent Substrate (Thermo Scientific, Waltham, MA). The chemiluminescent signals were exposed to autoradiography film (Genesee Scientific, San Diego, CA) using a Kodak film processor (Kodak, A Walsh Imaging, Inc, Pompton Lakes, NJ).

## RESULTS

### Seventeen Histone *H3* Genes were Identified from the Current Draft Switchgrass Genome

To identify putative switchgrass histone *H3* genes, we used the Arabidopsis histone *H3* gene (At1g09200) as a query to BLASTx search against the switchgrass draft genome (*Panicum virgatum* v1.1). Seventeen potential switchgrass histone *H3* genes, which showed significant similarity to the Arabidopsis histone *H3* gene, (the majority of *E*-values were <1.00e-30) were identified (Table 1).

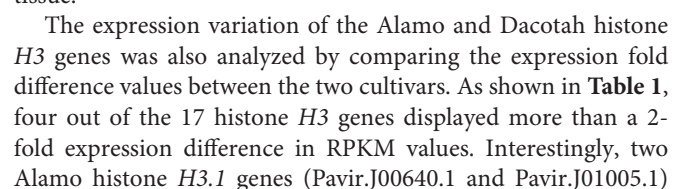
The seventeen switchgrass histone *H3* genes encode nine different proteins (Table 1). These nine proteins were aligned with human (NCBI Reference Sequence: NP\_003520.1, NP\_066403.2, and NP\_002098.1), mouse (NP\_659539.1, NP\_038576.1, and NP\_032236.1), Arabidopsis (At1g09200, At4g40030, and At1g01370), rice (Os01g0866200, Os03g0390600, and Os05g0489800) and maize (NP\_001131276.1, AFW71933.1, and NP\_001105520.1) histone *H3* and CENH3 proteins (Figure 1A). All of the switchgrass histone *H3* proteins, except for Pavir.J05674.2 and Pavir.J25829.1, which encode putative CENH3 proteins, display a high degree of homology and conservation to the histone proteins of these different species.

To further classify the switchgrass histone *H3* proteins, we used the switchgrass and Arabidopsis histone *H3* proteins to generate a phylogenetic tree (Figure 1B). The switchgrass histone *H3* proteins largely grouped into five clades. The first clade consisted of Pavir.J01005.1, Pavir.Ga01868.1, Pavir.J18804.1,

TABLE 1 | Summary of database search and expression analysis for 17 Switchgrass Histone H3 genes.

Protein ID	Gene ID (nt) <sup>a</sup>	H3 type	E-value <sup>b</sup>	Alamo RPKM <sup>c</sup> (mean)	Dacotah RPKM <sup>c</sup> (mean)	Fold Difference <sup>d</sup>	Length (aa)	Number of introns	Chromosome Location <sup>e</sup>	Number of SNPs between Alamo and Dacotah
PvCENH3.1	Pavir.J05674.2	CENH3	1.00E-03	0.43	0.57	1.17	167	6	contig08989	8
PvCENH3.2	Pavir.J25829.1	CENH3	0.18	0.33	0.98	2.97	167	6	contig28575	0
PvH3.1	Pavir.Db02133.1	H3.1	9.00E-41	0	0	—	136	0	Chr04b	0
	Pavir.J20671.1	H3.1	8.00E-32	0.06	0.14	2.48	136	0	contig224856	0
	Pavir.Ga01868.1	H3.1	2.00E-38	0	0.04	—	136	0	Chr07a	0
	Pavir.J18804.1	H3.1	3.00E-31	0	0	—	136	0	contig20530	0
	Pavir.J00640.1	H3.1	6.00E-48	1.95	0.07	-29.22	136	0	contig00653	4
	Pavir.J01005.1	H3.1	1.00E-49	5.61	0.51	-10.89	136	0	contig01099	14
PvH3.1-like-A	Pavir.Fa02085.1	H3.1-like	2.00E-41	0	0.07	—	137	0	Chr06a	0
PvH3.1-like-B	Pavir.J07529.1	H3.1-like	6.00E-45	0	0	—	136	0	contig110228	0
PvH3.3	Pavir.la03121.2	H3.3	3.00E-93	406.03	494.74	1.22	136	3	Chr09a	6
	Pavir.lb01857.1	H3.3	1.00E-95	670.22	514.42	-1.3	136	3	Chr09b	15
	Pavir.J05563.1	H3.3	8.00E-35	11.89	26.73	2.25	136	1	contig08697	24
PvH3.3-like-A	Pavir.J26857.1	H3.3-like	1.00E-95	49.51	32.06	-1.54	136	3	contig30080	11
	Pavir.J24812.1	H3.3-like	5.00E-98	100.67	88.05	-1.14	136	3	contig27272	25
PvH3.3-like-B	Pavir.J09299.1	H3.3-like	3.00E-93	1.17	1.96	1.67	136	3	contig123174	16
PvH3.3-like-C	Pavir.J10481.1	H3.3-like	1.00E-55	0.04	0	—	136	2	contig131996	0

<sup>a</sup>Full length genes were found at these locations in the Phytozome v 10.1 *Panicum virgatum* v1.1 genome.  
<sup>b</sup>E-value are shown in this table as determined by a TAIR CDS BLAST search using switchgrass histone H3 nucleotide sequences.  
<sup>c</sup>RPKM = reads per kilobase per million = Number of mapped reads/length of transcript in kilo base/million mapped reads.  
<sup>d</sup>Fold Difference = the amount of the mean expression values differ between Alamo and Dacotah.  
<sup>e</sup>Chromosome Location can find in Phytozome v10.1 *Panicum virgatum* v1.1 genome.



are expressed at 29.22 and 10.89 folds higher, respectively, than their homologous genes in *Dacotah*.

The current release of the switchgrass genome (*Panicum virgatum* v1.1) contains 636 Mb of sequencing data assembled onto 18 scaffolds with an additional 593 Mb remaining on unanchored contigs (<http://phytozome.jgi.doe.gov>). Five of the 17 histone *H3* genes identified in this study (Pavir.Fa02085.1, Pavir.Db02133.1, Pavir.Ga01868.1, Pavir.Ia03121.2, and Pavir.Ib01857.1) have been anchored onto a given switchgrass chromosome, while the others are located on contigs. Eight out of the 17 histone *H3* genes are identical in Alamo and *Dacotah*. The other nine genes have nucleotide sequence polymorphisms that contain anywhere from 4 to 25 SNPs (Table 1, Table S2).

## Cloning Switchgrass Histone *H3* Genes by RT-PCR

To validate the gene sequences and expression levels of the predicted switchgrass histone *H3* genes, we used a pair of conserved primers (Table S1) to perform RT-PCR on the leaf cDNAs of cv. Alamo. PCR products were cloned and 22 clones were randomly chosen for sequencing analysis. Fifteen clones (68%) carried the DNA sequences coding for *PvH3.3*, which suggests that the *PvH3.3* gene has a relatively higher expression level in switchgrass leaf tissue in comparison to the other histone *H3* genes. This is consistent with the RNA-seq data in which the two *PvH3.3* genes in both Alamo and *Dacotah* exhibited a relatively higher RPKM value than the RPKM value of the other histone genes (Table 1).

Four *H3.1* genes (Pavir.Ga01868.1, Pavir.J18804.1, Pavir.J00640.1, and Pavir.J01005.1), two *H3.3* genes (Pavir.Ia03121.2 and Pavir.Ib01857.1), two *H3.3-like* genes (Pavir.J26857.1 and Pavir.J24812.1) and one *CENH3* gene (Pavir.J05674.2) were also amplified from Alamo cDNA using the histone *H3* and *CENH3* specific primers (Table S1). In addition, one *H3.3* gene (*PvH3.3*) was amplified in our RT-PCR analysis (data not shown) that was not identified in the switchgrass genome (*Panicum virgatum* v1.1).

## PvH3s and PvCENH3 Fused to YFP are Predominately Located in the Plant Cell Nucleus

To test the subcellular localization of the switchgrass histone *H3* genes, we cloned Pavir.J01005.1 (*PvH3.1*), Pavir.Ib01857.1 (*PvH3.3*), Pavir.J24812.1 (*PvH3.3-like*), Pavir.J05674.2 (*PvCENH3*), and *PvH3.3* (not identified in our genome search, but identified by RT-PCR) into the binary vector pEarleygate101 (Earley et al., 2006), which fused a C-terminus YFP (yellow fluorescent protein) to each of the histone genes. The fusion genes were then transiently expressed in *Nicotiana benthamiana* plant cells. *PvH3s*-YFP and *PvCENH3*-YFP fusion proteins localized predominantly in the plant cell nucleus, whereas the control YFP was located in both the cytosol and the nucleus (Figure S2).

## Overexpression of Histone *H3*-YFP Triggers a Cell Death Phenotype in *N. Benthamiana*

When transiently overexpressed in *N. benthamiana*, both the *PvH3*-YFP and the *PvCENH3*-YFP proteins triggered a cell death phenotype 3 days post inoculation (dpi) (Figures 2A,B). Interestingly, overexpression of the histone *H3* genes cloned from other plant species, including *Arabidopsis H3.3* and *CENH3*, Italian ryegrass *H3.3* and *N. benthamiana H3.3* and *CENH3*, also triggered cell death in the transformed *N. benthamiana* plant cells (Figures 2B,C).

To test if the C-terminal YFP fusion affected the cell-death-triggering ability of histone *H3*, we fused YFP to the N-terminus of switchgrass *H3.3* by cloning the *H3.3* gene into the binary vector pEarleygate 104 (Earley et al., 2006). As shown in Figure 2C, YFP-*PvH3.3* and YFP-*PvCENH3* could also trigger cell death.

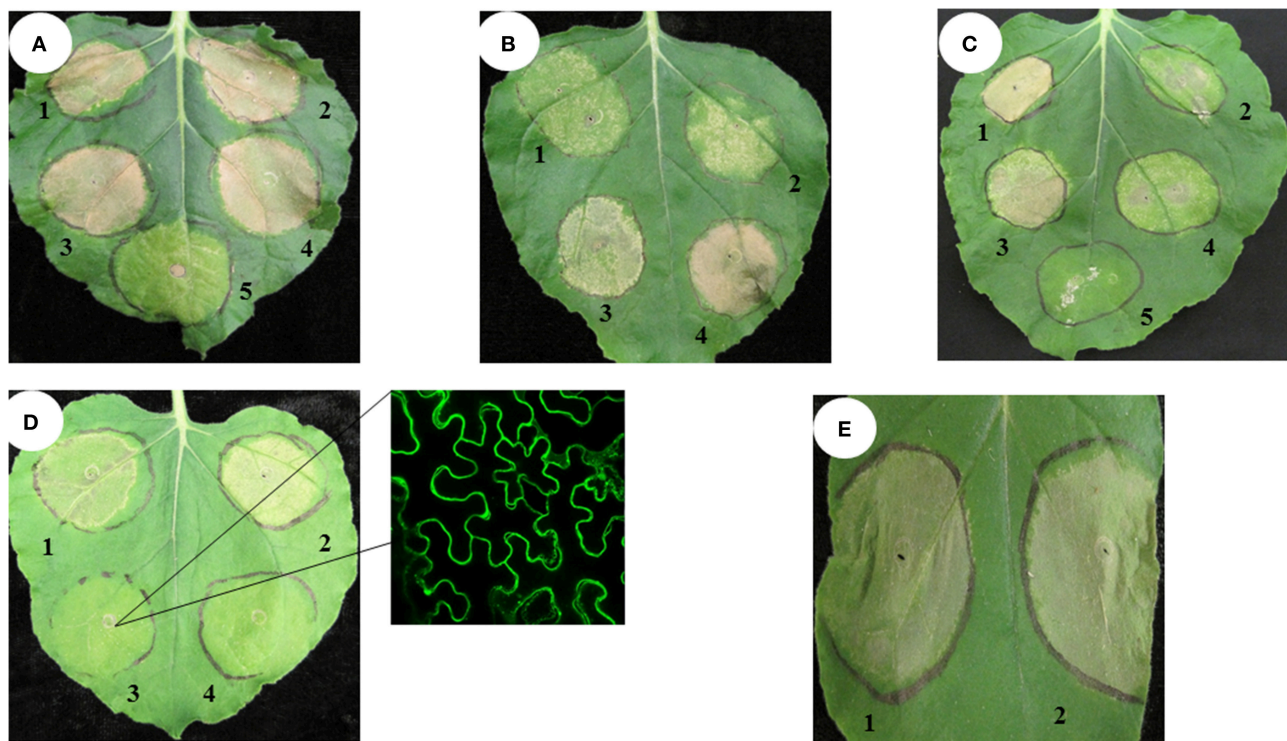
Fusion of YFP to either end of a protein may alter the native protein structure and function. Thus, YFP-tagged proteins can produce unexpected and unwanted phenotypes. To determine if YFP was contributing to the cell death phenotype of the histone *H3* genes, we also cloned *PvH3.3* into the pEAQ-HT-DEST3 vector, which fuses a smaller tag (6xHistidine) to the C-terminal end of the protein (Sainsbury et al., 2009). As shown in Figure 2E, transient expression of pEAQ-HT-*PvH3.3* was still able to trigger cell death. Therefore, we conclude that overexpression of solely the *PvH3.3* protein is able to trigger cell death in *N. benthamiana*.

Histone *H3* is one of the key elements of the nucleosome, which is predominately located in the nucleus. To test if nuclear localization of the histone *H3s* is required for its ability to trigger the cell death phenotype, we fused a myristoylation signal peptide to the N-terminus of the *PvH3.3* gene (Pavir.Ib01857.1) that also contained an N-terminal YFP tag (Figure S3). As shown in Figure 2D, the Myr-YFP-*PvH3.3* fluorescence signal was primarily detected on the plasma membrane with no obvious nuclear localization. Interestingly, the fusion of the myristoylation signal peptide could completely inhibit the cell death phenotype triggered by YFP-*PvH3.3*. This suggests that the *PvH3.3* protein needs to be localized in the plant nucleus in order to trigger cell death.

## Overexpression of the N-Terminal Tail of *PvH3.3* and *PvCENH3* Triggers Programmed Cell Death of Transformed Tobacco Cells

Histone *H3* proteins have two domains: an N-terminal tail and a histone fold domain (Luger et al., 1997; Malik and Henikoff, 2003). In order to determine the part of the *PvH3.3* protein that is essential for triggering cell death in *N. benthamiana*, we performed a deletion mutagenesis series from both the N- and C-terminal ends. Five different fragments of *PvH3.3* (Figure 3A) were used to generate *PvH3.3*-YFP fusion genes. As shown in Figure 3D, the N-terminal tail (1–43aa) is the part of the protein that maintains the ability to trigger cell death. All of the fragments that contained the N-terminal tail were predominately localized





**FIGURE 2 | Transient expression of switchgrass histone H3 in *N. benthamiana* triggered cell death phenotype.** Agrobacterium strains ( $OD_{600} = 0.5$ ) expressing different constructs were inoculated in tobacco leaves. The cell death phenotype was pictured at 4–7 days post inoculation. Pv, Switchgrass; At, Arabidopsis; Nb, tobacco; Lm, ryegrass. **(A)** 1. Pavir.J01005.1-YFP; 2. Pavir.lb01857.1-YFP; 3. Pavir.J24812.1-YFP; 4. PvH3.3-YFP; 5. YFP only (negative control); **(B)** 1. Pavir.J05674.2-YFP; 2. AtCENH3-YFP; 3. AtH3.3-YFP; 4. LmH3.3-YFP; **(C)** 1. NbH3.3-YFP; 2. NbCENH3-YFP; 3. YFP-Pavir.lb01857.1; 4. YFP-Pavir.J05674.2; 5. YFP only; **(D)** 1. Pavir.J05674.2-YFP; 2. Pavir.lb01857.1-YFP; 3. MyrYFP-Pavir.lb01857.1; 4. YFP only; the right one is MyrYFP-Pavir.lb01857.1 (YFP fluorescence signal was predominately localized in the plasma membrane of transformed plant cells); **(E)** 1. Pavir.lb01857.1-6xHistidine; 2. Pavir.lb01857.1-YFP.

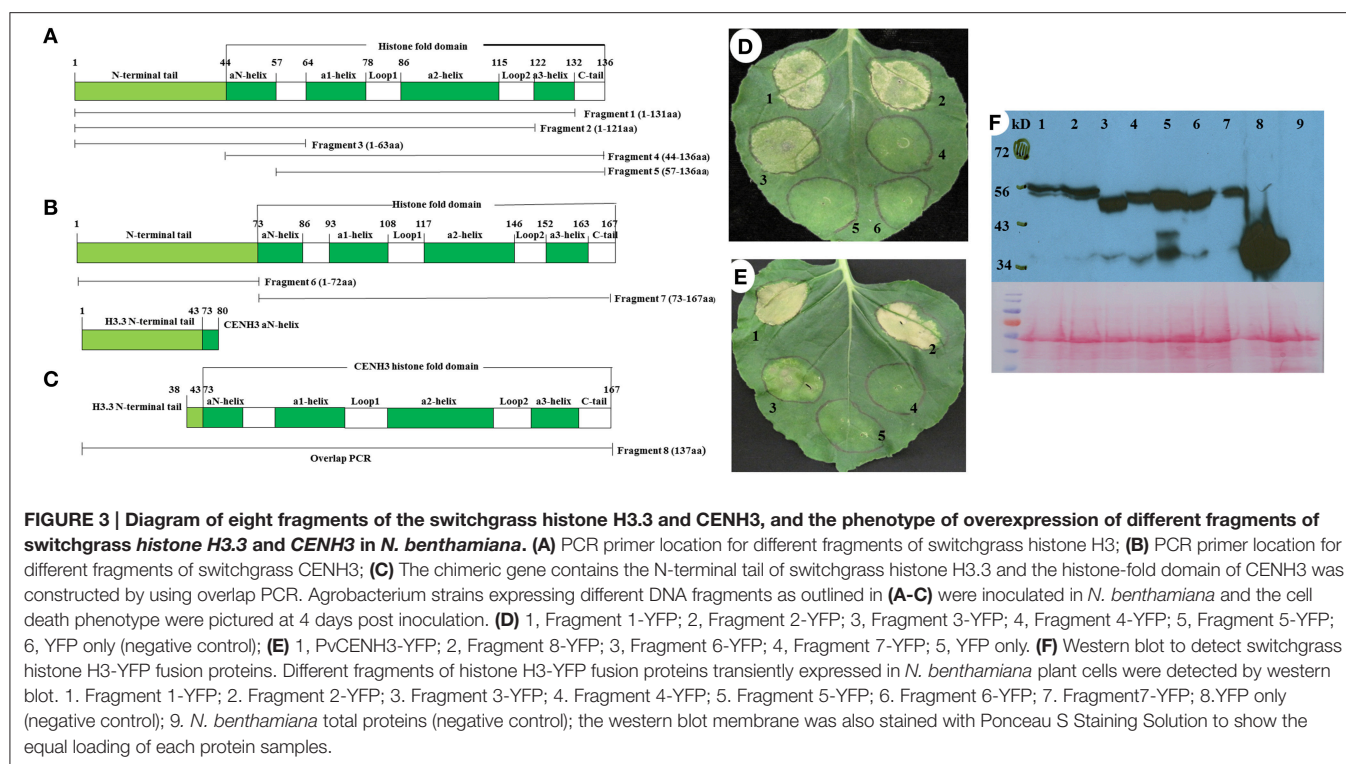
in the nucleus (**Figures S4A–C**). This suggests that there is an unidentified nuclear localization signal in the N-terminal tail sequence. The fragments that contained solely the PvH3.3 C-terminal histone fold domain, however, lost the ability to localize in the nucleus (**Figures S4D,E**) and could not trigger cell death in the transformed *N. benthamiana* plant cells (**Figure 3D**).

To investigate whether the N-terminal of *PvCENH3* is also sufficient enough to cause cell death, both the N-terminal tail and the histone fold domain of the *PvCENH3* gene were amplified (**Figure 3B**) and cloned into expression vectors that fused them with a C-terminal YFP tag. After transiently expressing each construct in tobacco, the fluorescence signal showed that the N-terminal tail of *PvCENH3* solely localized in the nucleus, whereas the histone fold domain was ambiguously located between the nucleus and the cytosol (**Figures S4F,G**). Therefore, *PvCENH3* must also contain an uncharacterized nuclear localization signal in its N-terminal tail. Similarly to the results for *PvH3.3*, overexpression of the N-terminal tail caused cell death after 3 dpi, whereas the histone fold domain failed to trigger cell death (**Figure 3E**). A previous report has suggested that a chimerical histone *H3-CENH3* fusion gene can trigger chromosome elimination in Arabidopsis (Ravi and Chan, 2010). In this study, we also fused the N-terminal tail of *PvH3.3* with

the histone fold domain of *PvCENH3* (**Figure 3C**). We found that transient expression of *PvH3.3-PvCENH3-YFP* could also trigger cell death in *N. benthamiana* plant cells (**Figure 3E**). The expressions of all different fragments of *PvH3.3* and *PvCENH3* in *N. benthamiana* were confirmed by Western blot analyses (**Figure 3F**).

### Silencing of the *SGT1* Gene in *N. Tabacum* Inhibits the Cell Death Phenotype Triggered by *PvH3.3* and *PvCENH3*

The cell death phenotype caused by the overexpression of *PvH3.3* and *PvCENH3* is similar to the hypersensitive response (HR) triggered by the interaction between plant pathogen effectors and cognate plant *R* genes (Coll et al., 2011). Since the HR-like cell death that is triggered by many *R* genes requires the function of *SGT1*, which is a conserved immune signaling component (Peart et al., 2002), we therefore tested whether or not *SGT1* is also required for the elicitation of cell death induced by overexpression of *PvH3.3* and *PvCENH3* in *Nicotiana tabacum*. Two independent *NtSGT1*-RNAi transgenic lines (Traore et al., under revision) were used for transient overexpression of *PvH3.3* and *PvCENH3*. As shown in **Figure 4A**, transient



expression of *PvH3.3*<sup>(1–63aa)</sup>-YFP and *PvCENH3*<sup>(1–72aa)</sup>-YFP triggered cell death phenotypes on the wild type *N. tabacum* plants but failed to trigger any phenotype on the *NtSGT1*-RNAi plants. Western blot analysis found that the *PvH3.3*<sup>(1–63aa)</sup>-YFP and *PvCENH3*<sup>(1–72aa)</sup>-YFP fusion proteins were both expressed in the wild type and the *NtSGT1*-RNAi transgenic plants (Figure 4B). Therefore, *NtSGT1* is essential for promoting histone H3-mediated cell death. Further studies are needed to investigate the role that SGT1 may play in histone gene-mediated cell death.

## DISCUSSION

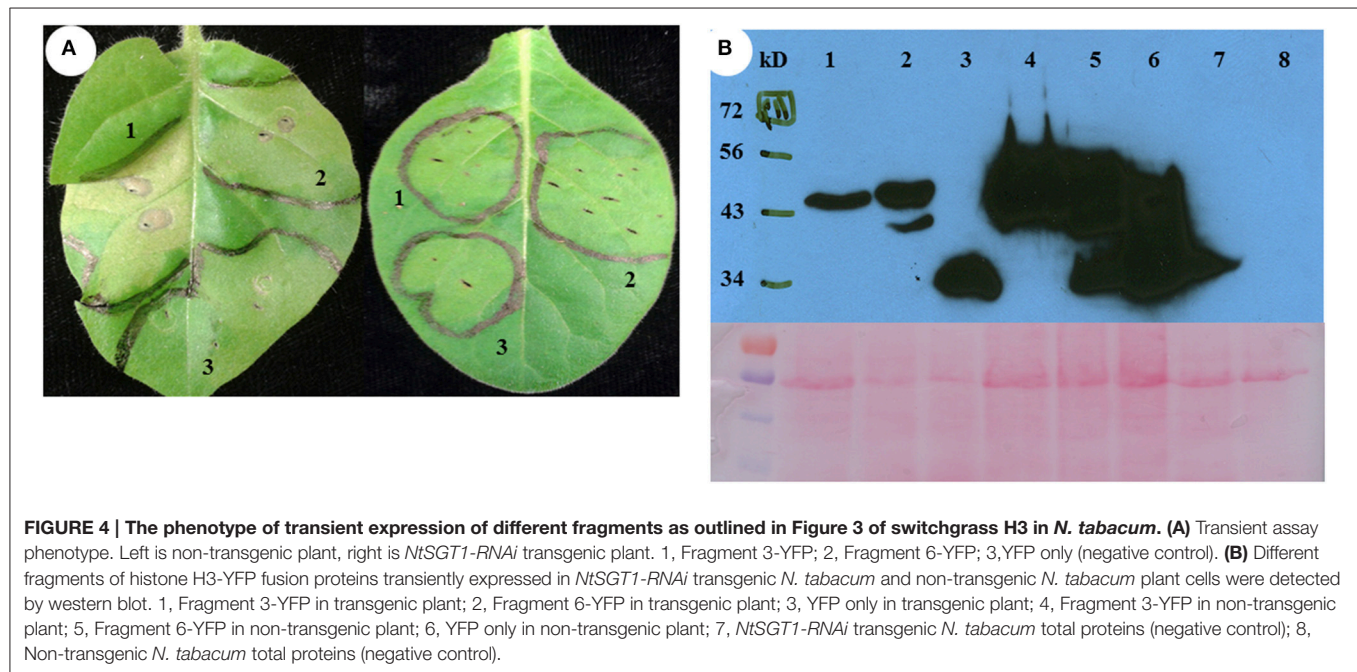
In this study, we identified 17 potential switchgrass histone *H3* genes in the draft switchgrass genome (v1.1). Although, their nucleotide sequences differ, these 17 genes encode for 9 individual proteins. Despite being a tetraploid, only two of these genes represent putative *CENH3* genes, which is the key histone component of the centromere. RNA-seq analysis of the histone *H3* genes in leaf tissue of Alamo and Dacotah found that some of the histone *H3* genes have sequence polymorphisms at the nucleotide level (Table S2). Alamo and Dacotah represent lowland and upland switchgrass ecotypes, respectively, and are genetically diverse from each other. Therefore, the identified SNPs could be developed into molecular markers that can distinguish histone *H3* gene alleles between the different ecotypes of switchgrass. We also found that transient overexpression of several of the putative switchgrass histone H3 proteins in tobacco leaves can trigger cell death.

## Histone H3 Genes in Switchgrass

Plant histone *H3* genes belong to a gene family with multiple members. For instance, maize and rice have approximately 14–15 histone *H3* genes (Ingouff and Berger, 2010). Additionally, *Arabidopsis* contains 15 histone *H3* genes including five *H3.1* genes, one *H3.1-like* gene, three *H3.3* genes, five *H3.3-like* genes, and one *CENH3* gene (Okada et al., 2005). In this study, we used a homology-based approach to identify histone *H3* genes in switchgrass. We found six *H3.1* genes, three *H3.3* genes, four *H3.3-like* genes, two *H3.1-like* genes, and two *CENH3* genes in the switchgrass genome. Core histones are among the most highly conserved proteins in eukaryotes, emphasizing their important role in maintaining genome stability and structure (Marino-Ramirez et al., 2011). In this study, we were also able to determine that histone H3 proteins from different species are highly conserved. The histone H3 proteins and the *CENH3* proteins share a relatively conserved histone fold domain; however, the N-terminal tail of each protein group shows significant sequence divergence, both among species and between each group (Figure 1A). This is consistent with previous studies showing that *CENH3* has evolved rapidly, particularly in its N-terminal tail, and the adaptive evolution that has occurred in the N-terminal tail may be in response to changing centromeric satellite repeats (Henikoff et al., 2001; Malik and Henikoff, 2001; Talbert et al., 2002).

In this study, 10 *PvH3* genes were cloned from switchgrass leaf cDNA. The other *PvH3* genes may be expressed at undetectable levels in switchgrass leaves but may be significantly expressed in other tissue types. For example, *AtMGH3*, which is an





**FIGURE 4 | The phenotype of transient expression of different fragments as outlined in Figure 3 of switchgrass H3 in *N. tabacum*.** (A) Transient assay phenotype. Left is non-transgenic plant, right is *NtSGT1-RNAi* transgenic plant. 1, Fragment 3-YFP; 2, Fragment 6-YFP; 3, YFP only (negative control). (B) Different fragments of histone H3-YFP fusion proteins transiently expressed in *NtSGT1-RNAi* transgenic *N. tabacum* and non-transgenic *N. tabacum* plant cells were detected by western blot. 1, Fragment 3-YFP in transgenic plant; 2, Fragment 6-YFP in transgenic plant; 3, YFP only in transgenic plant; 4, Fragment 3-YFP in non-transgenic plant; 5, Fragment 6-YFP in non-transgenic plant; 6, YFP only in non-transgenic plant; 7, *NtSGT1-RNAi* transgenic *N. tabacum* total proteins (negative control); 8, Non-transgenic *N. tabacum* total proteins (negative control).

Arabidopsis histone H3 gene, can only be detected in mature buds and mature bicellular and tricellular pollen (Okada et al., 2005).

### The Subcellular Localization of Histone H3

The fluorescent signals of the histone H3-YFP fusion proteins were mainly located in the nucleus. This is similar to the histone H3 that was studied in *Drosophila* (Ahmad and Henikoff, 2002b). The histone H3 fused with two tandem repeats of GFP in yeast cell is also in the nucleus (Mosammaparast et al., 2002). Using the pSORT software (<http://www.psort.org/>), we identified a putative nuclear localization signal (NLS), “KRVTIMPKDIQLARRIR,” which is conserved among the switchgrass histone H3 proteins. This NLS is located in the Loop2 and  $\alpha$ 3-helix of the protein. Two DNA binding motifs, “KAPRKQL” and “PFQRLVREI,” were identified in the N-terminal tail and the  $\alpha$ 1-helix, respectively (Figure 1A). In our transient assay experiments, however, only the N-terminal tail fused with YFP was predominantly localized in the plant nucleus and the fragments that did not contain the N-terminal tail lost the ability to localize in the nucleus (Figure S4). The result is similar with the H3 and H4 GFP fusion proteins in yeast cells. The amino-terminal domains of yeast H3-(1–58) and H4-(1–42) are necessary and sufficient for nuclear transport. While truncated H3 or H4 proteins that lacked amino terminal H3-(58–136) and H4-(42–103) resulted in reduced nuclear accumulation (Mosammaparast et al., 2002). Previous studies have indicated that the histone-fold domain of CENH3 controls kinetochore localization (Sullivan et al., 1994; Lermontova et al., 2006; Black et al., 2007). In this study, we demonstrated that the N-terminal tail of PvCENH3 is essential and sufficient for its nuclear localization (Figure S4). Interestingly, we identified a putative NLS signal, “PKKKLQE,” in the N-terminal tail of the switchgrass CENH3. Previous studies have reported that

NLS activity is found in the N-terminal tail domains of all core histones (Baake et al., 2001; Mosammaparast et al., 2001). Therefore, switchgrass histone H3 maybe also contains a NLS at its N-terminal tail. Site-directed mutagenesis on this putative NLS signal will be necessary in order to further validate its function.

### Overexpression of Histone H3 Triggers Cell Death in *N. Benthamiana* Plant Cells

In this study, we found that overexpression of switchgrass histone H3 genes could trigger a cell death phenotype in *N. benthamiana*. Since overexpression of histone H3 genes of *N. benthamiana*, Arabidopsis and Italian ryegrass also triggered a similar cell death, the cell death phenotype that we observed is not simply due to the heterologous expression of the switchgrass histone H3 genes in *N. benthamiana*. Therefore, aberrant expression of histone H3 genes appears to be toxic to plant cells. The cytotoxicity of histones has been previously reported in yeast and mammalian cells (Singh et al., 2010). A delicate balance between histone protein concentrations and DNA synthesis during the packaging of the genome into chromatin is essential for cell viability (Singh et al., 2010). Insufficient amounts of histone proteins inside the cell have been shown to be lethal (Han et al., 1987). On the other hand, excessive levels of histone proteins have also proved to be deleterious for cell growth as they promote genomic instability, increase DNA damage sensitivity and accelerate cytotoxicity (Gunjan and Verreault, 2003; Singh et al., 2009). The toxicity of large amounts of histone proteins could be attributed to their highly positive charge, which may exhibit non-specific electrostatic interactions with many negatively charged subcellular molecules including nucleic acids, such as DNA and RNA, and negatively charged proteins (Singh et al., 2010). Interestingly, mammalian

histones have been found in the extracellular space between cells. These extracellular histones may bind to cell membrane receptors and may activate multiple signaling pathways that can trigger diverse cellular responses including cytotoxicity, proinflammation, procoagulation and barrier dysfunction (Chen et al., 2014; Xu et al., 2015).

Chronic obstructive pulmonary disease (COPD) is a progressive disease that is characterized by extensive lung inflammation and apoptosis of pulmonary cells. Inflamed lung cells trigger an 8-fold increase in production of hyperacetylated histone H3.3, a modified version of this histone protein that is resistant to proteasomal degradation. As a result, the damaged cells release acetylated H3.3 into the extracellular space where it binds to lung structural cells and induces apoptosis (Barrero et al., 2013). Xu et al. found that a mixture of histones was cytotoxic to the endothelial cells (EA.hy926) of sepsis patients, and that the toxic effects were mainly due to histones H3 and H4 (Xu et al., 2009). In addition, sera from patients with sepsis directly induced histone-mediated cardiomyocyte death *ex vivo* (Alhamdi et al., 2015). *In vivo* studies on septic mice also confirmed the cause-effect relationship between circulating histones and the development of cardiac injury, arrhythmias and left ventricular dysfunction (Alhamdi et al., 2015). Therefore, the over-production of histones, either inside of cells or in the extracellular space, is toxic in mammalian cells. In the future, it will be interesting to test if increased levels of extracellular histones can also trigger cell death in plants.

The cytotoxicity caused by aberrant levels of histone proteins is normally avoided by regulating histone gene expression (Gunjan and Verreault, 2003). Correspondingly, other studies have also suggested that histone *H3* gene expression is tightly regulated, either at the transcriptional or translational levels, in different tissue types and during various developmental stages (Reichheld et al., 1998; Marino-Ramirez et al., 2005; Forcob et al., 2014). Further study on the regulation mechanisms underlying histone *H3* gene expression may help us gain a deeper understanding of the biological functions of histone H3 proteins.

The cell death phenotype observed in this study is similar to the hypersensitive response (HR) triggered by the interaction between plant disease resistance (R) proteins and their cognate effectors (Coll et al., 2011). The HR-like cell death is usually associated with ion fluxes across the plasma membrane and a burst of reactive oxygen species, such as H<sub>2</sub>O<sub>2</sub> and superoxide anion radicals. This leads to increased cytosolic Ca<sup>2+</sup> levels, activated protein kinase cascades, global transcriptional reprogramming, nuclear DNA cleavage, rapid cytoskeletal reorganization, organelle dismantling and eventually results in a rapid cell death (Pontier et al., 1998). In the future, it will be interesting to investigate if the cell death triggered by histone *H3* genes activates signaling cascades that overlap with the HR-like cell death triggered by R proteins. In this study, we demonstrated that silencing of *SGT1*, which promotes HR-like cell death in response to plant pathogens, in *N. tabacum* could completely abolish the cell death phenotype triggered by PvH3s (Figure 4A). Therefore,

*SGT1* may also function in promoting histone H3-mediated cell death.

A previous study in mice has suggested that CENH3 fused with GFP may have altered protein function and thus may affect mice embryo development (Kalitsis et al., 2003). For example, transgenic mice carrying the heterozygous *CENPA-GFP/CENPA* alleles were healthy, fertile and normal, whereas the mice carrying homozygous *CENPA-GFP/CENPA-GFP* alleles had delayed development and died during the embryo development stage (Kalitsis et al., 2003). It is possible the homozygous transgenic mice have increased CENPA-GFP protein accumulation that is lethal. We speculate that the GFP fused to switchgrass CENH3 may also have resulted in a high accumulation of PvCENH3-GFP proteins and causing cell death phenotype in the transient assay on *N. benthamiana* plants. However, further investigation is needed to understand the mechanism of cell death caused by CENPA-GFP in either mice or *N. benthamiana*.

Recent studies have shown that transgenic Arabidopsis plants expressing a chimeric GFP-tailswap protein (N-terminal tail of CENH3 was replaced by H3.3 N-terminal tail) display abnormal chromosome segregation during meiosis (Ravi and Chan, 2010; Lermontova et al., 2011). When the transgenic Arabidopsis carrying GFP-tailswap were crossed with a wild type plant, the GFP-tailswap derived genome was eliminated from the zygotes of some F<sub>1</sub> individuals, thus generating a high proportion of haploid (45%) and aneuploid (28%) progenies (Ravi and Chan, 2010). In this study, we also generated a chimerical switchgrass *CENH3* gene by replacing the N-terminal tail of *CENH3* with the *PvH3.3* N-terminal tail. Our transient assay results showed that this chimerical gene was also able to trigger a cell death phenotype (Figure 3E). Therefore, we speculate that the Arabidopsis chimeric GFP-tailswap gene may have altered native gene expression levels or resulted in histone protein accumulation in transgenic Arabidopsis, which ultimately was lethal to the plant cells. Further studies are needed to fully investigate the mechanisms underlying GFP-tailswap toxicity in Arabidopsis.

In this study, we revealed that the overexpression of the N-terminal tail of histone *PvH3.3* in the nucleus is essential and sufficient for triggering a cell death phenotype. Histone H3 proteins are important components of the nucleosome and play critical roles in the formation of higher-order chromatin (Dorigo et al., 2004; Kan et al., 2007; Sperling and Grunstein, 2009). Previous studies have demonstrated that histone N-terminal tails play an important role in the structure and stability of nucleosomes. The N-terminal truncation of histone H3 and CENH3 proteins enhances the transient unwrapping of DNA at the ribosomal entry/exit regions, which disrupts some histone-DNA contacts and thus reduces chromatin stability (Biswas et al., 2011; Tachiwana et al., 2011; Iwasaki et al., 2013). We speculate that overexpression of the histone *H3* N-terminal tail may compete with native histone H3s to interact with the nucleosomes, thus causing chromosomal instability. This interference may also lead to aberrant gene expression and disrupt faithful DNA replication, ultimately resulting in a cell



death phenotype. In the future, identification of genes that are either up- or down-regulated, through CHIP-seq or RNA-seq, by the overexpression of histone *H3* may allow us to understand how these proteins can contribute to cell death.

## ACCESSION NUMBERS

The RNA-seq data used in this article can be found in the GenBank database under the following accession numbers: SRR3473343, SRR3473344, SRR3467193, SRR3467194, SRR3467195, SRR3467196, SRR3467197, and SRR3467198.

## AUTHOR CONTRIBUTIONS

BZ, XZ designed the research projects. JM, TF performed the experiments. JM, TF, LH, XZ, and BZ wrote the manuscript. All authors reviewed and edited the manuscript before submission.

## ACKNOWLEDGMENTS

The authors thanks Drs. Hadrien Peyret and Lomonosoff for providing the vector pEAQ-HT-DEST3 used in this study. This project was partly supported by a grant from the Institute of Critical Technology and Applied Sciences (ICTAS) of Virginia Tech, the Bioprocess and Biodesign Research Center of the College of Agricultural and Life Science of Virginia Tech, a grant (2011-67009-30133) from the United States Department of Energy and National Institute of Food and Agriculture to BZ, the Foundation of American Electric Power, the National High

Technology Research and Development Program (863 Program) of China (No. 2012AA101801-02) and the Virginia Agricultural Research Station (VA135872).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.00979>

### Figure S1 | Switchgrass histone H3.3 variant exons and introns alignment.

The position of three introns were highlighted with black line on the top of the DNA sequence. The conserved nucleotides are highlighted in red. The polymorphic nucleotides A and G are in blue, while C and T are in black.

### Figure S2 | Subcellular localization of different histone H3-YFP fusion proteins. (A) PvH3.3-YFP predominately localized into plant nucleus of the transformed tobacco plant cells; (B) PvCENH3-YFP predominately localized into plant nucleus of the transformed tobacco plant cells; (C) YFP only (negative control) localized in both cytosol and nucleus of the transformed tobacco plant cells.

### Figure S3 | The diagram of binary vector pEG203-MyrYFP for expressing proteins fused with an N-terminal myristoylation signal peptide. A DNA fragment contains an N-terminal myristoylation signal peptide fused with YFP gene was amplified using overlap PCR method. The derived fragment was inserted into pEarleyGate 203 to generate pEG203-MyrYFP. Myr, Myristoylation signal peptide; TEV, TEV protease cleavage site.

### Figure S4 | Localization of different fragments outlined in Figure 3 of H3.3-YFP and CENH3-YFP fusion proteins in the transformed tobacco plant cells. (A) Fragment 1-YFP; (B) Fragment 2-YFP; (C) Fragment 3-YFP; (D) Fragment 4-YFP; (E) Fragment 5-YFP; (F) Fragment 6-YFP; (G) Fragment 7-YFP.

### Table S1 | Primer sequences used in this study.

### Table S2 | SNPs identified between cv. Alamo and cv. Dacotah. SNV, single nucleotide variant.

## REFERENCES

- Ahmad, K., and Henikoff, S. (2002a). Histone H3 variants specify modes of chromatin assembly. *Proc. Natl. Acad. Sci. U.S.A.* 99, 16477–16484. doi: 10.1073/pnas.172403699
- Ahmad, K., and Henikoff, S. (2002b). The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol. Cell* 9, 1191–1200. doi: 10.1016/S1097-2765(02)00542-7
- Alhamdi, Y., Abrams, S. T., Cheng, Z., Jing, S., Su, D., Liu, Z., et al. (2015). Circulating histones are major mediators of cardiac injury in patients with sepsis. *Crit. Care Med.* 43, 2094–2103. doi: 10.1097/CCM.0000000000001162
- Allshire, R. C., and Karpen, G. H. (2008). Epigenetic regulation of centromeric chromatin: old dogs, new tricks? *Nat. Rev. Genet.* 9, 923–937. doi: 10.1038/nrg2466
- Baake, M., Doenecke, D., and Albig, W. (2001). Characterisation of nuclear localisation signals of the four human core histones. *J. Cell. Biochem.* 81, 333–346. doi: 10.1002/1097-4644(20010501)81:2<333::AID-JCB1048>3.0.CO;2-D
- Barrero, C. A., Perez-Leal, O., Aksoy, M., Moncada, C., Ji, R., Lopez, Y., et al. (2013). Histone 3.3 participates in a self-sustaining cascade of apoptosis that contributes to the progression of chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* 188, 673–683. doi: 10.1164/rccm.201302-0342OC
- Bennett, M. D., Finch, R. A., and Barclay, I. R. (1976). The time rate and mechanism of chromosome elimination in *Hordeum* hybrids. *Chromosoma* 54, 175–200. doi: 10.1007/BF00292839
- Biswas, M., Voltz, K., Smith, J. C., and Langowski, J. (2011). Role of histone tails in structural stability of the nucleosome. *PLoS Comput. Biol.* 7:e1002279. doi: 10.1371/journal.pcbi.1002279
- Black, B. E., Jansen, L. E., Maddox, P. S., Foltz, D. R., Desai, A. B., Shah, J. V., et al. (2007). Centromere identity maintained by nucleosomes assembled with histone H3 containing the CENP-A targeting domain. *Mol. Cell* 25, 309–322. doi: 10.1016/j.molcel.2006.12.018
- Chen, Q., Chen, X., Wang, Q., Zhang, F., Lou, Z., Zhang, Q., et al. (2013). Structural basis of a histone H3 lysine 4 demethylase required for stem elongation in rice. *PLoS Genet.* 9:e1003239. doi: 10.1371/journal.pgen.1003239
- Chen, R., Kang, R., Fan, X. G., and Tang, D. (2014). Release and activity of histone in diseases. *Cell Death Dis.* 5, e1370. doi: 10.1038/cddis.2014.337
- Coll, N. S., Epple, P., and Dangel, J. L. (2011). Programmed cell death in the plant immune system. *Cell Death Differ.* 18, 1247–1256. doi: 10.1038/cdd.2011.37
- Costich, D. E., Friebe, B., Sheehan, M. J., Casler, M. D., and Buckler, E. S. (2010). Genome-size variation in switchgrass (*Panicum virgatum*): flow cytometry and cytology reveal rampant aneuploidy. *Plant Genome* 3, 130–141. doi: 10.3835/plantgenome2010.04.0010
- Dorigo, B., Schalch, T., Kulangara, A., Duda, S., Schroeder, R. R., and Richmond, T. J. (2004). Nucleosome arrays reveal the two-start organization of the chromatin fiber. *Science* 306, 1571–1573. doi: 10.1126/science.1103124
- Earley, K. W., Haag, J. R., Pontes, O., Oppen, K., Juehne, T., Song, K., et al. (2006). Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* 45, 616–629. doi: 10.1111/j.1365-313X.2005.02617.x
- Forcob, S., Bulic, A., Jonsson, F., Lipps, H., and Postberg, J. (2014). Differential expression of histone H3 genes and selective association of the variant H3.7 with a specific sequence class in *Stylonychia macronuclear* development. *Epigenetics Chromatin* 7:4. doi: 10.1186/1756-8935-7-4
- Gunjan, A., and Verreault, A. (2003). A Rad53 kinase-dependent surveillance mechanism that regulates histone protein levels in *S. cerevisiae*. *Cell* 115, 537–549. doi: 10.1016/S0092-8674(03)00896-1

- Han, M., Chang, M., Kim, U. J., and Grunstein, M. (1987). Histone H2B repression causes cell-cycle-specific arrest in yeast: effects on chromosomal segregation, replication, and transcription. *Cell* 48, 589–597.
- Henikoff, S., Ahmad, K., and Malik, H. S. (2001). The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* 293, 1098–1102. doi: 10.1126/science.1062939
- Hopkins, A. A., Taliaferro, C. M., Murphy, C. D., and Christian, D. A. (1996). Chromosome number and nuclear DNA content of several switchgrass populations. *Crop Sci.* 36, 1192–1195.
- Hultquist, S. J., Vogel, K. P., Lee, D. J., Arumuganathan, K., and Kaeppler, S. (1996). Chloroplast DNA and nuclear DNA content variations among cultivars of switchgrass, *Panicum virgatum* L. *Crop Sci.* 36, 1049–1052.
- Ingouff, M., and Berger, F. (2010). Histone3 variants in plants. *Chromosoma* 119, 27–33. doi: 10.1007/s00412-009-0237-1
- Iwasaki, W., Miya, Y., Horikoshi, N., Osakabe, A., Taguchi, H., Tachiwana, H., et al. (2013). Contribution of histone N-terminal tails to the structure and stability of nucleosomes. *FEBS Open Bio* 3, 363–369. doi: 10.1016/j.fob.2013.08.007
- Jacob, Y., Bergamin, E., Donoghue, M. T., Mongeon, V., LeBlanc, C., Voigt, P., et al. (2014). Selective methylation of histone H3 variant H3.1 regulates heterochromatin replication. *Science* 343, 1249–1253. doi: 10.1126/science.1248357
- Kalitsis, P., Fowler, K. J., Earle, E., Griffiths, B., Howman, E., Newson, A. J., et al. (2003). Partially functional Cenpa-GFP fusion protein causes increased chromosome missegregation and apoptosis during mouse embryogenesis. *Chromosome Res.* 11, 345–357. doi: 10.1023/A:1024044008009
- Kan, P. Y., Lu, X., Hansen, J. C., and Hayes, J. J. (2007). The H3 tail domain participates in multiple interactions during folding and self-association of nucleosome arrays. *Mol. Cell. Biol.* 27, 2084–2091. doi: 10.1128/MCB.02181-06
- Lermontova, I., Koroleva, O., Rutten, T., Fuchs, J., Schubert, V., Moraes, I., et al. (2011). Knockdown of CENH3 in Arabidopsis reduces mitotic divisions and causes sterility by disturbed meiotic chromosome segregation. *Plant J.* 68, 40–50. doi: 10.1111/j.1365-313X.2011.04664.x
- Lermontova, I., Schubert, V., Fuchs, J., Klatte, S., Macas, J., and Schubert, I. (2006). Loading of Arabidopsis centromeric histone CENH3 occurs mainly during G2 and requires the presence of the histone fold domain. *Plant Cell* 18, 2443–2451. doi: 10.1105/tpc.106.043174
- Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–260.
- Malik, H. S., and Henikoff, S. (2001). Adaptive evolution of Cid, a centromere-specific histone in Drosophila. *Genetics* 157, 1293–1298.
- Malik, H. S., and Henikoff, S. (2003). Phylogenomics of the nucleosome. *Na. Struct. Mol. Biol.* 10, 882–891. doi: 10.1038/nsb996
- Marino-Ramirez, L., Kann, M. G., Shoemaker, B. A., and Landsman, D. (2005). Histone structure and nucleosome stability. *Expert Rev. Proteomics* 2, 719–729. doi: 10.1586/14789450.2.5.719
- Marino-Ramirez, L., Levine, K. M., Morales, M., Zhang, S., Moreland, R. T., Baxeianis, A. D., et al. (2011). The Histone database: an integrated resource for histones and histone fold-containing proteins. *Database* 2011:bar048. doi: 10.1093/database/bar048
- McLaughlin, S. B., and Adams Kszos, L. (2005). Development of switchgrass (*Panicum virgatum*) as a bioenergy feedstock in the United States. *Biomass Bioenergy* 28, 515–535. doi: 10.1016/j.biombioe.2004.05.006
- Mosammaparast, N., Guo, Y., Shabanowitz, J., Hunt, D. F., and Pemberton, L. F. (2002). Pathways mediating the nuclear import of histones H3 and H4 in yeast. *J. Biol. Chem.* 277, 862–868. doi: 10.1074/jbc.M106845200
- Mosammaparast, N., Jackson, K. R., Guo, Y., Brame, C. J., Shabanowitz, J., Hunt, D. F., et al. (2001). Nuclear import of histone H2A and H2B is mediated by a network of karyopherins. *J. Cell Biol.* 153, 251–262. doi: 10.1083/jcb.153.2.251
- Mozzetta, C., Pontis, J., Fritsch, L., Robin, P., Portoso, M., Proux, C., et al. (2014). The histone H3 lysine 9 methyltransferases G9a and GLP regulate polycomb repressive complex 2-mediated gene silencing. *Mol. Cell* 53, 277–289. doi: 10.1016/j.molcel.2013.12.005
- Nielsen, E. L. (1944). Analysis of variation in *Panicum virgatum*. *J. Agric. Res.* 69, 327–353.
- Okada, T., Endo, M., Singh, M. B., and Bhalla, P. L. (2005). Analysis of the histone H3 gene family in Arabidopsis and identification of the male-gamete-specific variant AtMGH3. *plant J.* 44, 557–568. doi: 10.1111/j.1365-313X.2005.02554.x
- Peart, J. R., Lu, R., Sadanandom, A., Malcuit, I., Moffett, P., Brice, D. C., et al. (2002). Ubiquitin ligase-associated protein SGT1 is required for host and nonhost disease resistance in plants. *Proc. Natl. Acad. Sci. U.S.A.* 99, 10865–10869. doi: 10.1073/pnas.152330599
- Pontier, D., Balague, C., and Roby, D. (1998). The hypersensitive response. A programmed cell death associated with plant resistance. *Comptes. Rend. l'Acad. Des Sci. III* 321, 721–734.
- Porter, C. L. Jr. (1966). An analysis of variation between upland and lowland switchgrass, *Panicum virgatum* L., in central Oklahoma. *Ecology* 47, 980–992.
- Ravi, M., and Chan, S. W. (2010). Haploid plants produced by centromere-mediated genome elimination. *Nature* 464, 615–618. doi: 10.1038/nature08842
- Reichheld, J. P., Gigot, C., and Chaubet-Gigot, N. (1998). Multilevel regulation of histone gene expression during the cell cycle in tobacco cells. *Nucleic Acids Res.* 26, 3255–3262. doi: 10.1093/nar/26.13.3255
- Sainsbury, F., Thuenemann, E. C., and Lomonosoff, G. P. (2009). pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol. J.* 7, 682–693. doi: 10.1111/j.1467-7652.2009.00434.x
- Sanei, M., Pickering, R., Kumke, K., Nasuda, S., and Houben, A. (2011). Loss of centromeric histone H3 (CENH3) from centromeres precedes uniparental chromosome elimination in interspecific barley hybrids. *Proc. Natl. Acad. Sci. U.S.A.* 108, E498–E505. doi: 10.1073/pnas.1103190108
- Shafiq, S., Berr, A., and Shen, W. H. (2014). Combinatorial functions of diverse histone methylations in *Arabidopsis thaliana* flowering time regulation. *New phytol.* 201, 312–322. doi: 10.1111/nph.12493
- Singh, R. K., Kabbaj, M. H., Paik, J., and Gunjan, A. (2009). Histone levels are regulated by phosphorylation and ubiquitylation-dependent proteolysis. *Nat. Cell Biol.* 11, 925–933. doi: 10.1038/ncb1903
- Singh, R. K., Liang, D., Gajjalaiahvari, U. R., Kabbaj, M. H., Paik, J., and Gunjan, A. (2010). Excess histone levels mediate cytotoxicity via multiple mechanisms. *Cell Cycle* 9, 4236–4244. doi: 10.4161/cc.9.20.13636
- Sperling, A. S., and Grunstein, M. (2009). Histone H3 N-terminus regulates higher order structure of yeast heterochromatin. *Proc. Natl. Acad. Sci. U.S.A.* 106, 13153–13159. doi: 10.1073/pnas.0906866106
- Sullivan, K. F., Hechenberger, M., and Masri, K. (1994). Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. *J. Cell Biol.* 127, 581–592.
- Tachiwana, H., Kagawa, W., Shiga, T., Osakabe, A., Miya, Y., Saito, K., et al. (2011). Crystal structure of the human centromeric nucleosome containing CENP-A. *Nature* 476, 232–235. doi: 10.1038/nature10258
- Talbert, P. B., Masuelli, R., Tyagi, A. P., Comai, L., and Henikoff, S. (2002). Centromeric localization and adaptive evolution of an Arabidopsis histone H3 variant. *Plant Cell* 14, 1053–1066. doi: 10.1105/tpc.010425
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Traore, S. M., and Zhao, B. (2011). A novel gateway(R)-compatible binary vector allows direct selection of recombinant clones in *Agrobacterium tumefaciens*. *Plant Methods* 7:42. doi: 10.1186/1746-4811-7-42
- Wu, S., and Zhao, B. (2013). The selection of recombinant binary plasmids generated by gateway(R) LR cloning in the *Escherichia coli* strain C2110. *Mol. Biotechnol.* 54, 125–132. doi: 10.1007/s12033-012-9548-1
- Xu, J., Zhang, X., Pelayo, R., Monestier, M., Ammollo, C. T., Semeraro, F., et al. (2009). Extracellular histones are major mediators of death in sepsis. *Nat. Med.* 15, 1318–1321. doi: 10.1038/nm.2053
- Xu, Z., Huang, Y., Mao, P., Zhang, J., and Li, Y. (2015). Sepsis and ARDS: the dark side of histones. *Mediators Inflamm.* 2015:205054. doi: 10.1155/2015/205054

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Miao, Frazier, Huang, Zhang and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Identification, Characterization, and Expression Analysis of Cell Wall Related Genes in *Sorghum bicolor* (L.) Moench, a Food, Fodder, and Biofuel Crop

Krishan M. Rai<sup>1</sup>, Sandi W. Thu<sup>1</sup>, Vimal K. Balasubramanian<sup>1</sup>, Christopher J. Cobos<sup>1</sup>, Tesfaye Disasa<sup>1,2</sup> and Venugopal Mendu<sup>1\*</sup>

<sup>1</sup> Department of Plant and Soil Science, Fiber and Biopolymer Research Institute, Texas Tech University, Lubbock, TX, USA,

<sup>2</sup> National Agricultural Biotechnology Research Center, Ethiopian Institute of Agricultural Research, Addis Ababa, Ethiopia

## OPEN ACCESS

### Edited by:

Teresa Donze,  
West Chester University, USA

### Reviewed by:

Chuang Ma,  
Northwest A&F University, China  
Liezhaio Liu,  
Northwest A&F University, China

### \*Correspondence:

Venugopal Mendu  
venugopal.mendu@ttu.edu

### Specialty section:

This article was submitted to  
Crop Science and Horticulture,  
a section of the journal  
Frontiers in Plant Science

**Received:** 28 May 2016

**Accepted:** 11 August 2016

**Published:** 31 August 2016

### Citation:

Rai KM, Thu SW,  
Balasubramanian VK, Cobos CJ,  
Disasa T and Mendu V (2016)  
Identification, Characterization, and  
Expression Analysis of Cell Wall  
Related Genes in *Sorghum bicolor* (L.)  
Moench, a Food, Fodder, and Biofuel  
Crop. *Front. Plant Sci.* 7:1287.  
doi: 10.3389/fpls.2016.01287

Biomass based alternative fuels offer a solution to the world's ever-increasing energy demand. With the ability to produce high biomass in marginal lands with low inputs, sorghum has a great potential to meet second-generation biofuel needs. Despite the sorghum crop importance in biofuel and fodder industry, there is no comprehensive information available on the cell wall related genes and gene families (biosynthetic and modification). It is important to identify the cell wall related genes to understand the cell wall biosynthetic process as well as to facilitate biomass manipulation. Genome-wide analysis using gene family specific Hidden Markov Model of conserved domains identified 520 genes distributed among 20 gene families related to biosynthesis/modification of various cell wall polymers such as cellulose, hemicellulose, pectin, and lignin. Chromosomal localization analysis of these genes revealed that about 65% of cell wall related genes were confined to four chromosomes (Chr. 1–4). Further, 56 tandem duplication events involving 169 genes were identified in these gene families which could be associated with expansion of genes within families in sorghum. Additionally, we also identified 137 Simple Sequence Repeats related to 112 genes and target sites for 10 miRNAs in some important families such as cellulose synthase, cellulose synthase-like, and laccases, etc. To gain further insight into potential functional roles, expression analysis of these gene families was performed using publically available data sets in various tissues and under abiotic stress conditions. Expression analysis showed tissue specificity as well as differential expression under abiotic stress conditions. Overall, our study provides a comprehensive information on cell wall related genes families in sorghum which offers a valuable resource to develop strategies for altering biomass composition by plant breeding and genetic engineering approaches.

**Keywords:** cell wall polymers, cellulose, hemicellulose, lignin, pectin, plant biomass, sorghum, abiotic stress

## INTRODUCTION

Sorghum (*Sorghum bicolor*), a C4 grass species, is one of the world's important multipurpose cereal crops with uses in food, fodder, and biofuel industries. Sorghum with its relatively smaller genome size (~730 Mbp) makes it an ideal model bioenergy crop compared to other C4 crops such as switch grass, sugarcane or miscanthus with bigger and more complex genomes. Grain sorghum is grown worldwide with an annual production of 62 million tons of grain yield from an estimated area of 42 million hectares (FAOSTAT data 2013; <http://faostat3.fao.org>). Biomass yield of energy sorghum (fodder sorghum) is twice that of the grain sorghum due to longer vegetative growth period, increased leaf area which helps in greater radiation interception and efficiently converting the synthesized carbon into cell wall polysaccharides (Olson et al., 2012). Apart from food, livestock feed, and biofuel source, sorghum is also a source of malt for brewing and various food industries (Taylor et al., 2006). With a higher biomass yield potential (15–40 Mg/hc), sorghum can be used as high value energy source in second generation biofuel industry (Rooney et al., 2007). Due to its high adaptability to different environmental conditions such as drought, salinity, water-logging, ability to grow in marginal land areas, efficient light to biomass energy conversion rate, and nitrogen utilization rate, sorghum is emerging as favorite multipurpose crop in recent years (Taylor et al., 2010; Byrt et al., 2011).

The focus of second generation biofuels is to produce biofuels from lignocellulosic material, which is derived mainly from plant cell walls. Lignocellulosic material is mainly composed of cellulose (15–40%), hemicellulose (30–40%), lignin (20–30%), and pectins (Mendu et al., 2011b). These structural polymers composition varies between primary and secondary cell walls, different tissues of an individual plant and among different plant species (Mendu et al., 2011b; Welker et al., 2015). Primary cell wall (PCW) is present in all plant cell types whereas secondary cell wall (SCW) is present in specific cell types such as tracheary elements (TE) and sclerenchymal cells. Cellulose, the most abundant polymer on earth, is a homopolymer of  $\beta$ -(1,4)-linked glucose monomers (Cosgrove, 2005; Somerville, 2006) whereas hemicelluloses are branched heteropolymers of pentose and hexose sugar monomers (Burton et al., 2010; Ochoa-Villarreal et al., 2012). Pectin, a complex polymer consists of  $\alpha$ -(1–4)-linked D-galacturonic acid backbone, is another polysaccharide which is mainly present in primary cell walls. Three classes of pectins, based on the nature of the sugars on the branches have been known in plants; homogalacturonans (HG), rhamnogalacturonans-I (RG-I), and rhamnogalacturonans-II (RG-II; Burton et al., 2010). The cellulose microfibrils are cross-linked with various matrix polysaccharides such as hemicelluloses and pectins thereby forming a complex polymeric network to maintain the cell wall strength (Cosgrove, 2005; Muthamilarasan et al., 2015). In addition to cellulose, hemicellulose and pectin, plant secondary cell walls are enriched with lignin. Lignin is a complex aromatic heteropolymer synthesized mainly from three canonical hydroxycinnamyl alcohol monomers viz. *p*-coumaryl (H), coniferyl (G), and sinapyl (S) alcohols (Boerjan et al., 2003;

Vanholme et al., 2010; Welker et al., 2015). Lignin is ester- and ether-linked with cellulose and hemicellulose polysaccharides in the plant cell walls with the help of ferulic acid (Harris and Trethewey, 2010). Callose, another  $\beta$ -1,3-linked glucan polymer, is present in the cell walls of specialized structures involved in pollen development, cell wall formation during cytokinesis, and plasmodesmatal canals (Nedukha, 2015). Apart from developmental deposition, callose is deposited in response to various external stimuli including biotic and abiotic stresses (Chen and Kim, 2009; Muthamilarasan et al., 2015).

Cell wall biosynthesis, reassembly, and degradation are complex processes, which involves cell wall biosynthetic, modification, and degrading enzymes. Cellulose is synthesized by plasma membrane localized cellulose synthase complexes while other matrix polysaccharides such as hemicelluloses and pectins are synthesized in Golgi complex followed by their transport and cross-linking/embedding which involves cell wall biosynthetic, modifying and degrading enzymes. Cell wall hydrolyzing enzymes produced by bacteria, fungi, and nematodes (Rai et al., 2015) degrade plant cell walls to gain entry into the plant cell and access the sugars for their survival while the cell wall hydrolyzing enzymes produced by plant cells are primarily involved in controlled cleavage of wall polymers to facilitate cell growth and elongation (Cosgrove, 2005). Carbohydrate Active enZymes (CAZy; <http://www.CAZy.org/>) database broadly classified cell wall enzymes into 135 families of Glycoside Hydrolases (GHs), 98 families of Glycosyl Transferases (GTs), 24 families of Polysaccharide Lyases (PLs), 16 families of Carbohydrate Esterases (CEs), and 13 families of Auxiliary Activities (AAs) enzymes based on the presence of protein catalytic or functional domains (Lombard et al., 2014). Some other web based databases such as Cell Wall Navigator (Girke et al., 2004) and Cell Wall Genomics (<https://cellwall.genomics.purdue.edu/families/index.html>) further classified these enzymes into different groups based on biological processes in which they are involved.

Most of the enzymes involved directly in polysaccharide biosynthesis belong to the glycosyl transferases. Glycosyl transferases form glycosidic bonds by catalyzing the transfer of sugar moieties from donor to acceptor molecules (Scheible and Pauly, 2004). Cellulose microfibrils are synthesized exclusively by cellulose synthases A (CESA) protein complexes, which belong to GT2 family of enzymes. Apart from Cesa genes, *Cellulose synthase like* (Csl) genes are also found in plants which are involved in hemicellulose and other glucan biosynthesis (Lerouxel et al., 2006). Among the other hemicellulose biosynthetic enzymes, xyloglucan  $\alpha$ -1,6-xylosyltransferases (GT34), xyloglucan fucosyltransferases (GT37), xyloglucan galactosyltransferases (GT47) are involved in synthesis of various xylan and xyloglucan molecules (Zhong and Ye, 2003; Del Bem and Vincentz, 2010; Vuttipongchaikij et al., 2012; Zabortina et al., 2012; Voiniciuc et al., 2015). The pectin biosynthetic galacturonosyltransferases (GT8) genes such as *FRAGILE FIBER8*, *IRREGULAR XYLEM8*, and *IRREGULAR XYLEM9* are reported to be involved in glucuronoxylan biosynthesis (Lee et al., 2007; Yin et al., 2010). In addition to the regular cell wall polymers, callose, a  $\beta$ -1,3-glucan, which is deposited by



the callose synthase (glucan synthase like; Gsls) belongs to the GT48 family (Farrokhi et al., 2006; Muthamilarasan et al., 2015). Integration of new polymers into the cell wall through synergistic action of biosynthesis and wall loosening process is essential in order to maintain the integrity during the cell elongation process (Cosgrove, 2005). This loosening and reassembly is accomplished by the combined action of various degrading enzymes such as glycoside hydrolases (Buchanan et al., 2012; Glass et al., 2015; Wei et al., 2015), pectin lyases (Jiang et al., 2013), xyloglucan endotransglucosylases/hydrolases (XTH; Rose et al., 2002; Nishitani and Vissenberg, 2007), and cell wall loosening proteins such as expansins (Cosgrove, 2015; Marowa et al., 2016), and yieldins (Okamoto-Nakazato et al., 2001). In sorghum, a total of 12 Cesa and 36/37 Csl genes have been reported in previous studies (Paterson et al., 2009; Yin et al., 2009). Characterization of sorghum (1,3; 1,4)- $\beta$ -glucan biosynthetic gene subfamilies CslF and CslH showed that *CslF6* plays an important role in elongating cells while *CslH3* has a major role in cells that has stopped growth and started depositing storage compounds (Ermawar et al., 2015a). In a recent study, genes encoding cellulose, lignin, and glucuroarabinoxylan biosynthetic enzymes were dynamically expressed during the different development stages of sorghum (McKinley et al., 2016). The expansins and XTHs encoding genes were also shown to be differentially expressed in the growing stem internodes of sorghum. One of the glycosyl hydrolases gene families, endo-(1,4)- $\beta$ -glucanase (GH9) has been studied across 5 grass genomes and 24 members were reported from sorghum (Buchanan et al., 2012).

The focus of second-generation biofuel production from plant biomass is to utilize the sugars from lignocellulosic material for biofuels, in particular for bioethanol production. In order to utilize the lignocellulosic biomass for bioethanol production, the cell wall polysaccharides need to be separated from lignin, hydrolyzed by polysaccharide degrading enzymes to produce fermentable sugars, a process called saccharification (Lin and Tanaka, 2005). The presence of interlinked lignin around cell wall polysaccharides contributes to biomass recalcitrance by hindering the enzyme access to polysaccharides (Ermawar et al., 2015b). Separation of lignin from other cell wall polysaccharides requires pretreatment with concentrated acids at high temperatures. In addition, the presence of hydroxyl groups in the cellulose units allows intra and intermolecular hydrogen bonding which makes the structure more crystalline. Either decreasing the lignin content or reducing the cellulose crystallinity or both will improve saccharification efficiency. A comparative analysis of lignin biosynthesis related gene families have been done across plant kingdom including sorghum (Xu et al., 2009). In sorghum, several mutants (*bmr*, *brown midrib*, and *rg*, *red for green*) with reduced lignin content showed increase in saccharification and digestibility compared to control plants (Palmer et al., 2008; Xin et al., 2008; Saballos et al., 2009; Yan et al., 2012; Petti et al., 2013; Sattler et al., 2014). Among these *bmr* mutants, several loci have been identified which includes *bmr2* encoding 4-coumarate: coenzyme A ligase (4CL), *bmr6* encoding cinnamyl alcohol dehydrogenase (CAD), and *bmr12* and *bmr18* encoding caffeic acid O-methyltransferase (COMT) enzymes of monolignol pathways (Saballos et al., 2009; Sattler et al., 2009;

Scully et al., 2016). The sorghum biomass digestibility and saccharification efficiency can be further improved by targeting various genes involved in lignin biosynthesis coupled with genes that alter the cellulose crystallinity. Apart from lignin related gene families, Cesa, Csls, and Gsls are among the most studied cell wall related gene families in sorghum. As the research on function of cell wall genes in model crop *Arabidopsis* is advancing, it is now essential to identify and characterize the cell wall related genes in sorghum to engineer sorghum biomass for food, feed and biofuel and bioproduct applications.

The present study focuses on mining of publically available *S. bicolor* genome for identification and comprehensive analysis of gene families involved in the biosynthesis of cell wall biopolymers. In addition, various other gene families involved in degradation and reassembly of cell walls have also been analyzed. Further, phylogenetic analysis, physical mapping, and duplication analysis of identified genes have been performed in order to get insight into the relation among the genes and their origin. All the identified genes were also analyzed for the presence of SSR markers and miRNA target sites for molecular breeding and biotechnological applications. Publically available transcriptome datasets from various tissues were analyzed to study the expression pattern of these genes. Furthermore, to understand the expression pattern of cell wall related gene families under abiotic stress condition, differential expression analysis of exogenous abscisic acid (ABA), and polyethylene glycol (PEG) treated tissues were also performed. The identification and analysis of cell wall related gene families in the present study would help the research community in planning effective strategies for more efficient utilization of biomass for various applications.

## MATERIALS AND METHODS

### Data Retrieval and Identification of Cell Wall Related Gene Families

Publically available sequences of gene, protein, and chromosomes were downloaded from the Phytozome 11 database (<https://phytozome.jgi.doe.gov/pz/portal.html#>; Goodstein et al., 2012) for the identification and analysis of cell wall related gene families in sorghum. Protein sequences from other plants were downloaded from the Cell Wall Navigator database (Girke et al., 2004) to build the family specific HMM profile using HMMER v3.1b1 package (<http://www.ebi.ac.uk/Tools/hmmer/>). We first performed the multiple alignment of downloaded family specific sequences using Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) and saved the output alignment as \*.stockholm files. Using the family specific \*.stockholm alignment file as input for hmmbuild script we built the family specific HMM profiles (Data Sheet 1). Sorghum proteome was screened to identify the protein sequences related to various cell wall related families using HMMER with default parameters. All the identified proteins were screened for presence of their characteristic pfam domains. The successful candidate proteins were further verified for the presence of conserved

domains using NCBI's Conserved Domain Database (CDD; (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>; Marchler-Bauer et al., 2011). Additionally, *Arabidopsis* cell wall related proteins were used as query to search the sorghum genome using blastP with an *e*-value of  $10^{-5}$  and further validated with CDD search. The identified protein sequences from HMMER analysis were compared to blast identified proteins to prepare gene family specific non-redundant gene list. The coding and amino acid sequences of all the identified members were retrieved from sorghum genome dataset obtained from Phytozome database and used for the further analysis. The molecular weight and pI-values of all the identified proteins were calculated using online tool Compute pI/Mw ([http://web.expasy.org/compute\\_pi/](http://web.expasy.org/compute_pi/); Gasteiger et al., 2005).

### Phylogeny, Physical Mapping, and Duplication Analysis of Cell Wall Related Genes

The protein sequences of individual families were used for multiple sequence alignment using ClustalW program of MEGA v6 package (Tamura et al., 2013). Individual phylogenetic tree was constructed for the individual gene families with the MEGA v6 using neighbor-joining method. Bootstrap test was performed with 1000 iterations. To map physical locations of the identified cell wall related genes on sorghum chromosomes, their genomic coordinates along with chromosome number were retrieved from the file (.gff) downloaded from Phytozome database. The physical localization of genes was performed using the Mapchart 2.30 software (Voorrips, 2002). Furthermore, all the identified genes were analyzed for tandem duplications within the genome using the Plant Genome Duplication Database (<http://chibba.agtec.uga.edu/duplication/>) dataset (Lee et al., 2012).

### Identification of SSR Markers in Cell Wall Related Genes

Coding sequences of all the cell wall related genes were used for SSRs identification using microsatellite identification tool (MISA, <http://pgsc.ipk-gatersleben.de/misa/misa.html>). The criteria for SSR search was repeat stretches having a minimum of five repeat units for dinucleotide (DNRs), trinucleotide (TNRs), tetranucleotide (TtNRs), pentanucleotide (PNRs), and hexanucleotide (HNRs). Mononucleotide repeats (MNRs) were excluded from the analysis. The maximum distance between two markers in a compound microsatellite was set to 100.

### *In silico* Prediction of miRNA Target Sites on Cell Wall Related Genes

The identified cell wall related genes from individual families were analyzed for the presence of miRNA target sites using psRNATarget server (<http://plantgrn.noble.org/psRNATarget/>; Dai and Zhao, 2011). The maximum expectation values of 3.0 with other default parameters were used to perform the analysis.

## Expression Analysis of Cell Wall Related Genes at Various Developmental Stages of Sorghum

Publically available transcriptome datasets from different developmental stages of sorghum (stem, 20 days old leaves, vegetative meristem, floral meristem, spikelet, flowers, embryos, and seeds) were downloaded from NCBI's Short Read Archive (SRA) database (<http://www.ncbi.nlm.nih.gov/sra>). All the transcriptome datasets were mapped on cell wall related genes using the QSeq program of DNASTAR Lasergene package (<http://www.dnastar.com/t-nextgen-qseq.aspx>). For the mapping purpose, 520 gene sequences related to cell wall gene families were exclusively used as reference. Transcript abundance was visualized by MeV (<http://www.tm4.org/mev.html>) generated hierarchical clustered heat map for individual gene families using the self-normalized RPKM (Reads Per Kilobase per Million reads) values calculated by the QSeq program.

### Differential Gene Expression Analysis under Various Abiotic Stress Conditions

The role of identified cell wall related sorghum genes in abiotic stress conditions (exogenous ABA and PEG induced osmotic stress) in root and shoot was analyzed using publically available transcriptome datasets (Dugas et al., 2011). In brief, the published experiments were performed by germinating the *S. bicolor* BTx623 seeds and treating the seedlings on the 8th day after germination with 20  $\mu$ M ABA (dissolved in NaOH), 57.1  $\mu$ M NaOH (control for ABA), 20% PEG-8000, and Milli-Q (control for PEG treatment). After 27 h of treatment total RNA was extracted from the shoots and roots in three biological replicates and sequenced using the Illumina platform. Respective data sets of stress treated tissues along with controls were downloaded from SRA database of NCBI (Dugas et al., 2011). The expression pattern of cell wall genes was analyzed by using QSeq program of DNASTAR Lasergene package with self-normalized RPKM method. Fold change was calculated by using RPKM values of H<sub>2</sub>O and NaOH treated root and shoot tissues as controls for PEG and ABA, respectively. All the differentially expressed genes were analyzed for statistical significance using the Student's *t*-test with multiple-hypothesis testing at less than 0.05. Significantly differentially expressed genes (fold change  $\geq 2.0$ ,  $p < 0.05$ ) from different stress conditions were used to find commonly up or down regulated genes from root and shoot using the online tool Venny 2.1 (<http://bioinfogp.cnb.csic.es/tools/venny/>).

## RESULTS

### Identification of Cell Wall Related Genes from Sorghum

Cell wall related gene families have been shown to play crucial roles in various biological processes related to plant development, biotic and abiotic stress responses (Hamann, 2012; Lombard et al., 2014; Le Gall et al., 2015). Lignin biosynthetic gene families of sorghum have been analyzed elsewhere (Xu et al., 2009), hence, the present study focused on the gene families involved in various cell wall related processes such as polysaccharide synthesis and

reassembly and degradation (Girke et al., 2004). Additionally, previously unreported laccase genes that are involved in lignin biosynthesis, were also analyzed in the present study. All 47,205 protein-coding transcripts and proteins from publically available *S. bicolor* genome were downloaded and analyzed. HMMER search identified a total of 520 genes from 20 cell wall related gene families with an average of 26 genes per family (Table 1, Figure 1, and Supplementary Table 1). Among the analyzed gene families, expansin with 83 members was the largest gene family whereas rhamnogalacturonan I lyases (CAZy ID: PL4) was the smallest family with 6 members (Table 1 and Figure 1A). According to the CAZy distribution, the total identified genes were classified into glycosyl transferases (160), glycoside hydrolases (201), pectin/pectate lyases (16), carbohydrate esterases (35), auxiliary activity (25), and expansin (83) families (Table 1).

### I. Cell Wall Polysaccharide Biosynthetic Gene Families of Sorghum

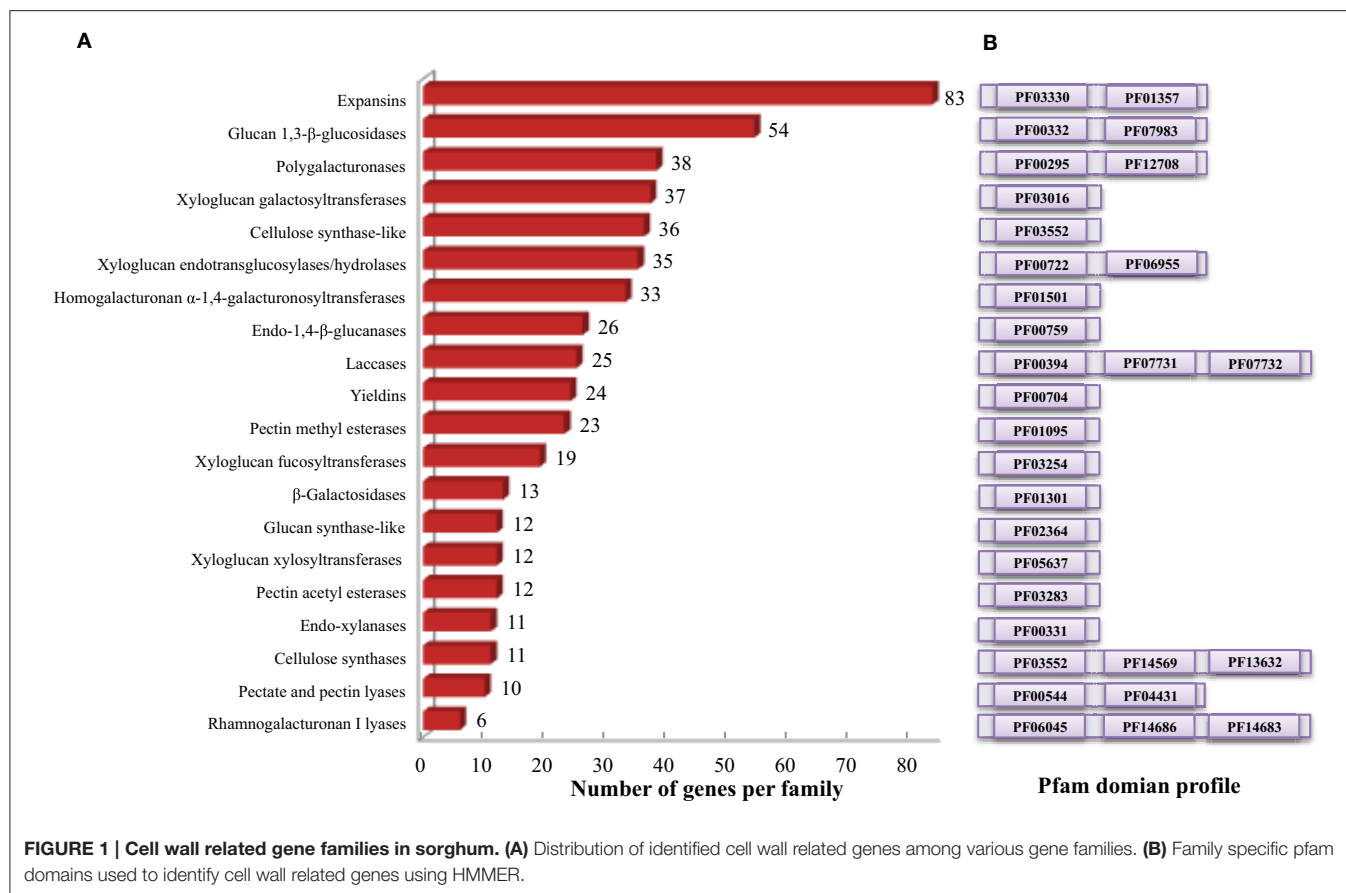
Plant cell wall biosynthesis is a complex process involving plethora of enzymes resulting in the biosynthesis of vast variety of cross linked cell wall polysaccharides. Majority of these enzymes belong to glycosyl transferases superfamily that is involved in the synthesis of cellulose, hemicellulose, pectin and callose.

### Cellulose biosynthetic genes

Genome-wide analysis of sorghum showed presence of 11 Cesa genes. In contrast to the 12 Cesa genes reported in the previous studies (Paterson et al., 2009; Yin et al., 2009), the present study identified only 11 SbCesa genes using conserved domain profiles based HMMER scanning of sorghum genome (Table 1 and Figure 1B). Further investigation showed, according to the updated sorghum genome (v3.1, phytozome), the previously predicted two Cesa genes (Sb03g004310.1 and Sb03g004320.1) are indeed a single gene (Sobic.003G049600.2). In addition, domain analysis of these 11 CESA proteins showed the presence of canonical cellulose synthase (CS, PF03552), zinc-binding RING-finger (PF14569), and glycosyl transferase 2 (PF13632) domains (Figure 1B). Majority of the CESAs (9/11) showed presence of all canonical domains, while Sobic.003G296400.1 showed lack of ZF domain and Sobic.010G183700.1 showed lack of ZF and GT2 domains. Eukaryotic Cesa genes were first cloned from cotton (Pear et al., 1996) and have been later reported from *Arabidopsis* (10), maize (12), poplar (18), and 14 in foxtail millet (Richmond and Somerville, 2000; Appenzeller et al., 2004; Djerbi et al., 2005; Muthamilarasan et al., 2015). Cluster analysis of CESA proteins found to be clustered with CSLD and CSLF proteins (Figure 2) consistent with earlier reports (Ermawar

**TABLE 1 | Details of identified cell wall related gene families in *Sorghum bicolor*.**

Substrates	Gene families	Abbre	CAZy ID	No. of genes
<b>POLYSACCHARIDE SYNTHESIS</b>				
Cellulose	Cellulose synthases	CESA	GT2	11
Hemicellulose	Cellulose synthase-like	CSL	GT2	36
	Xyloglucan xylosyltransferases and galactomannan gal-transferases	XXT	GT34	12
	Xyloglucan fucosyltransferases (MUR2)	XFT	GT37	19
	Xyloglucan galactosyltransferases (MUR3)	XGT	GT47	37
Pectin	Homogalacturonan $\alpha$ -1,4-galacturonosyltransferases	GAUT	GT8	33
Lignin	Laccases	LAC	AA1	25
Callose	Glucan synthase-like (Callose synthases)	GSL	GT48	12
<b>REASSEMBLY AND DEGRADATION</b>				
Cell wall loosening	Expansins	EXP		83
	Yieldins		GH18	24
	Xyloglucan endotransglucosylases/hydrolases	XTH	GH16	35
Glycoside hydrolases	Endo-1,4- $\beta$ -glucanases		GH9	26
	Endo-xylanases		GH10	11
	Glucan 1,3- $\beta$ -glucosidases		GH17	54
	Polygalacturonases	PGases	GH28	38
	$\beta$ -Galactosidases	BGAL	GH35	13
Pectin modifying	Pectate and pectin lyases		PL1	10
	Rhamnogalacturonan I lyases		PL4	6
	Pectin methyl esterases	PME	CE8	23
	Pectin acetyl esterases	PAE	CE13	12
Total	20			520



et al., 2015a). The clustering is due the presence of common conserved domains among cellulose synthase and cellulose synthase like family of proteins. Chromosomal distribution of Cesa genes showed presence of 4 and 3 genes on chromosome 2 and 1, respectively while remaining four are present on chromosome 3 (2 genes), 9 (1 gene), and 10 (1 gene; **Figure 4A**).

### Hemicellulose biosynthetic genes

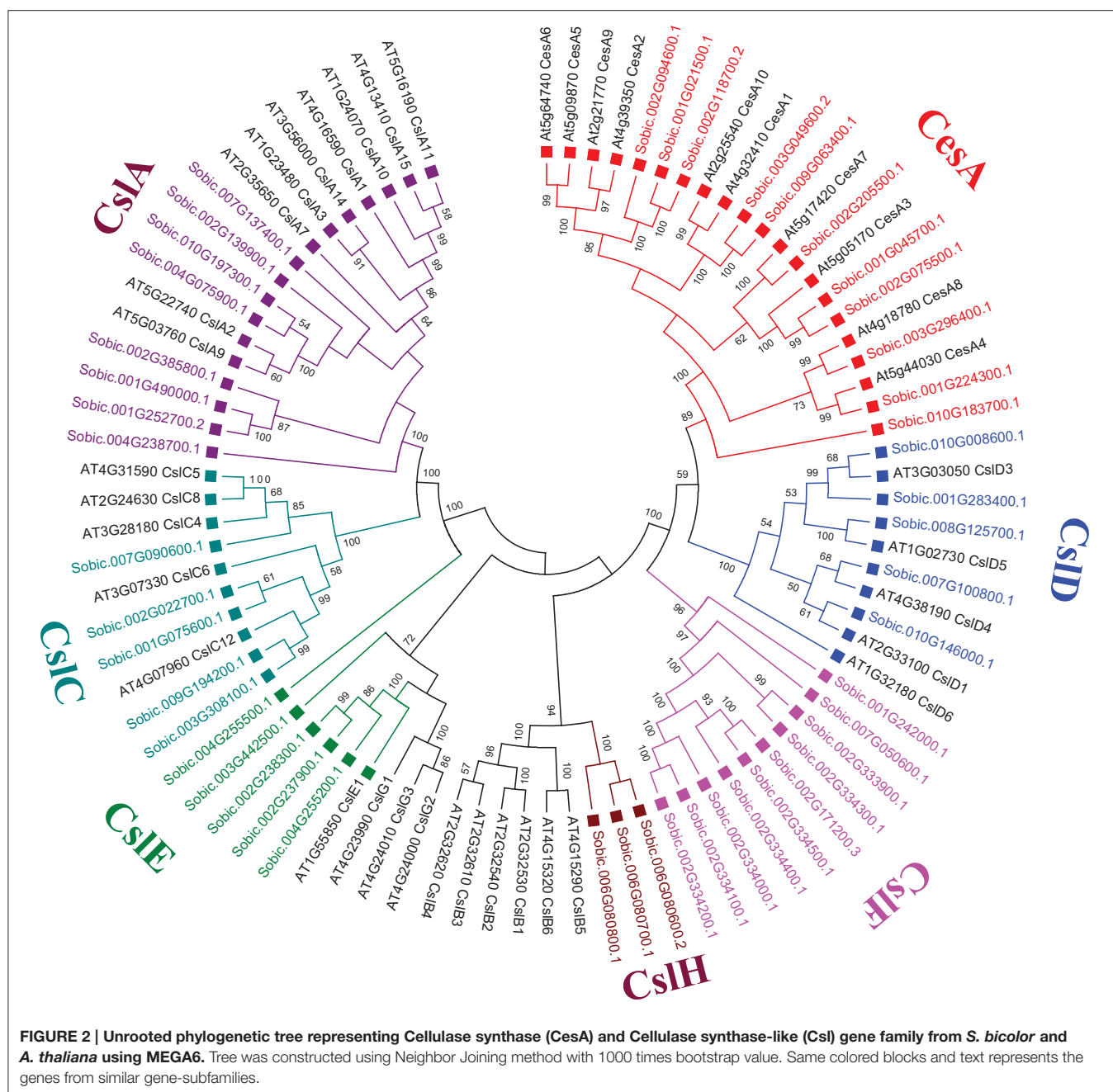
Sorghum genome showed presence of four hemicellulose biosynthetic enzyme families i.e., cellulose synthase like (Csl; GT2), xyloglucan xylosyltransferases (XXT; GT34), xyloglucan fucosyltransferases (MUR2, XFT; GT37), and xyloglucan galactosyltransferases (MUR3, XGT; GT47). A total of 104 genes representing, Csl (36), XXT (12), MUR2 (19), and MUR3 (37) gene families were identified (**Table 1**, **Figure 1A**, and Supplementary Table 1). Further, domain analysis of these genes revealed the presence of CS (PF03552) domain in CSL, presence of GT34 (PF05637) domain in XXT, presence of XG\_Ftase (PF03254) domain in MUR2, and exostosin (PF03016) domain in MUR3 family proteins (**Figure 1B**). Phylogenetic analysis of CSL proteins with *Arabidopsis* CSLs clustered them into 6 different sub-families namely, CSLA with 8, CSLC, CSLD, and CSLE with 5 each, CSLF with 10 and CSLH with 3 members (**Figure 2**). Phylogenetic analysis of other hemicellulose related gene families XXT (GT34) and XFT (GT37) showed their uniform clustering with *Arabidopsis*

homologs (**Figures 3A,B**). Phylogenetic analysis of sorghum XGT (GT47) members with *Arabidopsis* homologs further clustered them into 5 subfamilies, A, B, C, D, and E with 13, 7, 3, 4, and 10 members, respectively (**Figure 3F**). Physical mapping of Csl family genes showed its distribution over all chromosomes except Chr. 5 (**Figure 4A**). Majority of the Csl genes, almost one-third (13), were found to be present exclusively on Chr. 2. Physical mapping of GT34 family members showed their distribution over six chromosomes (Chr. 1, 2, 3, 4, 5, and 8) whereas GT37 members were found to be present on five chromosomes (Chr. 2, 4, 6, 8, and 10) with maximum of 8 genes on Chr. 4 (**Figure 4A**). Another hemicellulose specific family, GT47 members were found to be distributed on all the chromosomes except on Chr. 5. About one-third (12) of GT47 family members were found present on Chr. 1 (**Figure 4A**).

### Pectin biosynthetic genes

A total of 33 genes were identified as members of homogalacturonan  $\alpha$ -1,4-galacturonosyltransferases (GAUT), a GT8 family involved in pectin biosynthesis. Presence of glycosyl\_transferase\_8 domain (PF01501) in pfam analysis confirms the annotation of these genes as GT8 members (**Figure 1B**). Sorghum GAUT members were further clustered into 5 sub-families (A–E) based on phylogenetic analysis with *Arabidopsis* homologs (**Figure 3D**). Among sub-families, D



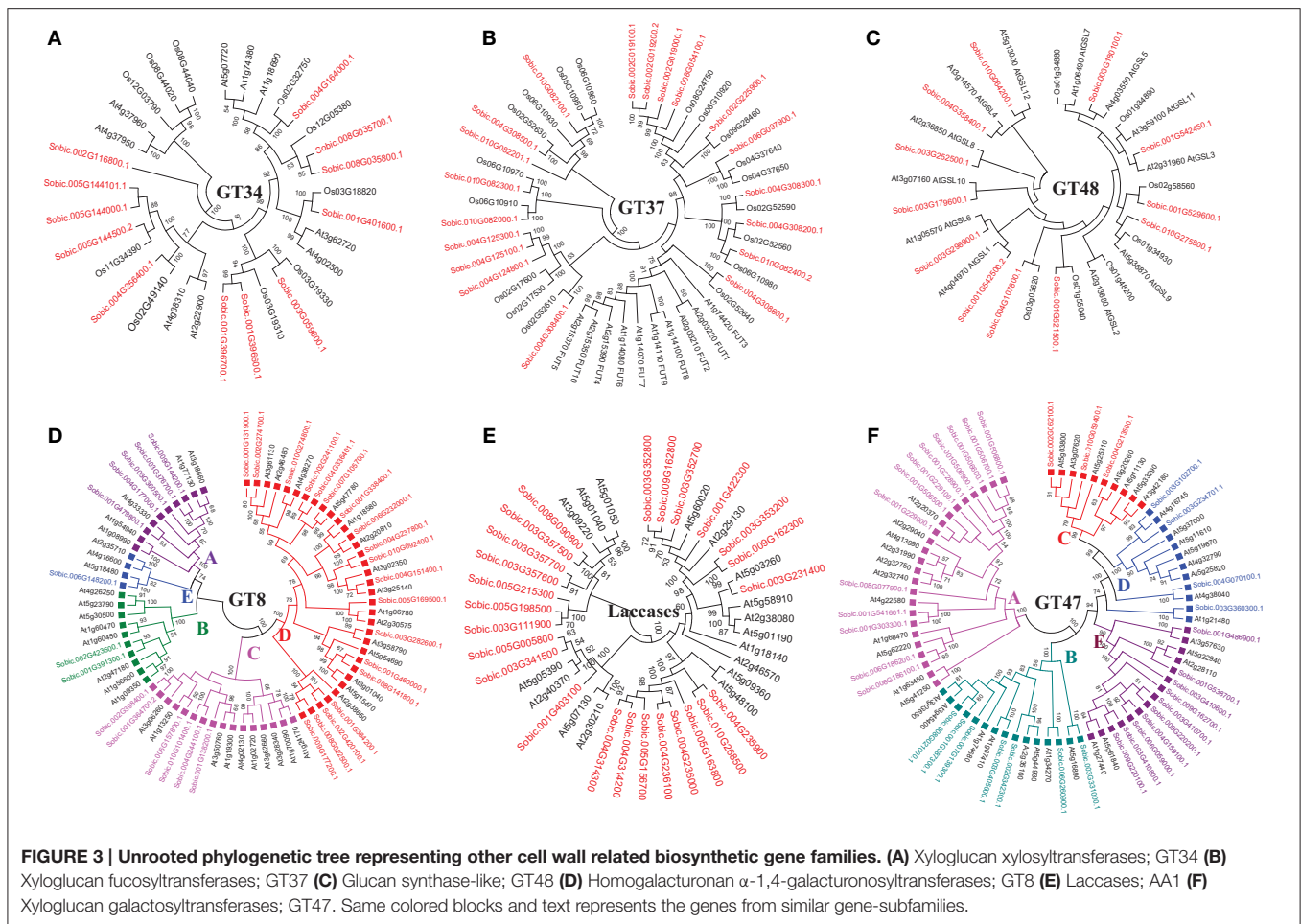


was the largest one with 19 members followed by C (6), A (5), B (2), and smallest sub-family E with a single member (Figure 3D). GT8 members were found to be distributed on all the chromosomes with 8 members exclusively present on Chr. 1 (Figure 4A).

#### Lignin biosynthetic genes

Sorghum lignin biosynthetic genes were analyzed along with other species (Xu et al., 2009) except laccase family. Here, we analyzed the laccase gene family that is involved in lignin biosynthesis. Laccases are among the CAZy AA1 class

of enzyme which play important role particularly in lignin metabolism. A total of 25 sorghum genes were identified as laccase family genes based on the presence of three copper containing conserved domains namely Cu<sub>2</sub>oxidase (PF00394), Cu<sub>2</sub>oxidase2 (PF07731), and Cu<sub>2</sub>oxidase<sub>3</sub> (PF07732) (Table 1 and Figure 1B). Phylogenetic analysis of laccase proteins showed alignment of some of them with *Arabidopsis* proteins whereas some of sorghum laccase proteins clustered in distinct clusters (Figure 3E). Physical mapping of laccase genes showed their presence on all the chromosomes except Chr. 2, 6, and 7 with a maximum of 9 genes on Chr. 3 (Figure 4A).



### Other cell wall biosynthetic genes

Glucan synthase-like (Gsl) gene family, a GT48 enzyme involved in biosynthesis of specialized polysaccharide callose, were found to have 12 members in sorghum genome based on the glucan\_synthase (PF02364) conserved domain (**Figure 1B**). Phylogenetic analysis of SbGSL proteins showed their uniform distribution with *Arabidopsis* and rice homologs (**Figure 3C**). Physical mapping of these genes showed their distribution limited to Chr. 1, 3, 4, and 10 with maximum 4 genes on Chr. 3 (**Figure 4A**).

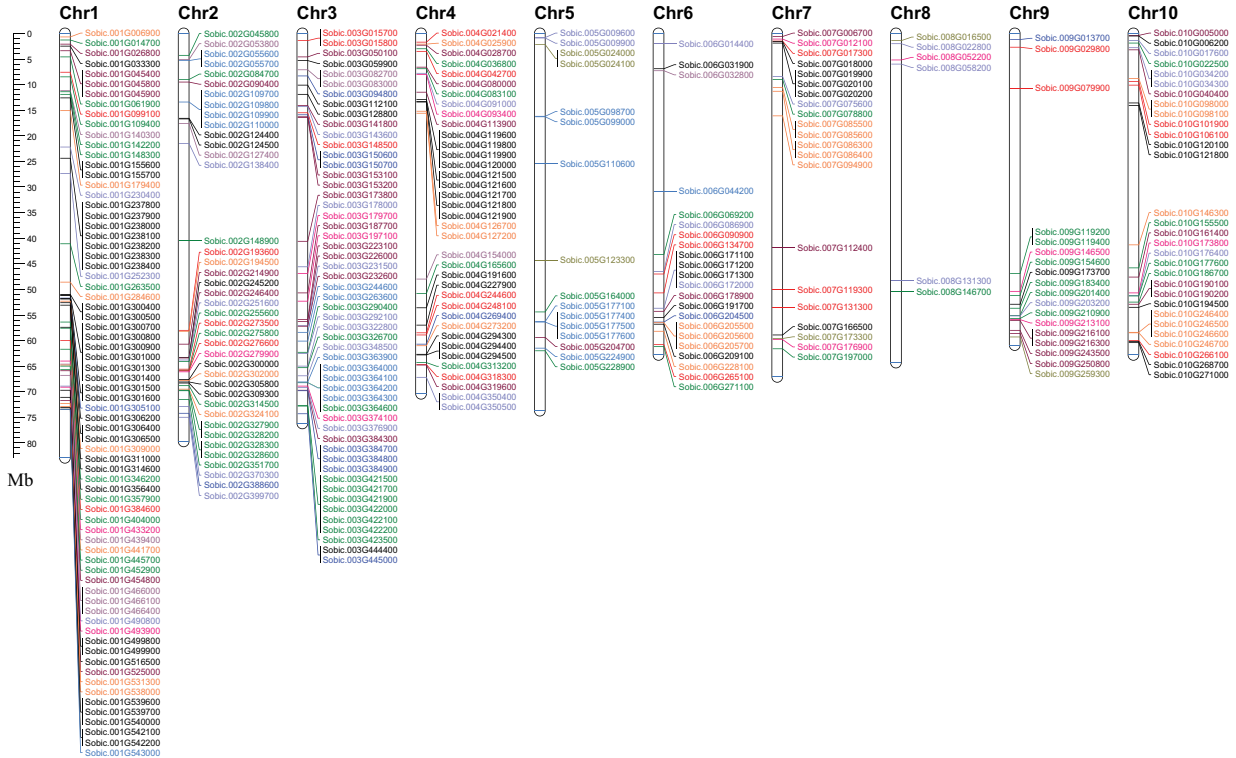
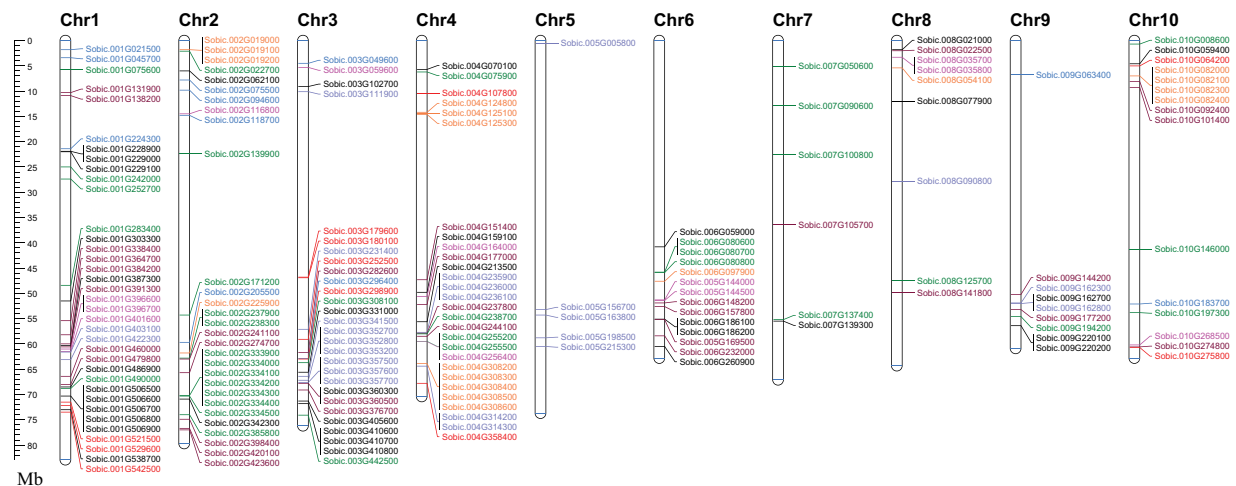
## II. Gene Families Involved in Cell Wall Reassembly and Degradation

Apart from cell wall biosynthetic enzymes, gene families involved in dynamic and complex cell wall extension and reassembly processes such as controlled degradation, loosening, and reassembly of cell wall polymers were also analyzed in the present study. A total of 335 genes were identified from 12 gene families that are involved in cell wall modifications.

### Cell wall loosening gene families

A total of 83 and 24 genes were identified as members of expansins and yieldins gene families which are primarily involved

in cell wall loosening (**Table 1**). The identification was based on the presence of conserved domains DPBB\_1 (PF03330) and pollen\_allerg\_1 (PF01357) observed in the pfam domain analysis (**Figure 1B**). Phylogenetic analysis of these proteins classified them into two major clusters, Expansin-A with 40 proteins and Expansin-B with 43 proteins (Supplementary Figure 1A). Further physical mapping and distribution analysis of these genes showed their presence on all the chromosomes except Chr. 5 and 8 (**Figure 4B**). Chromosome 1 was observed to have a maximum number of 34 expansins genes. Pfam domain profiling of another cell wall loosening protein family yieldins (GT18) showed presence of glyco\_hydro\_18 (PF00704) conserved domain (**Figure 1B**). Phylogenetic analysis of sorghum yieldins showed more similarity with rice yieldins proteins than *Arabidopsis* proteins indicating conservation of yieldins among monocots (Supplementary Figure 1B). Further, yieldins were found distributed on six sorghum chromosomes namely Chr. 1, 2, 3, 5, 6, and 7 with maximum of 8 genes on Chr. 5 (**Figure 4B**). Xyloglucan endotransglucosylases/hydrolases (XTH), another important cell wall loosening proteins, have dual role of hydrolyzing and extension of existing cell wall. A total of 35 sorghum genes were identified as XTH family members based on the observed conserved domains glyco\_hydro\_16 (PF00722) and



Genes with name in same color represents same gene family. A scale in the left represents length of chromosome in megabases (Mb).

**Glycoside hydrolases**

Among the identified cell wall modifying genes families in sorghum, there are 7 GH gene families (GH9, GH10, GH16, GH17, GH18, GH28, and GH35) with 201 genes (**Table 1**). Out of the identified 7 families, GH16 and GH18 have also been classified as cell wall loosening proteins. Among the GH families, GH17 (Glucan 1, 3- $\beta$ -glucosidases) was the largest with 54 genes



followed by GH28 (polygalacturonases) with 38 genes, GH9 (endo-1, 4- $\beta$ -glucanases) with 26 genes, GH35 ( $\beta$ -galactosidases) with 13 genes, and GH10 (endo-xylanases) with 11 genes (**Table 1**). Further, Pfam domain analysis showed the presence of conserved Glyco\_hydro\_9 (PF00759) in GH9, Glyco\_hydro\_10 (PF00331) and CBM\_4\_9 in GH10, Glyco\_hydro\_17 (PF00332) and X8 domain (PF07983) in GH17, Glyco\_hydro\_28 (PF00295) and pectate\_lyase\_3 (PF12708) in GH28 and Glyco\_hydro\_35 (PF01301) domains in GH35 (**Figure 1B**). Phylogenetic analysis of these gene families showed further classification of GH17 and GH28 into 5 and 7 sub-families, respectively whereas GH9, GH10, and GH35 showed clustering with *Arabidopsis* proteins without any sub-classification (Supplementary Figures 1D–G). GH17 has been further clustered into sub-families A with 13 genes, B with 13 genes, C with 16 genes, D with 9 genes, and sub-family E with 3 genes (Supplementary Figure 1F). GH28 has also been showed sub-clustering into 7 sub-families namely, A with 12 genes, C with 6 genes, D with 7 genes, E with 4 genes, F with 5 genes, and G with 2 genes (Supplementary Figure 1G) however, no sorghum proteins were found in sub-family B. Physical distribution of these genes on the sorghum genome showed the presence of GH9 genes on all chromosomes except Chr. 5 and 8, GH10 genes on five chromosomes (Chr. 1, 2, 3, 4, and 6), GH17 genes on all the chromosomes, GH28 genes on all the chromosomes except Chr. 8 and GH35 genes on all the chromosomes except Chr. 5 and 6 (**Figure 4B**).

### Pectin modifying enzymes

Among the identified cell wall related genes, 51 genes were classified as members of 4 gene families (2 pectin lyases and 2 pectin esterases) involved in pectin modification. Two pectin related lyases (PLs) namely pectate and pectin lyases (PL1) and rhamnogalacturonan I lyases (PL4) were found to have 10 and 6 genes respectively (**Table 1**). Conserved domain analysis of these proteins revealed the presence of Pec\_lyase\_C (PF00544) and Rhamno\_gal\_lyase (PF06045) domains in the PL1 and PL4, respectively (**Figure 1B**). An additional Pec\_lyase\_N (PF04431) domain was found in PL1 family while CBM\_like (PF14683) and Fn3\_3 (PF14686) domains were seen in PL4 family members. Phylogenetic analysis of PL1 family members along with *Arabidopsis* and rice homologs showed further clustering into three sub-families. PL1 sub family B was the largest one with 8 genes whereas sub family A and C was found to have one gene each (Supplementary Figure 1I). Phylogenetic analysis of PL4 genes showed more similarity to rice PL4 genes rather than *Arabidopsis* (Supplementary Figure 1J). Chromosomal distribution of these genes showed the presence of PL1 genes over six chromosomes (Chr. 1, 3, 4, 6, 8, and 10) whereas PL4 genes localization was limited to 4 chromosomes (Chr. 5, 7, 8, and 9) (**Figure 4B**).

Pectin esterases are another class of enzymes, which are involved in the cell wall reassembly. A total of 23 genes were identified as PME (CE8) homologs based on conserved pectin esterase (PF01095) domain whereas 12 genes were identified as PAE (CE13) family members based on pectin acetyl esterase domain (PF03283), respectively (**Table 1** and **Figure 1B**). Phylogenetic analysis of PME protein from sorghum showed

their even distribution with *Arabidopsis* PME proteins whereas sorghum PAE proteins showed more similarity to rice PAEs compared to *Arabidopsis* PAEs as expected (Supplementary Figures 1K,L). PME genes were distributed on all the 10 sorghum chromosomes with a maximum of 5 genes on Chr. 3 whereas PAE family members were limited to 6 chromosomes (Chr. 1, 2, 3, 4, 6, and 9) with a maximum of 7 genes on Chr. 3 (**Figure 4B**).

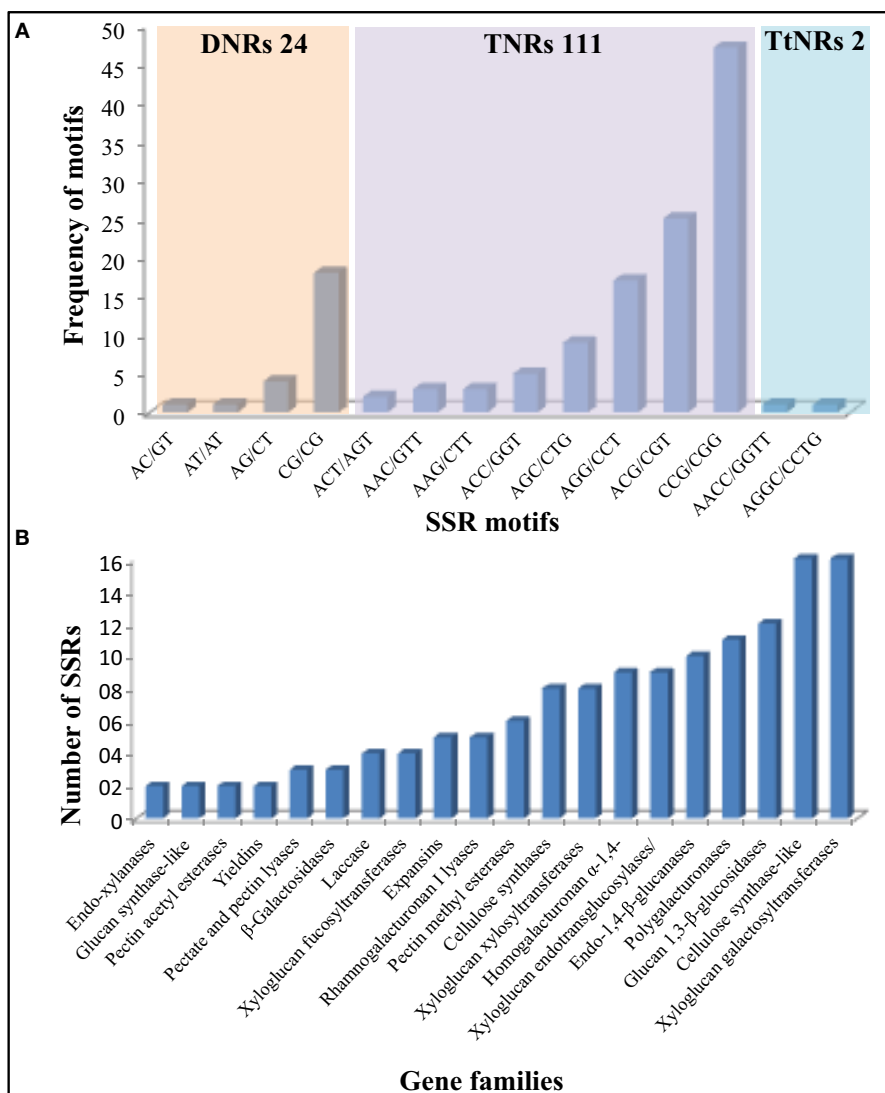
## Chromosomal Localization and Duplication Analysis

Chromosomal localization of identified cell wall related genes were performed on the 10 sorghum chromosomes using Mapchart 2.30 mapping software (**Figures 4A,B**). Approximately 65% (336) of cell wall related genes were present on 4 chromosomes namely, chromosome 1 with 22.3% (116), chromosome 3 with ~16% (83), chromosome 2 with 13.2% (69), and chromosome 4 with 13.1% (68). Remaining genes were found to be distributed on remaining six chromosomes with a minimum of 15 genes on chromosome 8 (**Figure 4**). Further, all the cell wall related families were analyzed for tandem duplication within the respective gene families to study their expansion. Out of 20 gene families analyzed, 56 tandem duplication events involving 169 genes were observed in 17 families (**Figures 4A,B**). No tandemly duplicated genes were observed in CesA, glucan synthase, and  $\beta$ -galactosidase gene families. Expansins gene family was observed to have highest number of tandem duplication events (14) involving 51 genes. Among other gene families, MUR3 (5 events/15 genes), XTH (5 events/13 genes), MUR2 (4 events/15 genes), GH28 (4 events/9 genes), yieldins (4 events/14 genes), laccases (4 events/10 genes), GH17 (3 events/10 genes), PAE (2 events/5 genes), and XXT with 3 events involving 6 genes were observed with significant number of duplications. Apart from this, Csl gene family was also found to have 2 duplication events involving 9 genes. Chromosome 1 was found to have a maximum number (14) of tandem duplications of the cell wall gene families, followed by chromosome 3 (9 duplications), 4 (8 duplications), and 2 with 6 duplications. Only one tandem duplication event among cell wall genes was observed on chromosome 8.

## Cell Wall Related Genes with SSR Markers in Sorghum

Microsatellites or SSR markers are short tandem DNA repeats which belongs to comparatively most efficient class of molecular markers with its genome wide distribution and high level of polymorphism. Expressed or coding sequence derived SSRs (ESSRs) have been reported to be comparatively more conserved than the genomic derived SSRs (Guo et al., 2006) which makes ESSRs as an important tool for marker assisted selection for various plant breeding programs. Considering the importance of cell wall related genes in developing sorghum mutants for biofuel applications, we analyzed all of the identified cell wall related genes for the presence of SSR markers. Out of 520 genes, 112 genes were identified with 137 SSRs (125 Simple and 6 compounds; **Figure 5A**, Supplementary Tables 2, 3). Among the identified SSRs, tri-nucleotide repeats (TNRs) were most





**FIGURE 5 | Details of identified SSR markers (137) in 520 cell wall related genes of sorghum. (A)** Marker distribution based on number and type of repeat motifs. **(B)** Gene family wise distribution of identified SSRs markers.

abundant with 111 occurrences followed by 24 DNRs. The identified SSRs were found to be present in all the 20 families analyzed with highest representation in Csl (16) and xyloglucan galactosyltransferases (16) gene families (**Figure 5B**).

### Cell Wall Related Genes with Putative miRNA Target Sites in Sorghum

MicroRNAs (miRNAs) are small and conserved non-coding RNA molecules which are known to regulate the gene expression at transcriptional and post-transcriptional levels. To understand the potential roles of miRNAs in regulating the cell wall related gene expression, all the 520 cell wall genes from various families were analyzed for the presence of miRNA target sites. A total of 10 genes were identified to have miRNA target sites out of which 6 belong to laccase (Sobic.001G422300.1,

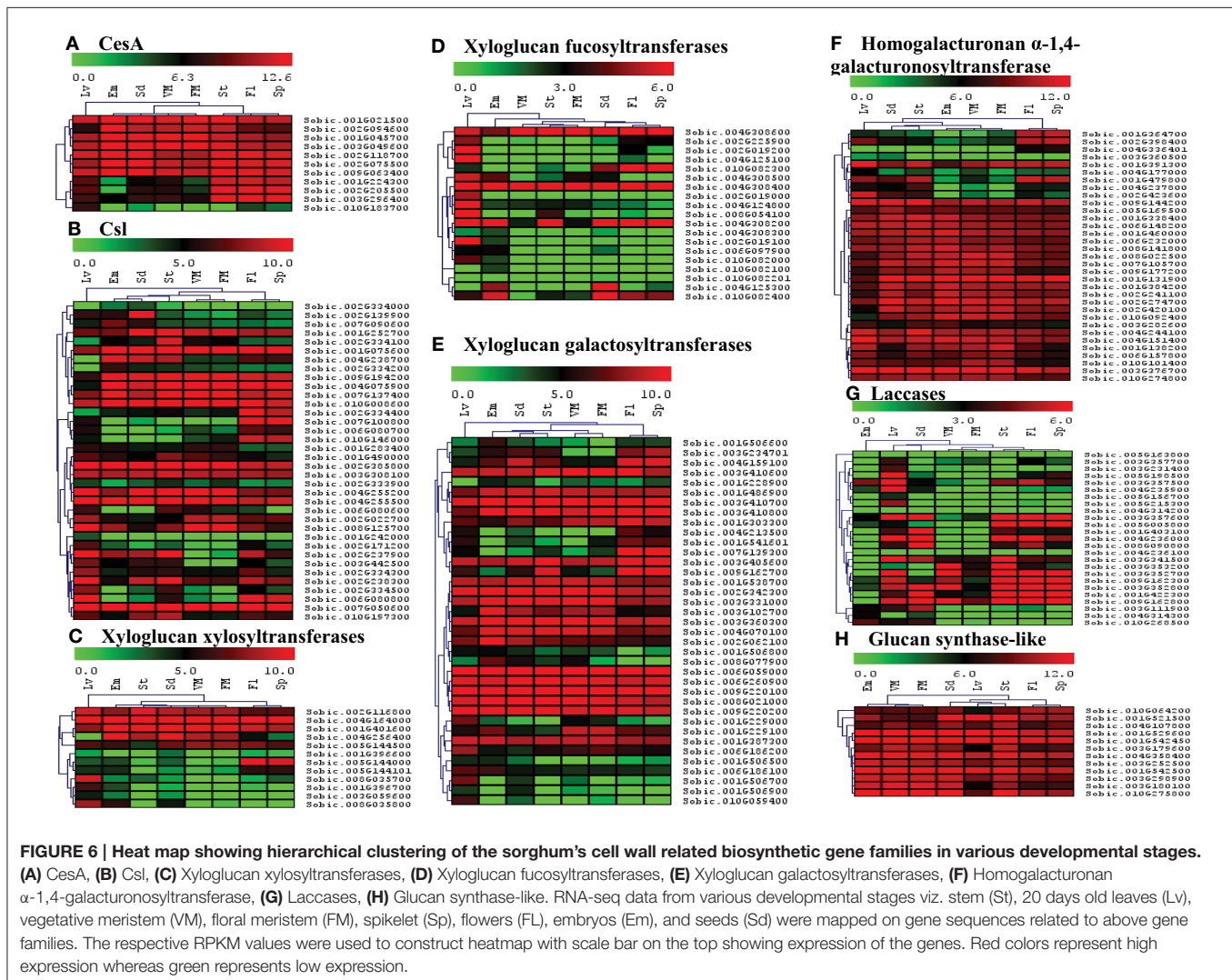
Sobic.003G352700.1, Sobic.003G352800.1, Sobic.003G353200.1, Sobic.005G198500.1, and Sobic.009G162800.1), 2 belong to Gsl (GT48) (Sobic.003G298900.1 and Sobic.004G107800.1) and one each to CesA (Sobic.003G049600.2) and Csl (Sobic.008G125700.1) (Supplementary Table 4). Six different miRNA families (miR156, miR164, miR397, miR528, miR5566, and miR6230) were identified to target these cell wall related genes.

### Expression Profile of Cell Wall Related Genes in Different Organs of Sorghum Plant

The availability of whole transcriptome data online presented an excellent opportunity to identify candidate genes that play key roles in specific organs during sorghum development. The

information on candidate genes can be further used to engineer cell walls in a cell/tissue specific manner to meet various industrial needs particularly in biofuel/feed industry. Publically available whole transcriptome datasets (Supplementary Table 5) were used to analyze spatial expression of the cell wall related genes in 8 different organs (leaves, embryo, seed, stem, spike, flower, vegetative as well as floral meristem) using RPKM values. The relative expression data was represented family wise using individual heat maps in order to better analyze the role of genes from each family (Figures 6, 7, Supplementary Table 6). In case of Cesa genes, 7 out of 11 genes were observed to have high expression in all analyzed organs, whereas, 3 genes were highly expressed in stem, flowers and spikes with moderate expression in leaves, seeds, and vegetative meristem (VM; Figure 6A). One Cesa gene (Sobic.010G183700) found to be expressed exclusively in leaves. A mixed pattern of expression was observed in Csl gene family with 11 genes showing high expression and 3 genes with very low expression in all the 8 tissues analyzed (Figure 6B). Among hemicellulose

biosynthetic genes, 12 xyloglucan xylosyltransferases genes were clustered in two main clusters, first with 5 genes having medium to high expression in all the 8 tissues, whereas second cluster of 7 genes with tissue specific expression mainly in leaves (Figure 6C). Xyloglucan fucosyltransferases gene family members, other than 3 genes (Sobic.004G308200, Sobic.004G308400, Sobic.004G308600) that showed higher expression in all the 8 tissues, expression of remaining genes was mostly limited to leaves (Figure 6D). Half (19) of the xyloglucan galactosyltransferases gene family, showed a higher expression level in all the tissues, whereas the other half (18) showed moderate to low expression in various tissues (Figure 6E). In homogalacturonan  $\alpha$ -1,4-galacturonosyltransferase gene family, 25 out of 33 genes showed high expression in all the tissues whereas remaining 8 genes showed tissue specific expression (Figure 6F). Lignin biosynthetic related laccase genes showed high expression in selective tissues like leaves, seeds, stem flower, and spikes (Figure 6G). Other than few laccase genes, most of the laccases were not expressed in embryo and meristematic tissues.



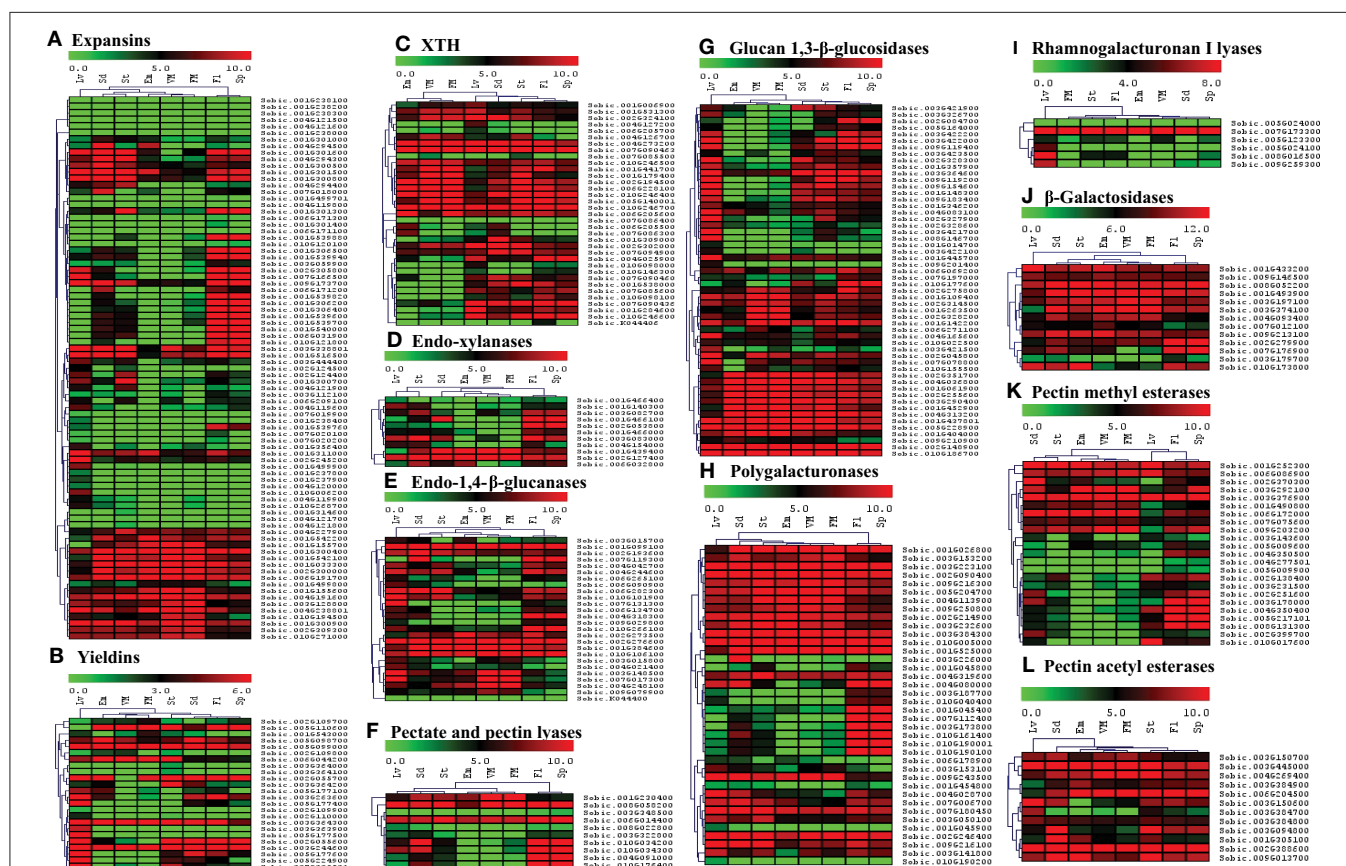
All the 12 genes of glucan synthase gene family showed high expression across all the tissue analyzed (**Figure 6H**).

Apart from polysaccharide biosynthetic gene families, the spatial expression of 12 gene families involved in degradation and reassembly was also analyzed. Among expansin genes, 16 genes from 2 clusters were observed with medium to high expression in all the tissues (**Figure 7A**). Remaining expansin genes mostly showed expression in flower, spike, leaves, seeds, and/or stem. Yieldins were observed to express consistently in leaves, whereas some of them showed high expression in all the other tissues (**Figure 7B**). Most of the XTH gene family showed moderate to high expression in almost all the tissues analyzed other than a cluster with no expression in embryo and meristematic tissues (**Figure 7C**). The five GH family genes (endo-xylanases, endo-1, 4- $\beta$ -glucanases, glucan 1, 3- $\beta$ -glucosidases, polygalacturonases and  $\beta$ -galactosidases) were highly expressed in leaves, stem, seed, flowers, and spikes apart from the clusters with high expression in all tissues (**Figures 7D,E,G,H,J**, respectively). Other than  $\beta$ -galactosidases, a significant proportion of genes from these families were not expressed in the embryo and meristematic

tissues. Among the 10 genes encoding pectate lyases, 3 showed expression in all the tissues whereas expression of remaining genes were limited to flower, spike, and seeds (**Figure 7F**). In the other pectin related rhamnogalacturonan I lyases gene family, only one gene showed consistent expression in all the tissues whereas remaining genes showed leaf specific expression (**Figure 7I**). Among the two families of esterases, gene encoding PME were majorly clustered into two clusters based on expression, first with moderate to high expression in almost all the tissues analyzed whereas second cluster with expression limited to tissues other than embryo and meristem (**Figure 7K**). All the PAE family genes showed medium to high expression in almost all the tissues analyzed (**Figure 7L**).

## Differential Expression Analysis of Cell Wall Related Genes under Different Abiotic Stress Conditions

Differential expression analysis of cell wall related genes under two abiotic stress treatments (ABA and osmotic stress) was



**FIGURE 7 |** Heat map showing hierarchical clustering of the sorghum's cell wall related gene families involved in reassembly and degradation in various developmental stages. (A) Expansins, (B) Yieldins, (C) XTH, (D) Endo-xylanases, (E) Endo-1, 4- $\beta$ -glucanases, (F) Pectate and pectin lyases, (G) Glucan 1, 3- $\beta$ -glucosidases, (H) Polygalacturonases, (I) Rhamnogalacturonan I lyases, (J)  $\beta$ -Galactosidases, (K) Pectin methyl esterases, (L) Pectin acetyl esterases. RNA-seq data from various developmental stages viz. stem (St), 20 days old leaves (Lv), vegetative meristem (VM), floral meristem (FM), spikelet (Sp), flowers (FL), embryos (Em), and seeds (Sd) were mapped on gene sequences related to above gene families. The respective RPKM values were used to construct heatmap with scale bar on the top showing expression of the genes. Red colors represent high expression whereas green represents low expression.

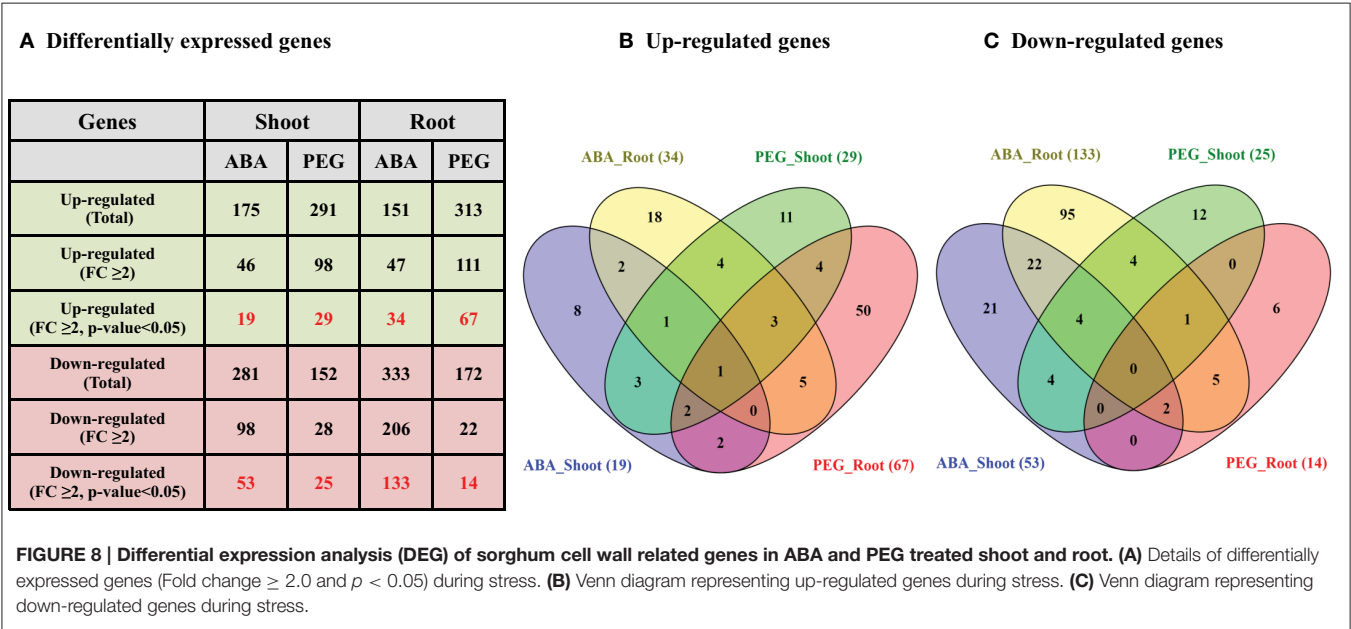
performed to analyze their response in seedling root and shoots. A total of 19 and 29 genes were found to be significantly up-regulated ( $FC \geq 2.0$  and  $p < 0.05$ ) in the sorghum shoots whereas 34 and 67 genes were found significantly up-regulated in the roots subjected to ABA and PEG treatment, respectively (Figures 8A,B, Supplementary Table 7, Supplementary Figures 2A,C, 3A,C). Similarly, 53 and 25 genes were significantly down-regulated in shoot whereas 133 and 14 genes were found down-regulated in root treated with ABA and PEG, respectively (Figures 8A,C, Supplementary Table 7, Supplementary Figures 2B,D, 3B,D). Relatively higher number of genes was down-regulated in the ABA treated shoot and root than PEG treatment. Comparative analysis of differentially expressed genes in root and shoot subjected to ABA and PEG treatment showed common up-regulation of 1 gene (Figure 8B) whereas no gene was found to be down-regulated in common (Figure 8C).

The ABA treated shoots showed up-regulation of polygalacturonases (4) whereas down-regulation of expansins (12), laccases (9), glucan 1,3- $\beta$ -glucosidases (8), and polygalacturonases (5) (Supplementary Figures 2A,B). The ABA treated roots showed up-regulation of glucan 1,3- $\beta$ -glucosidases (5), homogalacturonan  $\alpha$ -1,4-galacturonosyltransferases (4), laccases (3), and xyloglucan galactosyltransferases (3) whereas down-regulation of expansins (38), glucan 1,3- $\beta$ -glucosidases (18), laccases (11), Csls (11), and XTHs (9) (Supplementary Figures 3A,B). Large number of expansins showed down-regulation in ABA treated shoots as well as roots. Glucan 1,3- $\beta$ -glucosidases was the second most down-regulated family in ABA treated shoot as well as root. Similarly, most of the up-regulated genes in PEG treated roots mainly belong to expansins (21), XTHs (10), and glucan 1,3- $\beta$ -glucosidases (6) families whereas in PEG treated roots the down-regulated genes mainly belong to yieldins (3) and Csls (3) (Supplementary Figures 3C,D). The major up-regulated cell wall related gene

families in PEG-treated shoots were expansins (6), glucan 1, 3- $\beta$ -glucosidases (4), and XTHs (4). Similar to PEG-treated roots, yieldins were the most down-regulated gene family in PEG treated shoots (Supplementary Figures 2C,D).

DISCUSSION

Cell wall biogenesis is a dynamic process that involves synergistic action of multiple gene families that are involved in the biosynthesis as well as controlled degradation and reassembly of cell wall polymers. Broadly, glycosyl transferases are the major class of enzymes involved in cell wall polysaccharide biosynthesis while substrate specific glycoside hydrolases, pectin related lyases, various esterases together with cell wall loosening enzymes are responsible for cell wall extension. Among these families, cellulose synthase and cellulose synthase like genes are among the most studied in model, tree, and crop plants (Richmond and Somerville, 2000; Appenzeller et al., 2004; Djerbi et al., 2005; Muthamilarasan et al., 2015). Successful production of renewable biofuels and bioproducts from lignocellulose requires a comprehensive understanding on the genes involved in the biosynthesis of plant lignocellulosic material. A comprehensive report on cell wall related genes was missing in sorghum, which is an important food, fiber, bioproduct, and biofuel crop. Understanding the presence and distribution of cell wall related genes in sorghum would augment the plant breeding and biotechnological approaches to develop sorghum plants with altered cell wall composition for various industrial applications apart from crop improvement. In the present study, 520 genes from 20 cell wall related gene families have been identified and characterized *in silico*. Gene expression analysis of the identified genes was performed in different organs under normal and abiotic stress treated conditions to understand their role





in cell wall development and abiotic stress of sorghum. These candidate genes can be putative targets of reverse genetics for crop improvement apart from value addition to sorghum.

Lignocellulosic material deconstruction is important for bioethanol production from plant biomass. Current technology of bioethanol production involves separation of lignin from the lignocellulosic material, saccharification of sugars from wall polysaccharides and fermentation. Lignocellulosic based bioethanol production is technically and economically not competitive compared to fossil based gasoline with the existing conversion technologies. Further, the cost of bioethanol from plant biomass is higher (\$1.5/gal) than starch based (\$0.9/gal) bioethanol (<http://www.nrel.gov/docs/fy01osti/28893.pdf>). Improving the efficiency of the lignocellulosic biomass deconstruction, particularly separation of lignin from other wall polymers is essential to make bioethanol production economically feasible. Altering the biomass composition is essential to reduce biomass recalcitrance and improve the conversion technologies. Understanding the composition of lignocellulosic biomass and the genes involved in the lignocellulosic biomass facilitates the biomass engineering to improve the conversion efficiency.

Lignocellulosic material is mainly composed of cellulose, hemicellulose, lignin, and pectin. Cellulose is a linear homopolymer of  $\beta$ -1–4 linked glucose molecules occupying 30–40% of cell wall weight. CesA genes were first identified in bacteria and later in cotton followed by other plants (Saxena et al., 1990; Pear et al., 1996). The cellulose is synthesized by plasma membrane localized cellulose synthase complexes (CSCs) composed of multiple CESA proteins that produce individual glucan chains (Persson et al., 2007; Kumar and Turner, 2015). The individual glucan chains form a cellulose microfibrils and several microfibrils form a cellulose fiber hence the number of cellulose synthases present in individual species is important to understand the cellulose biosynthetic process. Moreover, cellulose synthases serve specific roles in plant development as they have cell, tissue, and developmental specific roles (Taylor et al., 2000; Mendu et al., 2011a). Hence there are multiple CesAs in each species and the number of genes varies based on the plant species; *Arabidopsis* (10), maize (12), poplar (18), and foxtail millet (14) (Richmond and Somerville, 2000; Appenzeller et al., 2004; Djerbi et al., 2005; Muthamilarasan et al., 2015). In the present study we identified 11 CesA genes in contrast to the reports of 12 CesA genes in sorghum, which is due to the removal of errors in the updated sorghum genome assemblies (Table 1, Figure 1). Though the number of CesAs present in sorghum is known, it is important to study the role of individual CesAs in primary and secondary cell wall biosynthesis to modify the biomass composition in specific organ or tissue.

Hemicellulose composition and biosynthesis is complex as they are composed of branched polysaccharides compared to homopolymeric cellulose. Hemicelluloses play an important role in cell wall polymer cross-linking and help in maintaining the cell wall integrity and strength. The hemicellulose content and composition is different in monocot and dicot plants. The cell walls of monocots such as sorghum contain 20–50% of hemicelluloses that makes it an attractive source for pentose

sugars (Welker et al., 2015). Bioethanol production from pentose sugars apart from hexoses is currently being heavily investigated (Unrean and Srienc, 2010). Understanding the hemicellulose biosynthesis and genes involved in the biosynthetic process will help to alter the biomass composition for easy deconstruction as well as to improve hexose to pentose ratio. The hemicellulose biosynthesis genes have not been studied very well other than cellulose synthase like genes. In the present investigation, the large family of sorghum Csls has been classified into 8 different groups (CslA to H) based on the phylogenetic studies (Ermawar et al., 2015a; Figure 2). Particularly, two clusters, CslF and CslH were observed be unique to sorghum with no *Arabidopsis* homologs, which is in agreement with the previous reports of these clusters as grass specific (Paterson et al., 2009; Ermawar et al., 2015a) while no sorghum Csl genes were clustered in Clusters CslB and CslG (Figure 2). Similar clustering of SbCsl genes has been reported previously in the sorghum draft genome report (Paterson et al., 2009). Apart from Csls, we also identified additional three hemicellulose gene families including xyloglucan xylosyltransferases (GT34), xyloglucan fucosyltransferases (GT37), and xyloglucan galactosyltransferases (GT47). Homogalacturonan  $\alpha$ -1,4-galacturonosyltransferases (GT8), a pectin biosynthesis related gene family was also observed as one of the sorghum's big cell wall biosynthetic gene families (Table 1, Figure 1). Pectin molecules play an important role in cell adhesion and contributes for biomass recalcitrance due to extensive interlinks with other cell wall polymers. Overall, information on cell wall biosynthetic genes will help to design customized biomass production for economical production of biofuels and bioproducts from sorghum. Apart from easier deconstruction and saccharification, enhancing the total sugars in the walls will help to improve the cost effectiveness of bioethanol production from sorghum biomass.

Cell wall biosynthesis is dynamic; it allows cell elongation while maintaining the wall integrity to withstand the internal turgor pressure. The degradation/assembly mechanism plays important roles in the cell wall building process, wall strength and integrity. Altering the process of wall degradation/assembly process will influence the cell wall deconstruction/digestibility hence identification and characterization of genes involved in cell wall degradation/assembly is important. The degradation/assembly related genes identified in this study has been distributed in to 3 cell wall loosening related gene families, 6 family of glycoside hydrolases, 2 pectin related lyases as well as 2 pectin related esterase gene families. The conserved domain analysis of the cell wall related gene families (Figure 1) along with clusters obtained from phylogenetic analysis with *Arabidopsis* and rice proteins suggests the evolutionary conserved nature of these proteins (Supplementary Figure 1). Physical mapping revealed presence of approximately 65% of cell wall related genes mainly confined to chromosomes 1–4 (1 with 22.3%, chromosome 3 with ~16%, chromosome 2 with 13.2%, and chromosome 4 with 13.1%; Figure 4, Supplementary Table 1). These chromosomes with hotspot of cell wall genes can be targeted in breeding and crop improvement programs to alter the cell wall composition. Further, 56 tandem duplication events observed in these genes were found to be distributed

across all the 10 chromosomes with maximum duplication observed on first 4 chromosomes. The excessive duplications observed on the chromosomes 1–4 could be the possible reason of presence of ~65% genes on these chromosomes. Further, a total of 137 SSR markers were found on ~22% cell wall related genes with highest representation in Csl (16) and xyloglucan galactosyltransferases (16) gene families (**Figure 5**, Supplementary Tables 2, 3). Among these, TNRs with ~81% share are the most abundant SSRs which are in agreement with the previous reports of TNRs abundance in plants. These molecular markers will help the breeding programs for selection of genes in a breeding population or introgression of a specific cell wall related genes. Further, *in silico* analysis for the presence of miRNA targets revealed presence of miR156, miR164, miR397, miR528, miR5566, and miR6230 target sites in 10 independent cell wall related genes (Supplementary Table 4). Three of these miRNA families viz. miR156, miR164, and miR528 have been reported to be differentially expressed in stem and leaves during sugar accumulation in sweet sorghum (Yu et al., 2015). Further, Yu et al. (2015) reported miR164 and miR528 as stem specific miRNA whereas miR156 was up-regulated in the leaves at dough stage. Over-expressed miR156 has been reported to cause the *Corngress1* (*Cg1*) phenotype in maize (Chuck et al., 2007). Further, four of the six sorghum laccase family genes found to have target site of miR397 and showed differential expression during the drought stress conditions (Hamza et al., 2016) indicating a potential change in the sorghum cell wall composition under stress.

The cell wall composition and gene expression varies among different tissues of the plant (stem, root, leaves, etc.) and among the cell types within a tissue (i.e., epidermal, xylem, phloem, fiber cells, etc.; Hatfield et al., 1999; McKinley et al., 2016). Expression analysis across different tissues will provide important insight into the role of cell wall related genes in that particular tissue. In the present study, we found a differential expression of genes among different tissues (**Figures 6, 7**). This analysis provides information on the tissue specific target genes for bioengineering purposes. A recent study of sorghum gene expression in pre- and post-anthesis stages of stem internodes showed differential expression of genes involved in growth, cell wall development and stem sugar accumulation (McKinley et al., 2016). CesA, Csls, callose synthases, XTHs, glucuroarabinoxylan biosynthetic genes, expansins, glucosyl hydrolases, pectin lyases/esterases, and lignin biosynthetic genes were among the major differentially expressed cell wall related genes (McKinley et al., 2016). In the present study, expression analysis of sorghum cell wall related genes showed that most of the genes from CesA, xyloglucan galactosyltransferases, homogalacturonan  $\alpha$ -1,4-galacturonosyltransferase, glucan synthase-like, glucan 1,3- $\beta$ -glucosidases, polygalacturonases,  $\beta$ -Galactosidases, and pectin acetyl esterases families expressed in all the stages studied. Remaining cell wall gene families showed genes with either stage specific to ubiquitously expressed genes or both.

Environmental conditions including temperature, drought, osmotic, and salinity, etc., have been shown to affect the gene expression and crop productivity (Tenhaken, 2014; Wang et al.,

2016). With the fast changing environmental conditions across the globe, studying the effect of stress on plants is important. In addition, to avoid food/fuel competition, the biofuel crops were advocated to be grown on marginal lands with limited irrigation and minimal input. Upon exposure to these adverse environmental conditions, the plants alter their gene expression and biochemical metabolism to survive in these conditions including cell wall composition. Most important component of the cell wall that adds enormous cost of bioethanol production is lignin. It has been reported that abiotic stress results in increased lignin content in plants (Moura et al., 2010). This results in increased cost of bioethanol production hence there is a need to develop bioenergy crops that do not accumulate higher lignin when grown in marginal lands with limited irrigation and low inputs. A better understanding of the cell wall gene expression under abiotic stress is important to design strategies to produce crops in marginal lands with less lignin accumulation. Analysis of sorghum transcriptome under abiotic stress showed differential expression of significant number of cell wall related genes (**Figure 8**). Comparatively, root was observed to have more altered expression of cell wall genes compared to shoot. Among the differentially expressed gene families, expansins, laccases, and glucan 1, 3- $\beta$ -glucosidases showed down-regulation in ABA treated root and shoot (Supplementary Figures 2B, 3B). Similarly, following PEG treatment, expansins, and XTHs were among up-regulated genes in root as well as shoot whereas yieldins were among the highly down-regulated genes in both the tissues (Supplementary Figures 2C,D, 3C,D). Since most of the cell wall related gene families are with multiple genes and each with either specific or redundant function, there is a need to characterize function of individual genes in order to develop a fine annotation of their function in normal growth and development as well as under abiotic stress conditions.

## CONCLUSIONS

Comprehensive information on cell wall related genes would facilitate biosynthetic pathway engineering for enhanced biomass production as well as efficient deconstruction and saccharification. Lignin content and cellulose crystallinity contribute to the poor separation and saccharification, which are the biggest hurdles in the cost efficient utilization of sorghum biomass for biofuel production. Here we have identified various cell wall related gene families and analyzed the gene expression pattern but the functional role of the individual genes is still not known. Cell wall related gene mutations in sorghum showed higher saccharification efficiency and are being used for animal feed hence further analysis and functional characterization will lead to development of more efficient sorghum lines for animal feed, biofuel and bioproduct industries. Apart from this, analyzing the cell wall composition of sorghum under abiotic stress conditions and their correlation with differentially expressed genes will also shed light on the mechanism involved in regulation of cell wall biosynthesis and degradation. The present study analyzed the gene expression of sorghum seedlings

exposed to abiotic stress, which provides valuable information, however a detailed study at different developmental stages that are critical for biomass harvest will provide information necessary to manipulate the biomass through plant breeding and genetic engineering. Overall, the comprehensive information developed in the present study can be used in expanding target genes as well as developing better strategies for the future sorghum crop improvement programs.

## AUTHOR CONTRIBUTIONS

KR designed the work, performed the analysis and wrote the manuscript. ST, VB, CC, and TD helped with bioinformatics analysis, prepared figures and wrote the manuscript. VM

conceived the idea, designed work and wrote the manuscript. All the authors have read and approved the manuscript.

## ACKNOWLEDGMENTS

This work was supported by the Department of Plant & Soil Sciences, Texas Tech University, the USDA/Agricultural Research Service (Ogallala Aquifer Initiative), and USDA-FAS.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.01287>

## REFERENCES

- Appenzeller, L., Doblin, M., Barreiro, R., Wang, H., Niu, X., Kollipara, K., et al. (2004). Cellulose synthesis in maize: isolation and expression analysis of the cellulose synthase (CesA) gene family. *Cellulose* 11, 287–299. doi: 10.1023/B:CELL.0000046417.84715.27
- Boerjan, W., Ralph, J., and Baucher, M. (2003). Lignin biosynthesis. *Annu. Rev. Plant Biol.* 54, 519–546. doi: 10.1146/annurev.arplant.54.031902.134938
- Buchanan, M., Burton, R. A., Dhugga, K. S., Rafalski, A. J., Tingey, S. V., Shirley, N. J., et al. (2012). Endo-(1,4)-beta-glucanase gene families in the grasses: temporal and spatial co-transcription of orthologous genes. *BMC Plant Biol.* 12:235. doi: 10.1186/1471-2229-12-235
- Burton, R. A., Gidley, M. J., and Fincher, G. B. (2010). Heterogeneity in the chemistry, structure and function of plant cell walls. *Nat. Chem. Biol.* 6, 724–732. doi: 10.1038/nchembio.439
- Byrt, C. S., Grof, C. P. L., and Furbank, R. T. (2011). C4 plants as biofuel feedstocks: optimising biomass production and feedstock quality from a lignocellulosic perspective. *J. Integr. Plant Biol.* 53, 120–135. doi: 10.1111/j.1744-7909.2010.01023.x
- Chen, X.-Y., and Kim, J.-Y. (2009). Callose synthesis in higher plants. *Plant Signal. Behav.* 4, 489–492. doi: 10.4161/psb.4.6.8359
- Chuck, G., Cigan, A. M., Saetern, K., and Hake, S. (2007). The heterochronic maize mutant Corngrass1 results from overexpression of a tandem microRNA. *Nat. Genet.* 39, 544–549. doi: 10.1038/ng2001
- Cosgrove, D. J. (2005). Growth of the plant cell wall. *Nat. Rev. Mol. Cell Biol.* 6, 850–861. doi: 10.1038/nrml1746
- Cosgrove, D. J. (2015). Plant expansins: diversity and interactions with plant cell walls. *Curr. Opin. Plant Biol.* 25, 162–172. doi: 10.1016/j.pbi.2015.05.014
- Dai, X., and Zhao, P. X. (2011). psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res.* 39, W155–W159. doi: 10.1093/nar/gkr319
- Del Bem, L. E., and Vincentz, M. G. (2010). Evolution of xyloglucan-related genes in green plants. *BMC Evol. Biol.* 10:341. doi: 10.1186/1471-2148-10-341
- Djerbi, S., Lindskog, M., Arvestad, L., Sterky, F., and Teeri, T. T. (2005). The genome sequence of black cottonwood (*Populus trichocarpa*) reveals 18 conserved cellulose synthase (CesA) genes. *Planta* 221, 739–746. doi: 10.1007/s00425-005-1498-4
- Dugas, D. V., Monaco, M. K., Olsen, A., Klein, R. R., Kumari, S., Ware, D., et al. (2011). Functional annotation of the transcriptome of *Sorghum bicolor* in response to osmotic stress and abscisic acid. *BMC Genomics* 12:514. doi: 10.1186/1471-2164-12-514
- Ermawar, R. A., Collins, H. M., Byrt, C. S., Betts, N. S., Henderson, M., Shirley, N. J., et al. (2015a). Distribution, structure and biosynthetic gene families of (1,3;1,4)-beta-glucan in *Sorghum bicolor*. *J. Integr. Plant Biol.* 57, 429–445. doi: 10.1111/jipb.12338
- Ermawar, R. A., Collins, H. M., Byrt, C. S., Henderson, M., O'Donovan, L. A., Shirley, N. J., et al. (2015b). Genetics and physiology of cell wall polysaccharides in the model C4 grass, *Setaria viridis* spp. *BMC Plant Biol.* 15:236. doi: 10.1186/s12870-015-0624-0
- Farrokhi, N., Burton, R. A., Brownfield, L., Hrmova, M., Wilson, S. M., Bacic, A., et al. (2006). Plant cell wall biosynthesis: genetic, biochemical and functional genomics approaches to the identification of key genes. *Plant Biotechnol. J.* 4, 145–167. doi: 10.1111/j.1467-7652.2005.00169.x
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. E., Wilkins, M. R., Appel, R. D., et al. (2005). “Protein identification and analysis tools on the ExPASy Server,” in *The Proteomics Protocols Handbook*, ed J. M. Walker (Totowa, NJ: Humana Press), 571–607.
- Girke, T., Lauricha, J., Tran, H., Keegstra, K., and Raikhel, N. (2004). The cell wall navigator database. A systems-based approach to organism-unrestricted mining of protein families involved in cell wall metabolism. *Plant Physiol.* 136, 3003–3008. doi: 10.1104/pp.104.049965
- Glass, M., Barkwill, S., Unda, F., and Mansfield, S. D. (2015). Endo-beta-1,4-glucanases impact plant cell wall development by influencing cellulose crystallization. *J. Integr. Plant Biol.* 57, 396–410. doi: 10.1111/jipb.12353
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., and Fazo, J. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186. doi: 10.1093/nar/gkr944
- Guo, W., Wang, W., Zhou, B., and Zhang, T. (2006). Cross-species transferability of G. arboreum-derived EST-SSRs in the diploid species of *Gossypium*. *Theor. Appl. Genet.* 112, 1573–1581. doi: 10.1007/s00122-006-0261-y
- Hamann, T. (2012). Plant cell wall integrity maintenance as an essential component of biotic stress response mechanisms. *Front. Plant Sci.* 3:77. doi: 10.3389/fpls.2012.00077
- Hamza, N. B., Sharma, N., Tripathi, A., and Sanan-Mishra, N. (2016). MicroRNA expression profiles in response to drought stress in *Sorghum bicolor*. *Gene Expr. Patterns* 20, 88–98. doi: 10.1016/j.gep.2016.01.001
- Harris, P., and Trethewey, J. K. (2010). The distribution of ester-linked ferulic acid in the cell walls of angiosperms. *Phytochem. Rev.* 9, 19–33. doi: 10.1007/s11101-009-9146-4
- Hatfield, R. D., Wilson, J. R., and Mertens, D. R. (1999). Composition of cell walls isolated from cell types of grain sorghum stems. *J. Sci. Food Agric.* 79, 891–899.
- Jiang, J., Yao, L., Miao, Y., and Cao, J. (2013). Genome-wide characterization of the pectate lyase-like (PLL) genes in *Brassica rapa*. *Mol. Genet. Genomics* 288, 601–614. doi: 10.1007/s00438-013-0775-3
- Kumar, M., and Turner, S. (2015). Plant cellulose synthesis: CESA proteins crossing kingdoms. *Phytochemistry* 112, 91–99. doi: 10.1016/j.phytochem.2014.07.009
- Lee, C., Zhong, R., Richardson, E. A., Himmelsbach, D. S., McPhail, B. T., and Ye, Z. H. (2007). The PARVUS gene is expressed in cells undergoing secondary wall thickening and is essential for glucuronoxylan biosynthesis. *Plant Cell Physiol.* 48, 1659–1672. doi: 10.1093/pcp/pcm155
- Lee, T.-H., Tang, H., Wang, X., and Paterson, A. H. (2012). PGDD: a database of gene and genome duplication in plants. *Nucleic Acids Res.* 41, D1152–D1158. doi: 10.1093/nar/gks1104
- Le Gall, H., Philippe, F., Domon, J.-M., Gillet, F., Pelloux, J., and Rayon, C. (2015). Cell wall metabolism in response to abiotic stress. *Plants* 4:112. doi: 10.3390/plants4010112



- Lerouxel, O., Cavalier, D. M., Liepman, A. H., and Keegstra, K. (2006). Biosynthesis of plant cell wall polysaccharides - a complex process. *Curr. Opin. Plant Biol.* 9, 621–630. doi: 10.1016/j.pbi.2006.09.009
- Lin, Y., and Tanaka, S. (2005). Ethanol fermentation from biomass resources: current state and prospects. *Appl. Microbiol. Biotechnol.* 69, 627–642. doi: 10.1007/s00253-005-0229-x
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, D490–D495. doi: 10.1093/nar/gkt1178
- Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., et al. (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* 39, D225–D229. doi: 10.1093/nar/gkq1189
- Marowa, P., Ding, A., and Kong, Y. (2016). Expansins: roles in plant growth and potential applications in crop improvement. *Plant Cell Rep.* 35, 949–965. doi: 10.1007/s00299-016-1948-4
- McKinley, B., Rooney, W., Wilkerson, C., and Mullet, J. (2016). Dynamics of biomass partitioning, stem gene expression, cell wall biosynthesis, and sucrose accumulation during development of *Sorghum bicolor*. *Plant J.* doi: 10.1111/tpj.13269. [Epub ahead of print].
- Mendu, V., Griffiths, J. S., Persson, S., Stork, J., Downie, A. B., Voiniciuc, C., et al. (2011a). Subfunctionalization of cellulose synthases in seed coat epidermal cells mediates secondary radial wall synthesis and mucilage attachment. *Plant Physiol.* 157, 441–453. doi: 10.1104/pp.111.179069
- Mendu, V., Harman-Ware, A. E., Crocker, M., Jae, J., Stork, J., Morton, S., et al. (2011b). Identification and thermochemical analysis of high-lignin feedstocks for biofuel and biochemical production. *Biotechnol. Biofuels* 4, 1–14. doi: 10.1186/1754-6834-4-43
- Moura, J. C., Bonine, C. A., de Oliveira Fernandes Viana, J., Dornelas, M. C., and Mazzafera, P. (2010). Abiotic and biotic stresses and changes in the lignin content and composition in plants. *J. Integr. Plant Biol.* 52, 360–376. doi: 10.1111/j.1744-7909.2010.00892.x
- Muthamilarasan, M., Khan, Y., Jaishankar, J., Shweta, S., Lata, C., and Prasad, M. (2015). Integrative analysis and expression profiling of secondary cell wall genes in C4 biofuel model *Setaria italica* reveals targets for lignocellulose bioengineering. *Front. Plant Sci.* 6:965. doi: 10.3389/fpls.2015.00965
- Nedukha, O. M. (2015). Callose: localization, functions, and synthesis in plant cells. *Cytol. Genet.* 49, 49–57. doi: 10.3103/S0095452715010090
- Nishitani, K., and Vissenberg, K. (2007). “Roles of the XTH protein family in the expanding cell,” in *The Expanding Cell*, eds J.-P. Verbelen and K. Vissenberg (Berlin, Heidelberg: Springer), 89–116.
- Ochoa-Villarreal, M., Aispuro-Hernández, E., Vargas-Arispuro, I., and Martínez-Téllez, M. A. (2012). *Plant Cell Wall Polymers: Function, Structure and Biological Activity of Their Derivatives, Polymerization*. Rijeka: InTech.
- Okamoto-Nakazato, A., Takahashi, K., Katoh-Semba, R., and Katou, K. (2001). Distribution of yieldin, a regulatory protein of the cell wall yield threshold, in etiolated cowpea seedlings. *Plant Cell Physiol.* 42, 952–958. doi: 10.1093/pcp/pce121
- Olson, S. N., Ritter, K., Rooney, W., Kemanian, A., McCarl, B. A., Zhang, Y., et al. (2012). High biomass yield energy sorghum: developing a genetic model for C4 grass bioenergy crops. *Biofuels Bioprod. Biorefining* 6, 640–655. doi: 10.1002/bbb.1357
- Palmer, N. A., Sattler, S. E., Saathoff, A. J., Funnell, D., Pedersen, J. F., and Sarath, G. (2008). Genetic background impacts soluble and cell wall-bound aromatics in brown midrib mutants of sorghum. *Planta* 229, 115–127. doi: 10.1007/s00425-008-0814-1
- Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., et al. (2009). The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457, 551–556. doi: 10.1038/nature07723
- Pear, J. R., Kawagoe, Y., Schreckengost, W. E., Delmer, D. P., and Stalker, D. M. (1996). Higher plants contain homologs of the bacterial celA genes encoding the catalytic subunit of cellulose synthase. *Proc. Natl. Acad. Sci. U.S.A.* 93, 12637–12642. doi: 10.1073/pnas.93.22.12637
- Persson, S., Paredes, A., Carroll, A., Palsdottir, H., Doblin, M., Poindexter, P., et al. (2007). Genetic evidence for three unique components in primary cell-wall cellulose synthase complexes in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15566–15571. doi: 10.1073/pnas.0706592104
- Petti, C., Harman-Ware, A. E., Tateno, M., Kushwaha, R., Shearer, A., Downie, A. B., et al. (2013). Sorghum mutant RG displays antithetic leaf shoot lignin accumulation resulting in improved stem saccharification properties. *Biotechnol. Biofuels* 6:146. doi: 10.1186/1754-6834-6-146
- Rai, K. M., Balasubramanian, V. K., Welker, C. M., Pang, M., Hii, M. M., and Mendu, V. (2015). Genome wide comprehensive analysis and web resource development on cell wall degrading enzymes from phyto-parasitic nematodes. *BMC Plant Biol.* 15:187. doi: 10.1186/s12870-015-0576-4
- Richmond, T. A., and Somerville, C. R. (2000). The cellulose synthase superfamily. *Plant Physiol.* 124, 495–498. doi: 10.1104/pp.124.2.495
- Rooney, W. L., Blumenthal, J., Bean, B., and Mullet, J. E. (2007). Designing sorghum as a dedicated bioenergy feedstock. *Biofuels Bioprod. Biorefining* 1, 147–157. doi: 10.1002/bbb.15
- Rose, J. K. C., Braam, J., Fry, S. C., and Nishitani, K. (2002). The XTH family of enzymes involved in xyloglucan endotransglucosylation and Endohydrolysis: current perspectives and a new unifying nomenclature. *Plant Cell Physiol.* 43, 1421–1435. doi: 10.1093/pcp/pcf171
- Saballos, A., Ejeta, G., Sanchez, E., Kang, C., and Vermerris, W. (2009). A genomewide analysis of the cinnamyl alcohol dehydrogenase family in sorghum [*Sorghum bicolor* (L.) Moench] identifies SbCAD2 as the brown midrib6 gene. *Genetics* 181, 783–795. doi: 10.1534/genetics.108.098996
- Sattler, S. E., Saathoff, A. J., Haas, E. J., Palmer, N. A., Funnell-Harris, D. L., Sarath, G., et al. (2009). A nonsense mutation in a cinnamyl alcohol dehydrogenase gene is responsible for the Sorghum brown midrib6 phenotype. *Plant Physiol.* 150, 584–595. doi: 10.1104/pp.109.136408
- Sattler, S. E., Saballos, A., Xin, Z., Funnell-Harris, D. L., Vermerris, W., and Pedersen, J. F. (2014). Characterization of novel Sorghum brown midrib mutants from an EMS-mutagenized population. *G3 (Bethesda)* 4, 2115–2124. doi: 10.1534/g3.114.014001
- Saxena, I. M., Lin, F. C., and Brown, R. M. (1990). Cloning and sequencing of the cellulose synthase catalytic subunit gene of *Acetobacter xylinum*. *Plant Mol. Biol.* 15, 673–683. doi: 10.1007/BF00016118
- Scheible, W. R., and Pauly, M. (2004). Glycosyltransferases and cell wall biosynthesis: novel players and insights. *Curr. Opin. Plant Biol.* 7, 285–295. doi: 10.1016/j.pbi.2004.03.006
- Scully, E. D., Gries, T., Funnell-Harris, D. L., Xin, Z., Kovacs, F. A., Vermerris, W., et al. (2016). Characterization of novel Brown midrib 6 mutations affecting lignin biosynthesis in sorghum. *J. Integr. Plant Biol.* 58, 136–149. doi: 10.1111/jipb.12375
- Somerville, C. (2006). Cellulose synthesis in higher plants. *Annu. Rev. Cell Dev. Biol.* 22, 53–78. doi: 10.1146/annurev.cellbio.22.022206.160206
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Taylor, J. R. N., Schober, T. J., and Bean, S. R. (2006). Novel food and non-food uses for sorghum and millets. *J. Cereal Sci.* 44, 252–271. doi: 10.1016/j.jcs.2006.06.009
- Taylor, N. G., Laurie, S., and Turner, S. R. (2000). Multiple cellulose synthase catalytic subunits are required for cellulose synthesis in Arabidopsis. *Plant Cell* 12, 2529–2539. doi: 10.1105/tpc.12.12.2529
- Taylor, S. H., Hulme, S. P., Rees, M., Ripley, B. S., Ian Woodward, F., and Osborne, C. P. (2010). Ecophysiological traits in C3 and C4 grasses: a phylogenetically controlled screening experiment. *New Phytol.* 185, 780–791. doi: 10.1111/j.1469-8137.2009.03102.x
- Tenhaken, R. (2014). Cell wall remodeling under abiotic stress. *Front. Plant Sci.* 5:771. doi: 10.3389/fpls.2014.00771
- Unrean, P., and Srienc, F. (2010). Continuous production of ethanol from hexoses and pentoses using immobilized mixed cultures of *Escherichia coli* strains. *J. Biotechnol.* 150, 215–223. doi: 10.1016/j.jbiotec.2010.08.002
- Vanholme, R., Demedts, B., Morreel, K., Ralph, J., and Boerjan, W. (2010). Lignin biosynthesis and structure. *Plant Physiol.* 153, 895–905. doi: 10.1104/pp.110.155119
- Voiniciuc, C., Schmidt, M. H., Berger, A., Yang, B., Ebert, B., Scheller, H. V., et al. (2015). MUCILAGE-RELATED10 produces galactoglucomannan that maintains pectin and cellulose architecture in arabidopsis seed mucilage. *Plant Physiol.* 169, 403–420. doi: 10.1104/pp.15.00851
- Voorrips, R. E. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Hered.* 93, 77–78. doi: 10.1093/jhered/93.1.77



- Vuttipongchaikij, S., Brocklehurst, D., Steele-King, C., Ashford, D. A., Gomez, L. D., and McQueen-Mason, S. J. (2012). Arabidopsis GT34 family contains five xyloglucan  $\alpha$ -1,6-xylosyltransferases. *New Phytol.* 195, 585–595. doi: 10.1111/j.1469-8137.2012.04196.x
- Wang, T., McFarlane, H. E., and Persson, S. (2016). The impact of abiotic factors on cellulose synthesis. *J. Exp. Bot.* 67, 543–552. doi: 10.1093/jxb/erv488
- Wei, H., Brunecky, R., Donohoe, B. S., Ding, S.-Y., Ciesielski, P. N., Yang, S., et al. (2015). Identifying the ionically bound cell wall and intracellular glycoside hydrolases in late growth stage Arabidopsis stems: implications for the genetic engineering of bioenergy crops. *Front. Plant Sci.* 6:315. doi: 10.3389/fpls.2015.00315
- Welker, C., Balasubramanian, V., Petti, C., Rai, K., DeBolt, S., and Mendu, V. (2015). Engineering plant biomass lignin content and composition for biofuels and bioproducts. *Energies* 8:7654. doi: 10.3390/en8087654
- Xin, Z., Wang, M. L., Barkley, N. A., Burow, G., Franks, C., Pederson, G., et al. (2008). Applying genotyping (TILLING) and phenotyping analyses to elucidate gene function in a chemically induced sorghum mutant population. *BMC Plant Biol.* 8:103. doi: 10.1186/1471-2229-8-103
- Xu, Z., Zhang, D., Hu, J., Zhou, X., Ye, X., Reichel, K. L., et al. (2009). Comparative genome analysis of lignin biosynthesis gene families across the plant kingdom. *BMC Bioinformatics* 10(Suppl. 11):S3. doi: 10.1186/1471-2105-10-S11-S3
- Yan, L., Liu, S., Zhao, S., Kang, Y., Wang, D., Gu, T., et al. (2012). Identification of differentially expressed genes in sorghum (*Sorghum bicolor*) brown midrib mutants. *Physiol. Plant.* 146, 375–387. doi: 10.1111/j.1399-3054.2012.01646.x
- Yin, Y., Chen, H., Hahn, M. G., Mohnen, D., and Xu, Y. (2010). Evolution and function of the plant cell wall synthesis-related glycosyltransferase family 8. *Plant Physiol.* 153, 1729–1746. doi: 10.1104/pp.110.154229
- Yin, Y., Huang, J., and Xu, Y. (2009). The cellulose synthase superfamily in fully sequenced plants and algae. *BMC Plant Biol.* 9:99. doi: 10.1186/1471-2229-9-99
- Yu, H., Cong, L., Zhu, Z., Wang, C., Zou, J., Tao, C., et al. (2015). Identification of differentially expressed microRNA in the stems and leaves during sugar accumulation in sweet sorghum. *Gene* 571, 221–230. doi: 10.1016/j.gene.2015.06.056
- Zabotina, O. A., Avci, U., Cavalier, D., Pattathil, S., Chou, Y. H., Eberhard, S., et al. (2012). Mutations in multiple XXT genes of Arabidopsis reveal the complexity of xyloglucan biosynthesis. *Plant Physiol.* 159, 1367–1384. doi: 10.1104/pp.112.198119
- Zhong, R., and Ye, Z. H. (2003). Unraveling the functions of glycosyltransferase family 47 in plants. *Trends Plant Sci.* 8, 565–568. doi: 10.1016/j.tplants.2003.10.003

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Rai, Thu, Balasubramanian, Cobos, Disasa and Mendu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Integrative analysis and expression profiling of secondary cell wall genes in C<sub>4</sub> biofuel model *Setaria italica* reveals targets for lignocellulose bioengineering

Mehanathan Muthamilarasan<sup>1†</sup>, Yusuf Khan<sup>1†</sup>, Jananee Jaishankar<sup>1</sup>, Shweta Shweta<sup>1</sup>, Charu Lata<sup>2</sup> and Manoj Prasad<sup>1\*</sup>

<sup>1</sup> National Institute of Plant Genome Research, New Delhi, India, <sup>2</sup> Division of Plant-Microbe Interactions, CSIR-National Botanical Research Institute, Lucknow, India

## OPEN ACCESS

### Edited by:

Gautam Sarath,  
United States Department of  
Agriculture - Agricultural Research  
Service, USA

### Reviewed by:

Lam-Son Tran,  
RIKEN Center for Sustainable  
Resource Science, Japan  
Erin D. Scully,  
United States Department of  
Agriculture - Agricultural Research  
Service, USA

### \*Correspondence:

Manoj Prasad  
manoj\_prasad@nipgr.ac.in

<sup>†</sup>These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Crop Science and Horticulture,  
a section of the journal  
Frontiers in Plant Science

**Received:** 26 June 2015

**Accepted:** 22 October 2015

**Published:** 04 November 2015

### Citation:

Muthamilarasan M, Khan Y,  
Jaishankar J, Shweta S, Lata C and  
Prasad M (2015) Integrative analysis  
and expression profiling of secondary  
cell wall genes in C<sub>4</sub> biofuel model  
*Setaria italica* reveals targets for  
lignocellulose bioengineering.  
Front. Plant Sci. 6:965.  
doi: 10.3389/fpls.2015.00965

Several underutilized grasses have excellent potential for use as bioenergy feedstock due to their lignocellulosic biomass. Genomic tools have enabled identification of lignocellulose biosynthesis genes in several sequenced plants. However, the non-availability of whole genome sequence of bioenergy grasses hinders the study on bioenergy genomics and their genomics-assisted crop improvement. Foxtail millet (*Setaria italica* L.; Si) is a model crop for studying systems biology of bioenergy grasses. In the present study, a systematic approach has been used for identification of gene families involved in cellulose (*CesA/CsI*), callose (*GsI*) and monolignol biosynthesis (*PAL*, *C4H*, *4CL*, *HCT*, *C3H*, *CCoAOMT*, *F5H*, *COMT*, *CCR*, *CAD*) and construction of physical map of foxtail millet. Sequence alignment and phylogenetic analysis of identified proteins showed that monolignol biosynthesis proteins were highly diverse, whereas *CesA/CsI* and *GsI* proteins were homologous to rice and *Arabidopsis*. Comparative mapping of foxtail millet lignocellulose biosynthesis genes with other C<sub>4</sub> panicoid genomes revealed maximum homology with switchgrass, followed by sorghum and maize. Expression profiling of candidate lignocellulose genes in response to different abiotic stresses and hormone treatments showed their differential expression pattern, with significant higher expression of *SiGsI12*, *SiPAL2*, *SiHCT1*, *SiF5H2*, and *SiCAD6* genes. Further, due to the evolutionary conservation of grass genomes, the insights gained from the present study could be extrapolated for identifying genes involved in lignocellulose biosynthesis in other biofuel species for further characterization.

**Keywords:** foxtail millet (*Setaria italica* L.), secondary cell wall biosynthesis, lignocellulose, bioenergy grasses, genomics, comparative mapping

## INTRODUCTION

Cell wall polymers of living plants constitute a predominant proportion of their biomass, which is formed by fermentable linked sugars. These polymers form a major structural component of plant cell wall and particularly, secondary cell walls provide mechanical strength and rigidity to vascular plants (Wang et al., 2013; Zhong and Ye, 2015). Secondary cell walls are present

in tracheary elements, xylem, phloem, extraxylary and interfascicular fibers, sclereids and seed coats, and are made of cellulose, hemicelluloses and lignin. Cellulose, the primary unit, cross-links with hemicelluloses including xylan and glucomannan, and impregnated with phenolic polymer lignin, and altogether, this complex polymeric network forms secondary cell wall. The proportion of cellulose, hemicelluloses, and lignin varies among different plant species and of note, the composition may also vary in response to diverse developmental and environmental conditions (Zhong and Ye, 2015). Being the prime constituents of wood and fiber, secondary cell walls have been extensively studied to understand and exploit their biofuel prospects. Biochemical and genomic methods have identified the genes encoding for enzymes which participate in the biosynthesis of secondary cell wall components.

Pear et al. (1996) was the first to identify *cellulose synthase* (*CesA*) genes in cotton and following this, *CesA* genes in other plants have been identified and their numbers were shown to vary between plant species. In *Arabidopsis*, 10 *CesA* genes have been identified (Richmond and Somerville, 2000), whereas 12 in maize (Appenzeller et al., 2004), 16 in barley (Burton et al., 2004), 18 in poplar (Djerbi et al., 2005) have been reported. The *CesA* enzymes belong to glycosyltransferase-2 (GT-2) superfamily, which is defined by an eight-transmembrane topology and conserved cytosolic substrate binding and catalytic residues (McFarlane et al., 2014). In addition to *CesA*, plants also have *cellulose synthase-like* (*Csl*) genes, which can be involved in biosynthesis of hemicellulose and other glucans (Lerouxel et al., 2006). *Csl* genes can synthesize other polysaccharides that are not components of the hemicellulose matrix (Lerouxel et al., 2006). So far, several types of *Csl* genes have been identified, denoted as *CslA* to *CslK*. *CslA* encodes for (1,4)- $\beta$ -D-mannan synthases (Dhugga et al., 2004; Liepman et al., 2005), *CslF* and *CslH* encode the mixed linkage glucan synthases for (1,3;1,4)- $\beta$ -glucan biosynthesis (Burton et al., 2006; Doblin et al., 2009), *CslC* genes are involved in xyloglucan biosynthesis (Cocuron et al., 2007), and *CslD* in xylan and homogalacturonan synthesis (Hamann et al., 2004; Bernal et al., 2008a,b; Li et al., 2009), whereas the functional roles of other *Csl* genes remain elusive (Yin et al., 2009). Noteworthy, *CslB* and *CslG* are specific to dicots whereas *CslF* and *CslH* are found only in monocots (Fincher, 2009; Doblin et al., 2010), but recently two *CslG* genes were identified in *Panicum virgatum* (*Pavirv00027268m* and *Pavirv00027269m*; Yin et al., 2014).

Callose is a (1,3)- $\beta$ -D-glucan, which is not present in cell walls but deposited in the walls of specialized tissues such as pollen mother cell walls, plasmodesmatal canals, and sieve plates in dormant phloem during normal growth and development (Stone and Clarke, 1992). In addition, callose is also deposited in response to environmental stimuli including abiotic stress, wounding, and pathogen challenge (Stone and Clarke, 1992; Muthamilarasan and Prasad, 2013). Callose is synthesized by callose synthases, which are encoded by *glucan synthase-like* (*Gsl*) genes (Saxena and Brown, 2000; Cui et al., 2001). To date, 12 *Gsl* genes have been identified in *Arabidopsis*, 13 in rice, 9 in poplar, and 8 in barley (Farrokhi et al., 2006).

In the case of lignin biosynthesis, phenylalanine is metabolized through the phenylpropanoid pathway to produce hydroxycinnamoyl-CoA esters, which enter the lignin branch of this pathway and are converted to monolignols. The process requires the involvement of phenylalanine ammonia lyase (PAL), trans-cinnamate 4-hydroxylase (C4H), 4-coumarate CoA ligase (4CL), hydroxycinnamoyl CoA:shikimate/quinate hydroxycinnamoyl transferase (HCT), *p*-coumaroyl shikimate 3'-hydroxylase (C3H), caffeoyl CoA 3-O-methyltransferase (CCoAOMT), ferulate 5-hydroxylase (F5H), caffeic acid O-methyltransferase (COMT), cinnamoyl CoA reductase (CCR), and cinnamyl alcohol dehydrogenase (CAD) (Bonawitz and Chapple, 2010; Zhong and Ye, 2015). Of these enzymes, PAL is the first enzyme of phenylpropanoid pathway which catalyzes the deamination of phenylalanine to generate cinnamic acid and C4H hydroxylates cinnamic acid to generate *p*-coumaric acid (Harakava, 2005). 4CL performs CoA esterification of *p*-coumaric acid and caffeic acid, whereas HCT catalyzes the conversion of *p*-coumaroyl-CoA and caffeoyl-CoA into corresponding shikimate or quinate esters and C3H converts these esters to corresponding caffeoyl esters. Following this, CCoAOMT catalyzes methylation of caffeoyl CoA to produce feruloyl CoA, whereas CCR converts hydroxycinnamoyl CoA esters to their corresponding aldehydes (Harakava, 2005). F5H has been assumed to catalyze the conversion of ferulic acid to 5-hydroxyferulic acid but recombinant DNA studies in *Arabidopsis* and *Liquidambar styraciflua* revealed that F5H converts coniferaldehyde and coniferyl alcohol to synapaldehyde and sinapyl alcohol, respectively (Humphreys et al., 1999; Osakabe et al., 1999). COMT is involved in the conversion of 5-hydroxyconiferaldehyde and/or 5-hydroxyconiferyl alcohol to sinapaldehyde and/or sinapyl alcohol, respectively (Osakabe et al., 1999; Parvathi et al., 2001), while CAD catalyzes the conversion of cinnamyl aldehydes into their corresponding alcohols (Harakava, 2005). The genes encoding these enzymes have recently been identified and characterized in several plant species (Raes et al., 2003; Vanholme et al., 2012; Shen et al., 2013; Carocha et al., 2015; van Parijs et al., 2015).

With the raise in the impacts of global climate change, reduction of greenhouse gases is essential, which could be facilitated through generating biorenewables. Importantly, production of lignocellulosic biofuels from secondary cell wall biomass has become a strategic research area, as it holds the potential to enhance energy security.  $C_4$  grasses, namely switchgrass (*P. virgatum*), napier grass (*Pennisetum purpureum*), pearl millet (*P. glaucum*), and foxtail millet (*Setaria italica*) have recently gained momentum in lignocellulosic biofuel research due to their high-efficiency  $CO_2$  fixation and efficient conversion of solar energy to biomass through  $C_4$  photosynthesis and photorespiration-suppressing modifications, respectively (Schmer et al., 2008; Byrt et al., 2011; van der Weijde et al., 2013). In addition, these grasses also possess better water use efficiency (WUE), higher nitrogen use efficiency (NUE), capacity to grow in arid and semi-arid regions and relatively high tolerance to environmental constraints including heat, drought, salinity and water-logging. For these reasons,  $C_4$  photosynthesis is an

important trait for lignocellulosic biofuel crops (Byrt et al., 2011; van der Weijde et al., 2013).

Recently, foxtail millet (*S. italica*) and its wild progenitor, green foxtail (*S. viridis*) have been recognized as the suitable experimental models for biofuel research owing to their genetic relatedness to several biofuel grasses (Li and Brutnell, 2011; Zhang et al., 2012; Lata et al., 2013; Petti et al., 2013; Diao et al., 2014; Warnasooriya and Brutnell, 2014; Muthamilarasan and Prasad, 2015). The genomes of both foxtail millet and green foxtail have been sequenced (Bennetzen et al., 2012; Zhang et al., 2012), and the availability of foxtail millet draft genome sequence in public domains has facilitated various genetic and genomic studies in this model crop pertaining to stress response and crop improvement (Diao et al., 2014; Muthamilarasan and Prasad, 2015; Muthamilarasan et al., 2015) though no comprehensive genome-wide study on biofuel traits has been performed. Recently, Petti et al. (2013) has compared the lignocellulosic feedstock composition, cellulose biosynthesis inhibitor response, saccharification dynamics and *CesA* gene family of green foxtail with sorghum, maize and switchgrass. The study identified eight potential *CesA* gene family members for functional genomic characterization (Petti et al., 2013).

The present study has been performed to identify the gene families participating in lignocellulose biosynthesis using computational approaches. Further, qRT-PCR analysis of few genes has been performed to understand their expression patterns in response to different abiotic stress treatments.

## MATERIALS AND METHODS

### Identification of Lignocellulose Biosynthesis Gene Families

Protein sequences of enzymes involved in cellulose biosynthesis, namely *CesA*, *Csl*, and *Gsl* of rice and *Arabidopsis* were retrieved from cell wall genomics webserver (<https://cellwall.genomics.purdue.edu/intro/index.html>). The sequences for PAL, C4H, 4CL, HCT, C3H, CCoAMT, F5H, COMT, CCR, and CAD reported in other crops (Appenzeller et al., 2004; Burton et al., 2004; Carocha et al., 2015; Zhong and Ye, 2015) were retrieved from respective literatures and HMM profile has been generated for individual families. Precisely, the sequences of respective families were aligned using Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) and HMM profiles were built using hmmbuilt command (<http://hmmer.janelia.org/>). HMMER tool was used to identify respective homologous proteins in foxtail millet protein dataset retrieved from Phytozome v10.2 (<http://phytozome.jgi.doe.gov/>) under default parameters (Muthamilarasan et al., 2014a). The protein sequences were confirmed using HMMSCAN (<http://www.ebi.ac.uk/Tools/hmmer/search/hmmscan>) analysis, and respective genomic, transcript, and CDS sequences were downloaded from Phytozome by BLAST searching the retrieved protein sequences against *S. italica* database under default parameters ([http://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org\\_Sitalica](http://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Sitalica)).

### Protein Properties and Phylogenetic Analysis

The properties of identified cell wall-related proteins including molecular weight, pI, and instability index were identified using ExPASy ProtParam tool (<http://web.expasy.org/protparam/>). The amino acid sequences of respective families were imported into MEGA v6 (Tamura et al., 2013) for multiple sequence alignment and phylogenetic tree construction using neighbor-joining method after bootstrap analysis for 1000 replicates (Muthamilarasan et al., 2014b). Sequence alignment and analysis was performed using BioEdit v7.2.5 (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>).

### Physical Mapping and Gene Structure Analysis

The chromosomal location of cell wall biosynthesis genes including chromosome number, position of gene start and end, gene length and orientation were obtained from Phytozome and a physical map was constructed using MapChart (Voorrips, 2002). Gene duplications, namely tandem and segmental were identified by performing MCScanx (Wang et al., 2012) according to the protocol of Plant Genome Duplication Database (Lee et al., 2012). Gene structure was predicted using Gene Structure Display Server v2.0 (<http://gsds.cbi.pku.edu.cn/>).

### Promoter Analysis, Targeting miRNA, and Marker Prediction

The upstream genomic sequence (~2 kb) of lignocellulose pathway genes of foxtail millet were retrieved from Phytozome and the presence of *cis*-regulatory elements were identified by Signal Scan Search using New PLACE web server (<https://sogo.dna.affrc.go.jp/cgi-bin/sogo.cgi?page=analysis&lang=en>). Mature miRNA sequences of foxtail millet were downloaded from miRBase v21 (Kozomara and Griffiths-Jones, 2014) and FmMiRNaDb (Khan et al., 2014). This information along with the miRNA data of a dehydration stress library (Yadav et al., unpublished data) were used to identify the miRNAs targeting the transcripts of lignocellulose pathway genes using psRNAtarget server (Dai and Zhao, 2011) under default parameters. The large-scale genome-wide molecular markers namely simple sequence repeats (SSR; Pandey et al., 2013), expressed sequence tag (EST)-SSR (eSSR; Kumari et al., 2013), and intron-length polymorphic markers (Muthamilarasan et al., 2014c) were retrieved from the Foxtail millet Marker Database (<http://www.nipgr.res.in/foxtail.html>; Suresh et al., 2013) and searched for their presence in the genic and promoter regions of lignocellulose biosynthesis genes using in-house perl script.

### Comparative Genome Mapping and Evolutionary Analysis

Protein sequences of lignocellulose pathway genes of foxtail millet were BLASTP searched against the protein sequences of switchgrass (*Panicum virgatum*), rice (*Oryza sativa*), and poplar (*Populus trichocarpa*), and hits with more than 80% identity were selected. The genomic and CDS sequences along with chromosomal locations for these proteins were retrieved



by performing BLAST searches against the corresponding genomes retrieved from Gramene (<http://www.gramene.org/>) under default parameters and comparative maps were visualized using Circos (Krzywinski et al., 2009). Reciprocal BLAST was also performed to ensure the unique relationship between the homologous genes (Mishra et al., 2013). Estimation of nonsynonymous substitutions per non-synonymous site (Ka) and synonymous substitutions per synonymous site (Ks) for paralogous (tandem and segmentally duplicated genes) as well as homologous (comparative mapping data) gene pairs were calculated by codeml program in PAML using PAL2NAL (Suyama et al., 2006). The Ka/Ks ratios along with estimation of duplication and divergence (as  $T = Ks/2\lambda$ , where,  $\lambda = 6.5 \times 10^{-9}$ ) were performed according to Puranik et al. (2013).

### ***In silico* Expression Profiling in Tissues and Drought Stress**

The transcriptome data of different tissues, namely root (SRX128223), stem (SRX128225), leaf (SRX128224), spica (SRX128226), and a drought stress library (SRR629694) as well as its control (SRR629695) were retrieved from European Nucleotide Archive (<http://www.ebi.ac.uk/ena>) (Zhang et al., 2012; Qi et al., 2013). The reads were filtered using NGS Toolkit (<http://www.nipgr.res.in/ngsqctoolkit.html>), mapped on foxtail millet genome using CLC Genomics Workbench v4.7.1, normalized by RPKM method and a heat map was generated using MultiExperiment Viewer (MeV) v4.9 (Saeed et al., 2003).

### **Plant Materials, Stress and Hormone Treatments and Quantitative Real-time PCR Analysis**

Seeds of foxtail millet cv. “IC-403579” (dehydration and salinity tolerant) were grown under optimum conditions following Lata et al. (2014). Twenty one day-old seedlings were exposed to 250 mM NaCl (salinity), 20% PEG6000 (dehydration), 4°C (cold), 100 mM abscisic acid (ABA), 100 mM methyl jasmonate (MeJA), and 100 mM salicylic acid (SA) treatments (Mishra et al., 2013; Puranik et al., 2013; Kumar et al., 2015) and whole seedlings were collected at 0 h (h) (control), 1 h (early), and 24 h (late) (Yadav et al., 2015). The samples were frozen immediately in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ . RNA isolation, cDNA synthesis and RT-PCR analysis were performed according to Puranik et al. (2013) in three technical replicates for each biological triplicate using the primers mentioned in **Supplementary Table S1**. All qRT-PCR data were the means of at least three independent experiments and the results were presented as the mean values  $\pm$  SE. The significance of differences between mean values of control and each stressed samples were statistically performed using One-Way analysis of variance (ANOVA) and comparison among means was carried out through Tukey-Kramer multiple comparisons test using GRAPHPAD INSTAT software v3.10 (<http://www.graphpad.com>). The differences in the effects of stress treatments on various parameters in 16 foxtail millet genes under study were considered statistically significant at  $*P < 0.05$ ,  $**P < 0.01$ ,  $***P < 0.001$ .

## **RESULTS**

### **CesA/Csl and Gsl Superfamily of Foxtail Millet**

HMM searches identified the presence of 14 CesA (SiCesA) and 39 Csl (SiCsl) proteins in foxtail millet (**Supplementary Table S2**). Among the 14 SiCesA proteins, one was found to be an alternate transcript (Si028766m), whereas in SiCsl, three alternate transcripts (Si029554m, Si035399m, and Si035101m) were identified. Domain analysis of SiCesA proteins revealed the presence of both the cellulose synthase domain (CS; PF03552) and the zinc finger structure (ZF; PF14569) in all the proteins except SiCesA8 and SiCesA10, which have only the CS domain (**Supplementary Table S3**). In addition, all the SiCesA proteins except SiCesA8 had Glycosyl transferase 2 (GT2; PF13632) domain. In the case of SiCsl proteins, 36 proteins (primary transcripts) were identified, of which 10 belonged to SiCslA, 6 to SiCslC, 5 to SiCslD, 4 to SiCslE, 7 to SiCslF, 2 each to SiCslH and SiCslJ families (**Supplementary Table S2**). Interestingly, two members of CslJ have been identified in foxtail millet, which was previously considered to be a cereal-specific gene family (Doblin et al., 2010). Domain analysis showed that all the SiCslA and SiCslC proteins possess GT2 domain (PF13641, PF13632, PF00535, and PF13506) (**Supplementary Table S3**).

All 5 SiCslD proteins possess CS (PF03552) and GT2 (PF13632) domain, and interestingly, SiCslD2, SiCslD4, and SiCslD5 were evidenced to have an additional RING/Ubox like zinc-binding domain (PF14570), whereas SiCslD3 has two CS domains (**Supplementary Table S4**). All the SiCslE proteins except SiCslE2 have more than one CS domain and SiCslE3 has an additional GT2 domain (PF13641). In the case of SiCslF proteins, all of the members except SiCslF6 have two CS domains and in addition, SiCslF1, SiCslF3, and SiCslF7 possess GT2 domain (PF13632). Two members each belonging to CslH and CslJ family proteins were identified and both the group members have two CS domains (**Supplementary Table S4**).

A total of 12 Gsl (SiGsl) proteins were identified in foxtail millet and all possessed glucan synthesis (GS) domain (1,3-beta-glucan synthase component; PF02364) (**Supplementary Table S5**). The number of GS domain within these proteins also varied as SiGsl1, SiGsl6, and SiGsl12 have two GS domains, whereas SiGsl11 had three domains. In addition, SiGsl2, SiGsl3, SiGsl5, SiGsl7, SiGsl8, SiGsl10, and SiGsl11 have a 1,3-beta-glucan synthase subunit FKS1, domain-1 (PF14288). Furthermore, SiGsl08, SiGsl10, and SiGsl11 have an additional Vta1 (VPS20-associated protein 1) like domain (PF04652) (**Supplementary Table S5**).

### **Monolignol Pathway Proteins of Foxtail Millet**

HMM profiling of PAL (SiPAL), C4H (SiC4H), 4CL (Si4CL), HCT (SiHCT), C3H (SiC3H), CCoAOMT (SiCCoAOMT), F5H (SiF5H), COMT (SiCOMT), CCR (SiCCR), and CAD (SiCAD) proteins in foxtail millet identified 10, 3, 20, 2, 2, 6, 2, 4, 33, and 13 members, respectively (**Supplementary Table S6**). Splice variants were evidenced among these members, including

three each in SiCL16 and SiCCR14, two in SiCCR11 and one each in Si4CL5, SiCCoAOMT1, SiCOMT, SiCCR11, and SiCCR17. HMMSCAN revealed a diverse domain organization of these proteins (**Supplementary Table S7**). All of the SiPAL proteins possess aromatic amino acid lyase (PF00221) domain, whereas Cytochrome P450 (PF00067) was present in all SiC4H, SiC3H, and SiF5H proteins. AMP-binding enzyme (PF00501) and AMP-binding enzyme C-terminal (PF13193) domains were present in all the Si4CL proteins except Si4CL13, which has only an AMP-binding enzyme domain. Both SiHCT1 and SiHCT2 have transferase family (PF02458) domains, and SiCCoAOMT proteins were evidenced to possess O-methyltransferase (PF01596) and methyltransferase (PF13578) domains with an exception of SiCCoAOMT, which has two O-methyltransferase domains (**Supplementary Table S7**). O-methyltransferase domain was also found to be present in SiCOMT proteins, whereas SiCOMT2 has an additional dimerisation domain (PF13578). A diverse domain composition was observed among SiCCR proteins in addition to the presence of signature NAD-dependent epimerase/dehydratase family (PF01370) and 3-beta hydroxysteroid dehydrogenase/isomerase family (PF01073) domains. Almost all the SiCCR proteins possess additional domains including GDP-mannose-4,6-dehydratase (PF16363), Male sterility protein (PF07993), NmrA-like family (PF05368), NAD(P)H-binding (PF13460), Polysaccharide biosynthesis protein (PF02719), and KR domains (PF08659). Of note, SiCCR7 was devoid of any of these domains except the NAD-dependent epimerase/dehydratase family domain, and SiCCR3 has an additional Alcohol dehydrogenase GroES-like domain (PF08240) (**Supplementary Table S7**). The presence of Alcohol dehydrogenase GroES-like and Zinc-binding dehydrogenase (PF00107) domains is the characteristic feature of SiCAD proteins and in addition to these, D-isomer specific 2-hydroxyacid dehydrogenase, NAD-binding domain (PF02826) was present in SiCAD4, SiCAD9, and SiCAD12. Moreover, an alanine dehydrogenase/PNT, C-terminal domain (PF01262) was found to be present in SiCAD12 and SiCAD13 (**Supplementary Table S7**).

## Properties of Lignocellulose Pathway Proteins

Among the SiCesA proteins, SiCesA4 was the largest protein with 1095 amino acids (aa), followed by SiCesA2 (1092 aa), SiCesA11 (1090 aa) and SiCesA3 (1088 aa), and the smallest was SiCesA8 (884 aa) (**Supplementary Table S2**). The molecular weight of these proteins also varied accordingly, ranging from SiCesA8 (95.5 kDa) to SiCesA11 (123.2 kDa), with an isoelectric pH (pI) of 6.03 (SiCesA10) to 8.15 (SiCesA1). The protein instability index was between 36.07 (SiCesA11) to 50.62 (SiCesA8), which signified that all the SiCesA proteins except SiCesA2, SiCesA8, and SiCesA10 were stable. In the case of SiCsl proteins, the smallest protein was SiCslE2 with 144 aa and the largest was SiCslD1 (1217 aa), and their respective molecular weights ranged from 16.4 kDa (SiCslE2) to 132.2 kDa (SiCslD1). The pI of SiCsl proteins ranged from 4.61 (SiCslE2) to 9.32 (SiCslF7), and their instability index range (31.44–67.71) revealed that a maximum of SiCsl proteins (~33%) were stable. The size and molecular

weights of SiGsl proteins ranged from 418 aa (47.8 kDa in SiGsl9) to 1956 aa (225.2 kDa in SiGsl8). Similarly, pI range of these proteins was between 8.61 (SiGsl12) and 9.69 (SiGsl9). The instability index range between 28.89 and 52.08 indicated that ~46% of SiGsl proteins were stable and the rest are unstable (**Supplementary Table S2**).

The SiPAL class of monolignol pathway proteins showed a narrow range of protein properties, as their sizes varied from 699 (SiPAL1 and SiPAL2) to 851 aa (SiPAL10), with molecular weights from 74.9 kDa (SiPAL2) to 91.1 kDa (SiPAL10) (**Supplementary Table S6**). The pI range of SiPAL was between 5.82 and 6.52, and their instability index range (28.82–39.84) showed that all the proteins except SiPAL5 were stable. The three members of SiC4H, namely SiC4H1, SiC4H2, and SiC4H3 had molecular sizes of 530 aa (59.7 kDa), 430 aa (49.3 kDa), and 506 aa (57.9 kDa), respectively. Their respective pI were 9.26, 7.72, and 9.33, and their instability index (46.46, 49.84, and 48.61) revealed that SiC4H proteins were stable. Among the Si4CL proteins, Si4CL4 and Si4CL10 were the smallest proteins with 198 aa (21.8 and 21.7 kDa in size, respectively) and the largest was Si4CL9 (642 aa; 68.5 kDa). Their pI range was between 5.14 and 8.98. The protein instability index ranged from 24.76 (Si4CL3) to 47.96 (Si4CL6) hinting that all the Si4CL proteins except Si4CL3 were stable. SiHCT, SiC3H, and SiF5H proteins have two members each, with a narrow range of protein properties, and all these proteins were found to be stable as indicated by their stability index. A significant difference was observed with the sizes of SiF5H members since SiF5H1 was 158 aa (16.7 kDa) and SiF5H2 was 524 aa (57.7 kDa) (**Supplementary Table S6**). Among SiCCoAOMT proteins, the smallest protein was SiCCoAOMT1 with 243 aa (25.7 kDa) and the largest was SiCCoAOMT5 with 307 aa (33.4 kDa). The pI range was between 5.04 and 8.94, and the protein instability index range (27.69–51.49) showed that except SiCCoAOMT4, all others were stable. The three-member SiCOMT class proteins have molecular sizes of 247 aa (25.8 kDa; SiCOMT1), 402 aa (43.53 kDa; SiCOMT2), and 153 aa (16.71 kDa; SiCOMT3). The pI values were 5.09, 5.97, and 9 for SiCOMT1, SiCOMT2, and SiCOMT3, respectively. The instability index range (42.24–52.75) hinted that all SiCOMT proteins are stable. Among the monolignol pathway proteins, SiCCR class has the highest number (26 members) and their sizes ranged from 27.2 kDa (251 aa; SiCCR26) to 69.13 kDa (625 aa; SiCCR9), with a pI range of 4.72 (SiCCR23) to 9.32 (SiCCR19). The protein instability index ranged from 24.86 (SiCCR18) to 54.11 (SiCCR13), which points out that ~77% of SiCCR proteins were stable. In the case of SiCAD proteins, SiCAD9 and SiCAD13 were the smallest proteins with 336 aa (35.6 and 36.4 kDa in size, respectively) and SiCAD8 was the largest with 495 aa (52.7 kDa). The pI ranged from 5.05 to 9.24, and the instability index (19.35–39.79) showed that ~50% of SiCAD proteins are unstable (**Supplementary Table S6**).

## Sequence Alignment and Phylogenetic Analysis of Cesa/Csl and Gsl Proteins

SiCesA and SiCsl proteins were aligned individually, and the alignment revealed the presence of conserved “DXD, D,

QXXRW” motif in both the superfamilies. All the SiCesA proteins except SiCesA8 have a “DCD, D, QVLRW” consensus sequence, whereas SiCesA8 had a unique “DYD, D” sequence and the motif “QXXRW” was absent (**Supplementary Figure S1**). Noteworthy, SiCesA8 protein has only the CS domain, while the other SiCesA proteins possess CS, ZF, and GT2 domains (**Supplementary Table S3**). In the case of SiCsl proteins, the “DXD” motif is absent in all the members of SiCslA, SiCslC and SiCslE2 (**Supplementary Figure S2**). This motif was predominantly “DCD,” except in SiCslF1 and SiCslF2, which have “DGD.” The second consensus “D” amino acid is present in all the SiCsl members (as “ED”), except SiCslA6, SiCslE2, and SiCslF4 (**Supplementary Figure S2**). In addition, SiCslA6 and SiCslE2 did not possess the “QXXRW” motif also, whereas a subgroup-wise conservation was evidenced in this motif in rest of the members. The majority of SiCslA (7) and all the SiCslC members have “QXHRW” motif, whereas SiCslE proteins have “QHKRW,” SiCslH and SiCslJ proteins have “QYKRW” and “QNKRW” motifs, respectively (**Supplementary Figure S2**). The unrooted phylogenetic tree constructed using the amino acid sequences of SiCesA/Csl proteins along with Cesa/Csl proteins of rice and *Arabidopsis* (<https://cellwall.genomics.purdue.edu/intro/index.html>) showed 2 distinct clusters, namely I and II (**Figure 1**). Cluster I was resolved into six branches including Cesa, CslD, CslE, CslF, CslH, and CslJ, whereas cluster II had two branches, CslA and CslC.

Sequence alignment of SiGsl proteins showed that the N-terminal region of all these proteins was diverse, whereas the C-terminal region was conserved (**Supplementary Figure S3**). Prediction of transmembrane (TM) helices in these proteins using TMHMM Server v2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) showed the presence of 7–16 TM helices in SiGsl proteins (**Supplementary Figure S4**). Phylogeny of foxtail millet, rice and *Arabidopsis* Gsl proteins showed three clusters (**Figure 2**). Cluster I included SiGsl4, SiGsl5, and SiGsl7, whereas cluster II comprised SiGsl2 and SiGsl3. SiGsl1, SiGsl6, SiGsl8, SiGsl10, SiGsl11, and SiGsl12 were included in cluster III.

## Sequence Alignment and Phylogenetic Analysis of Monolignol Biosynthesis Pathway Proteins

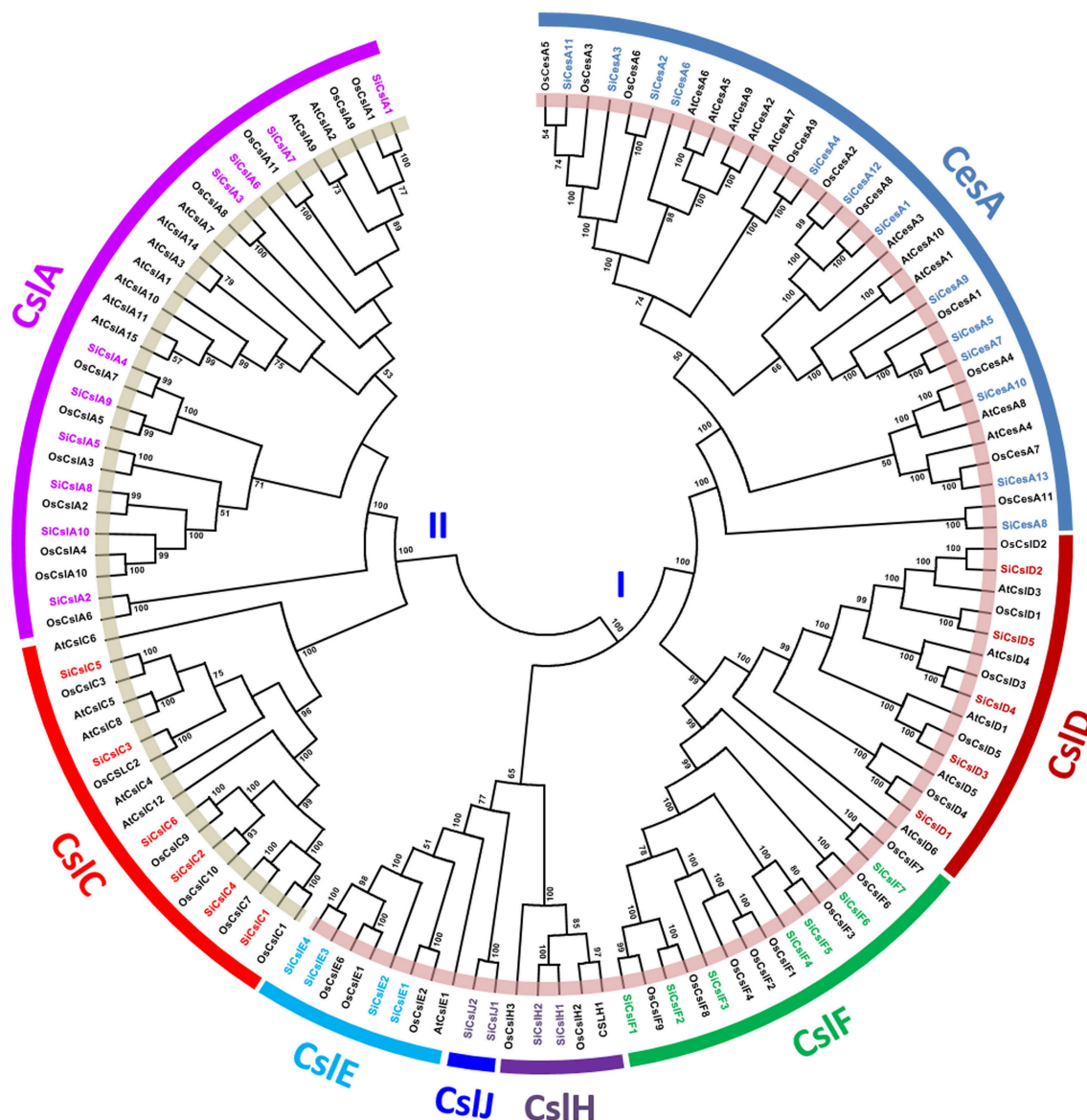
Sequence alignment and analysis of SiPAL proteins showed that all the members are almost completely conserved (**Supplementary Figure S5**). SiPAL2 was found to possess an extended N-terminal sequence of about 135 amino acids, which is unique to this class of protein. A phylogenetic tree constructed with PAL sequences of foxtail millet, eucalyptus, poplar, tobacco, medicago and *Arabidopsis* showed that the SiPAL proteins are phylogenetically divergent from the rest (**Figure 3A**). Sequence alignment of SiC4H showed that all the members share the conserved P450 superfamily domain and P450-featured motifs, namely, haem-iron binding motif (PFGVGRRSCPG), the T-containing binding pocket motif (AAIETT, the E-R-R-E-R-E-R), for optimal orientation of the enzyme (**Supplementary Figure S5**). Further, presence of conserved substrate recognition sites (SRSs)

of C4H/CYP73A5 enzymes, including SRS1 (SRTRNVVFDIFTGKGQDMVFTVY), SRS2 (LSQSFEYNY), SRS4 (IVENINVAAIETTLWS), and SRS5 (RMAIPLLVPH) was also evidenced (**Supplementary Figure S5**). Phylogeny of SiC4H along with C4H protein sequences of other organisms showed the grouping of SiC4H1 with C4H1 proteins of eucalyptus and *Phaseolus vulgaris*, whereas SiC4H2 and SiC4H3 were found to be more divergent (**Figure 3B**).

Si4CL protein sequence alignment showed the presence of 2 highly conserved peptide motifs “box I” (LPYSSGTTGLPKGV; AMP binding signature) and “box II” (GEICIRG), in addition to other conserved regions (**Supplementary Figure S5**). Phylogeny of 4CL proteins showed grouping of Si4CL1, Si4CL2, Si4CL15, and Si4CL16 with switchgrass (Pvi4CL1), demonstrating their close proximity and similarly, Si4CL11 was found to be grouped with Pvi4CL2, whereas other Si4CL proteins formed their own distinct cluster (**Figure 3C**). Alignment of SiHCT sequences showed that all the proteins have the conserved motifs for the acyl transferase family, namely “HXXXDG” and “DFGWG” (**Supplementary Figure S5**). Multiple sequence alignment of SiC3H proteins showed the presence of Cytochrome P450 cysteine heme-iron ligand signature [FW]-[SGNH]-x-[GD]-{F}-[RKHPT]-{P}-C-[LIVMFAP]-[GAD] (**Supplementary Figure S5**). The conserved motifs including three putative S-adenosyl-L-methionine binding motifs (A, B, and C) and CCoAOMT signature motifs (D, E, F, G, and H) were identified through multiple sequence alignment of SiCCoAOMT proteins (**Supplementary Figure S5**). Phylogenetic analysis of SiHCT, SiC3H, and SiCCoAOMT proteins with their respective family members of other organisms revealed the dissimilarity of foxtail millet proteins compared to their homologs (**Figures 3D–F**). In the case of CCoAOMT, SiCCoAOMT2 formed a distinct clade, whereas other SiCCoAOMT members were grouped together in one clade (**Figure 3F**).

Being truncated proteins, alignment of SiF5H1 with SiF5H2, and SiCOMT2 with SiCOMT1 and SiCOMT3 were not performed (**Supplementary Figure S5**). Protein sequence alignment between SiCOMT1 and SiCOMT3 did not highlight any consensus motif and their phylogenetic analysis with COMT proteins of other plants showed grouping of SiCOMT with ZmaCOMT of maize (**Figure 3G**). Sequence alignment of SiCCR proteins revealed that the conserved “KNWYCYGK” motif, catalytic site or the binding site for the cofactor NADPH (Larsen, 2004) has been diversified in foxtail millet (**Supplementary Figure S5**). Except SiCCR1 and SiCCR24, other SiCCR proteins have at least one amino acid change in this motif, which could be attributed to the substrate affinity of CCR proteins (Pichon et al., 1998). Phylogenetic analysis of SiCCR proteins showed that a maximum of these proteins were clustered in a separate group, whereas few proteins were grouped with CCR proteins of maize, switchgrass and poplar (**Figure 3H**). Alignment results of SiCAD highlighted a high degree of similarity in conserved domains and binding residues, including Zn-1 binding domain motif GHE(X)<sub>2</sub>G(X)<sub>5</sub>G(X)<sub>2</sub>V, NADP(H) co-substrate-binding motif GXG(X)<sub>2</sub>G (glycine-rich repeat) and Zn-2 metal ion binding motif GD(X)<sub>9,10</sub>C(X)<sub>2</sub>C(X)<sub>2</sub>C(X)<sub>7</sub>C





**FIGURE 1 |** Unrooted protein phylogenetic tree constructed with CesaA/Csl proteins of *Setaria italica* (Si), *Oryza sativa* (Os), and *Arabidopsis thaliana* (At).

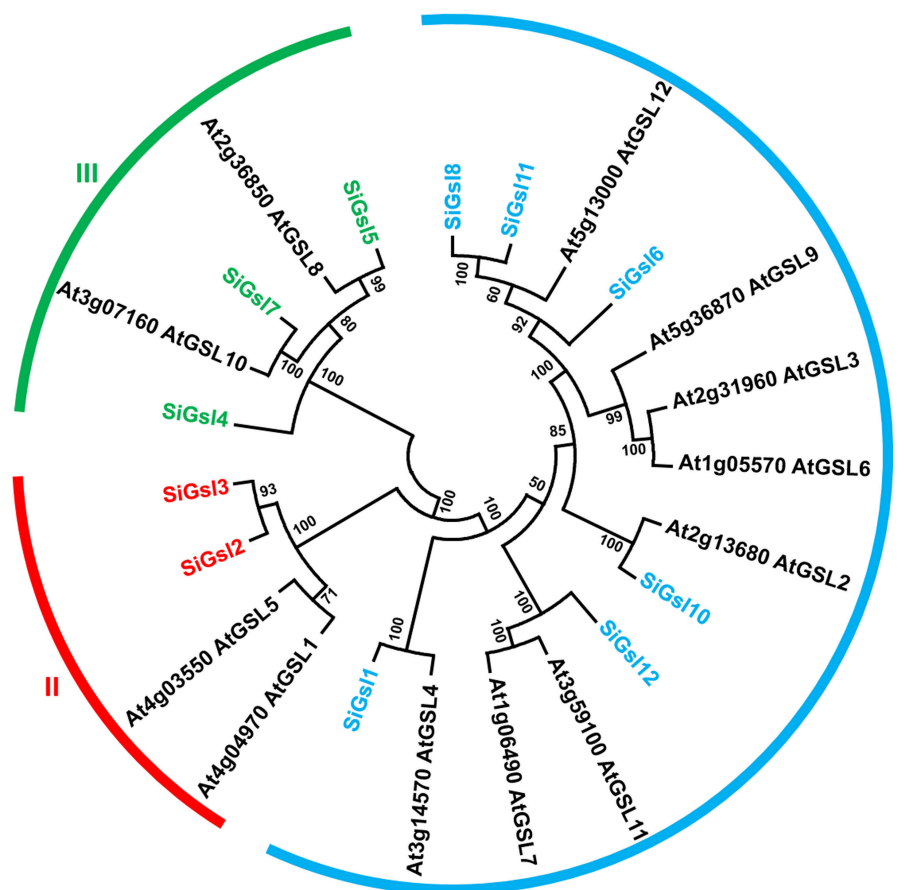
(Supplementary Figure S5). Phylogenetic tree of SiCAD with CAD proteins of other plant species showed clustering of a maximum of SiCAD proteins in one clade with complete out-grouping of SiCAD10. SiCAD1 and SiCAD11 were found to cluster with poplar CAD proteins (Figure 3I).

## Gene Structure of Lignocellulose Pathway Genes

The sequence data of genomic DNA, transcript and CDS along with chromosomal locations of confirmed protein sequences of identified lignocellulose biosynthesis pathway enzymes were retrieved and analyzed for gene size, intron-exon and physical position (Supplementary Tables S2, S6). The size of *SiCesa*

genes ranged from 3.1 (*SiCesa8*) to 6.9 kb (*SiCesa9*) and few genes including *SiCesa3*, *SiCesa7*, *SiCesa5*, and *SiCesa9* have a maximum of 13 introns, whereas *SiCesa12* was intronless (Supplementary Figure S6). The gene sizes of *SiCsl* ranged from 1.7 (*SiCslA6* and *SiCslE2*) to 6.6 kb (*SiCslA1* and *SiCslF6*), and their gene structure analysis revealed that *SiCsl* genes have up to eight introns (Supplementary Figure S7). The only intronless gene of *SiCsl* superfamily was *SiCslE2*. Among the *SiGsl* gene family members, *SiGsl3* was the smallest gene (3.2 kb), whereas the largest one was *SiGsl4* (17 kb). Interestingly, *SiGsl* genes were evidenced to contain numerous introns. *SiGsl7* has a maximum of 49 introns, whereas *SiGsl2* and *SiGsl3* were intronless (Supplementary Figure S8).



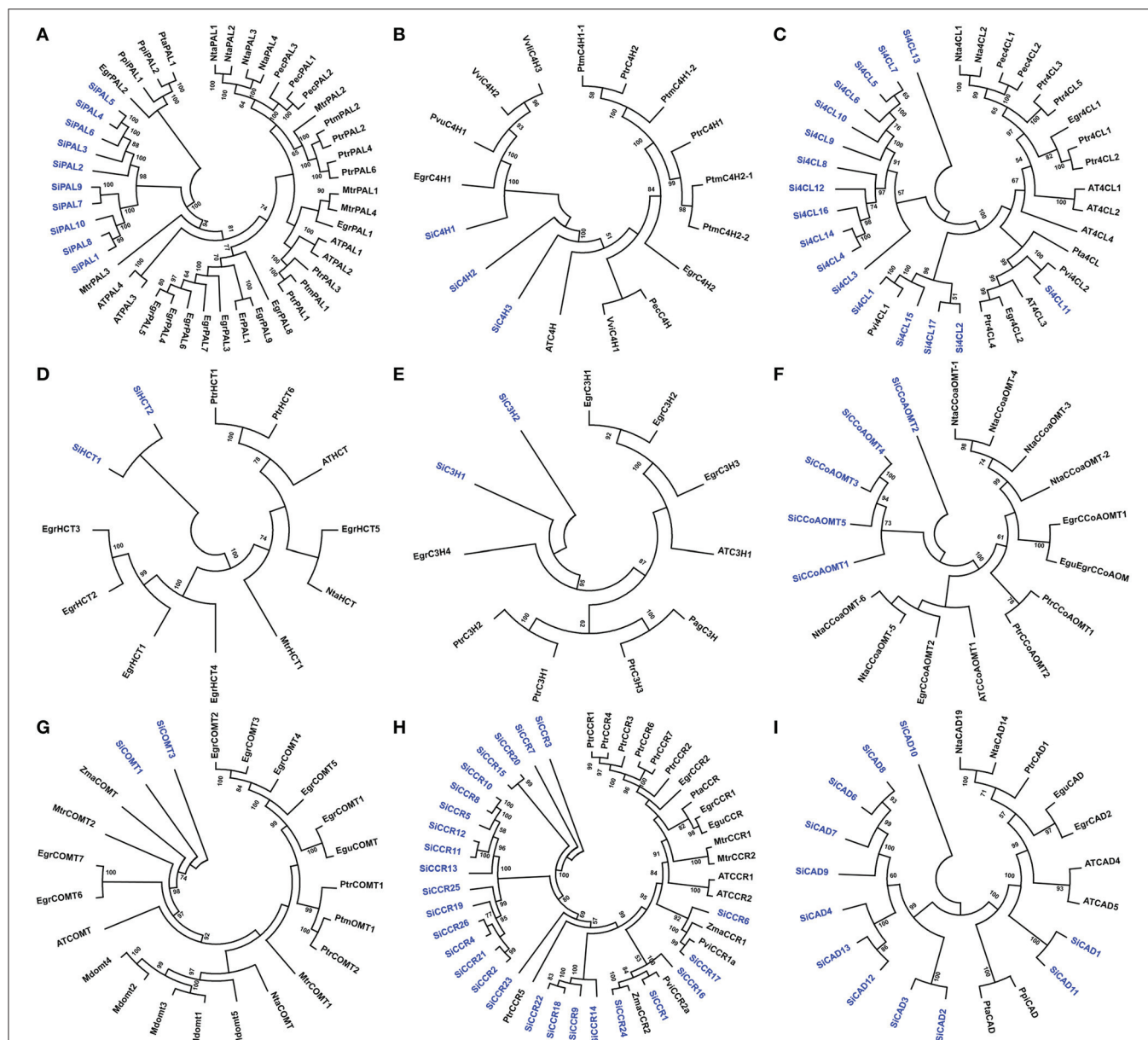


**FIGURE 2 |** Unrooted protein phylogenetic tree constructed with Gsl proteins of *Setaria italica* (Si) and *Arabidopsis thaliana* (At).

*SiPAL* gene sizes ranged from 2.1 (*SiPAL4*) to 4.6 kb (*SiPAL3*), of which *SiPAL4*, *SiPAL5*, and *SiPAL6* were intronless, *SiPAL2* has two introns and other *SiPAL* genes have 2 introns each (**Supplementary Figure S9**). Among the *Si4CL* genes, *Si4CL3* was the smallest gene (2 kb), whereas *Si4CL15* was the largest (6.7 kb). A total of 10 *Si4CL* genes have 5 introns each, while maximum number of introns was found in *Si4CL5* (6 introns). *Si4CL3* has the least number of one intron in its gene (**Supplementary Figure S9**). The size of *SiCCoAOMT* genes ranged from 0.8 (*SiCCoAOMT4*) to 3 kb (*SiCCoAOMT2*) with a maximum number of introns (7) in *SiCCoAOMT2*. *SiCCoAOMT3* and *SiCCoAOMT4* have one intron each (**Supplementary Figure S9**). Among the *SiCCR* genes, *SiCCR3* was 1.3 kb in size and though it is the smallest gene of this class, it has eight introns. *SiCCR9* and *SiCCR22* are the largest genes with a size of 5.8 kb and both the genes have 4 introns each. *SiCCR2* has a maximum of 10 introns, while *SiCCR7* is the only intronless gene in this group. The size of *SiCAD* genes ranged from 1.4 (*SiCAD9*) to 4.2 kb (*SiCAD1* and *SiCAD8*), with *SiCAD7*, *SiCAD8*, and *SiCAD9* having a minimum of 2 introns each whereas *SiCAD5* has a maximum of 6 introns (**Supplementary Figure S9**).

## Chromosomal Location and Gene Duplication of Lignocellulose Pathway Genes

The identified secondary cell wall biosynthesis genes were plotted onto the nine chromosomes of foxtail millet to generate the physical map (**Figure 4**), which showed that the majority of lignocellulose biosynthesis pathway genes (31; ~22%) were present in chromosome 2, followed by chromosome 9 (24 genes; ~17%) and chromosome 1 (21 genes; ~15%), and a minimum of 4 genes (~3%) were mapped on chromosome 8. Expansion of respective gene families within the genome were analyzed by investigating tandem and segmental duplication, which showed that 7 genes underwent tandem duplication, whereas segmental duplication did not occur among the lignocellulose pathway genes (**Figure 4**). *SiCesA* members were distributed on chromosomes 2 (4 genes), 4 (1), 5 (2), and 9 (3) and none of the genes were evidenced to undergo tandem or segmental duplication. *SiCsl* genes were found to be present in all the chromosomes except chromosome 8, and duplication analysis revealed that *SiCslE3* and *SiCslE4* were tandemly duplicated gene pairs on chromosome 2. *SiGsl* members were distributed on chromosomes 1 (2 genes), 2 (1), 4 (2), 5 (4),

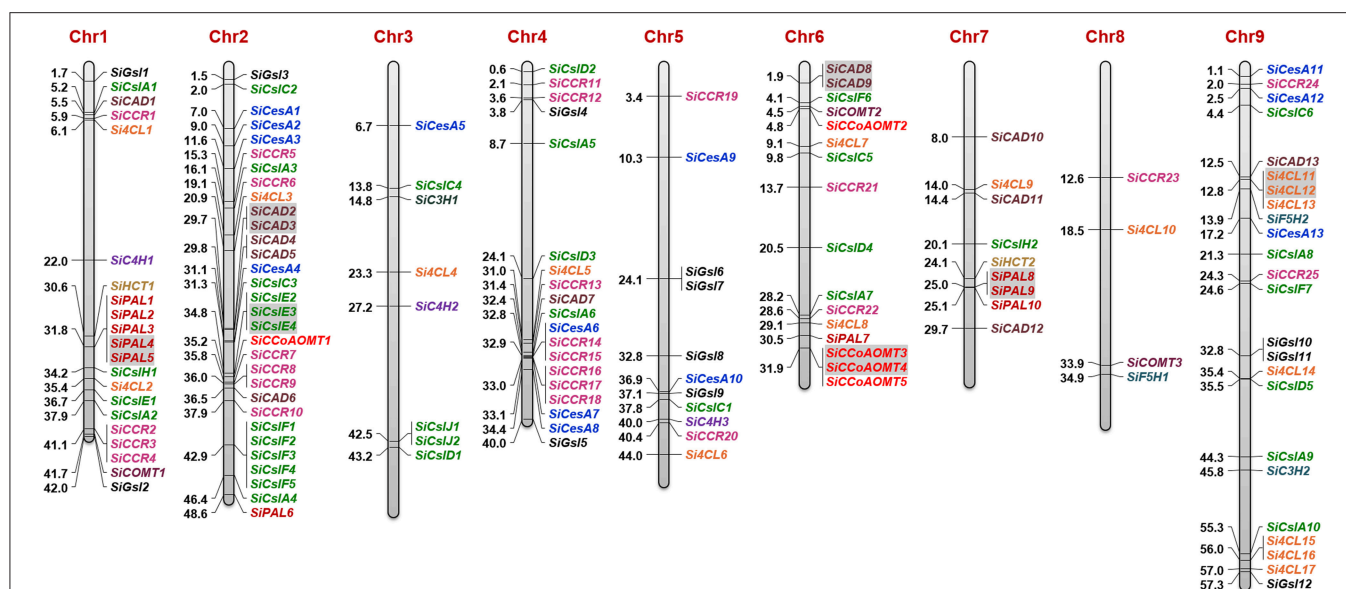


**FIGURE 3 |** Unrooted protein phylogenetic trees constructed with (A) PAL, (B) C4H, (C) 4CL, (D) HCT, (E) C3H, (F) CCoAOMT, (G) COMT, (H) CCR, and (I) CAD proteins of *Setaria italica* (Si), *Eucalyptus gunnii* (Egu), *E. grandis* (Egr), *Nicotiana tabacum* (Nta), *Populus trichocarpa* (Ptr), *Pinus pinaster* (Ppi), *Pinus taeda* (Pta), *Medicago truncatula* (Mtr), *Panicum virgatum* (Pvi), *Zea mays* (Zma), *Malus domestica* (Mdom), *Vitis vinifera* (Vvi), *Eucalyptus globulus* (Egl), *Populus alba* x *Populus grandidentata* (Pag), *Petroselinum crispum* (Pec), *Populus tremuloides* (Ptm), *Phaseolus vulgaris* (Pvu), and *Eucalyptus robusta* (Er).

and 9 (3) and no duplication pattern in this gene family was observed.

Among the monolignol biosynthesis genes, the majority of *SiPAL* genes were present in chromosome 1 (5) and 7 (3), and interestingly, *SiPAL4* and *SiPAL5* as well as *SiPAL8* and *SiPAL9* were identified to be tandem duplicates. Each of the three *SiC4H* genes were found in chromosome 1, 3, and 5 (Figure 4). A higher number of *Si4CL* genes were present in chromosome 9 (7 genes), of which *Si4CL11* and *Si4CL12* were tandemly

duplicated gene pairs. Chromosome 1 and 6 have two *Si4CL* members each and one member each in chromosome 2, 3, 4, 5, 7, and 8. Two members of *SiHCT*, *SiC3H*, and *SiF5H* as well as three genes of *SiCOMT* were present in chromosome 1, 3, 6, 7, 8, and 9 (Figure 4). Four out of five *SiCCoAOMT* genes were present in chromosome 6 and *SiCCoAOMT1* was mapped on chromosome 2, and duplication analysis revealed that *SiCCoAOMT3* and *SiCCoAOMT4* were tandemly duplicated gene-pairs. Among the *SiCCR* genes, *SiCCR26* could not be



**FIGURE 4 | Physical map showing the chromosomal locations of lignocellulose biosynthesis genes.** Bars represent chromosomes and the numbers at the left corresponds to location (in Mb). Gene IDs are provided in the right. Tandemly duplicated gene pairs are highlighted with gray shade.

mapped due to non-availability of its co-ordinates in Phytozome database. Of the 25 *SiCCR* genes mapped, a maximum of 8 genes were found to be present in chromosome 4 (8), followed by chromosome 2 (6) and 1 (4). Of the 13 *SiCAD* genes, maximum was in chromosome 2 (5) and a minimum of one each in chromosomes 1, 4, and 9. *SiCAD2* and *SiCAD3* on chromosome 2 as well as *SiCAD8* and *SiCAD9* on chromosome 6 were found to be tandemly duplicated gene-pairs (Figure 4).

## Promoter Analysis on Lignocellulose Pathway Genes

*In silico* analysis for predicting putative *cis*-regulatory elements showed the presence of universal as well as gene-specific promoter sequences in the upstream of lignocellulose pathway genes (Supplementary Tables S8, S9). A total of 271 *cis*-elements were found in *CesA/Csl* and *Gsl* genes, of which 15 (5.5%) elements, namely ACGTATERD1, ARR1AT, CAATBOX1, CACTFTPPCA1, DOFCOREZM, EBOXBNNAPA, GATABOX, GT1CONSENSUS, GTGANTG10, MYCONSENSUSAT, NODCON2GM, OSE2ROOTNODULE, POLLEN1LELAT52, WBOXNTERF3, and WRKY71OS were present in all these genes (Supplementary Table S8). Thirty-nine unique *cis*-elements (~14%) which were present in any one gene of *CesA/Csl* and *Gsl* superfamilies were also found, such as ABADES11 (*SiCslF6*), CEREBLUBOX3PSLEGA (*SiCesA2*), GBOXLERBCS (*SiCslA8*), ZDNAFORMINGATCAB1 (*SiCslA6*), TATCCACHVAL21 (*SiGsl3*), etc. In addition, few promoter sequences were found to be present in all the genes except one or two genes and this includes BIHD1OS (*SiCslC4*), CCAATBOX1 (*SiCslA1*, *SiCslF4*), CURECORECR (*SiCslC3*, *SiGsl1*), DPBFCORED CDC3 (*SiCslC2*, *SiCslC4*), EECRCRAH1 (*SiGsl5*), MYBCORE (*SiCesA3*, *SiCslD4*), RAV1AAT (*SiCslD1*), and SORLIP1AT (*SiCesA4*, *SiGsl8*). Of

note, no superfamily specific regulatory elements were identified (Supplementary Table S8).

A total of 293 *cis*-elements were detected in the upstream region of monolignol pathway genes, of which 10 (3.4%) were present in all the genes and 37 (~13%) were unique to any one gene (Supplementary Table S9). The elements which were present in all the genes include ARR1AT, CAATBOX1, CACTFTPPCA1, DOFCOREZM, EBOXBNNAPA, GATABOX, GT1CONSENSUS, GTGANTG10, WBOXNTERF3, and WRKY71OS. Few *cis*-regulatory elements were found to be present in all except one or two genes and it includes ACGTATERD1 (*SiPAL2*), CURECORECR (*SiPAL2*, *SiPAL10*), and MYBCORE (*SiPAL7*, *SiCCR16*). Similar to *CesA/Csl* and *Gsl*, no monolignol genes have superfamily specific regulatory elements (Supplementary Table S9).

## MicroRNAs and Molecular Markers of Lignocellulose Pathway Genes

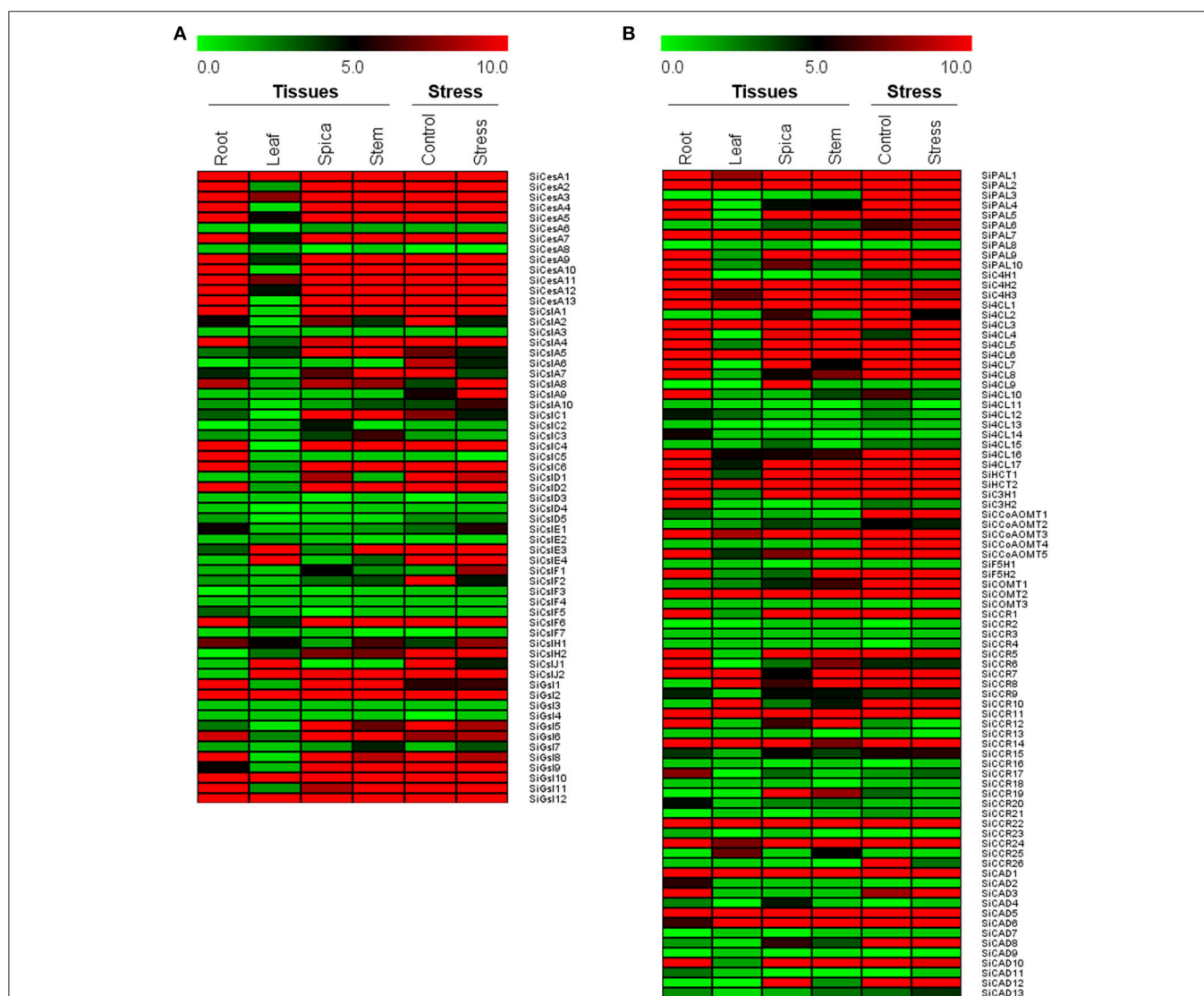
*In silico* scanning of lignocellulose pathway gene transcripts to identify their targeting miRNAs showed that the transcripts of *SiCslC2*, *SiGsl10*, and *SiF5H2* could be targeted by the miRNAs sit-miRn29, sit-miR114-npr and sit-miR395b, respectively (Supplementary Table S10). *SiGsl3* was predicted to be targeted by two foxtail millet miRNAs, namely sit-miR156d-1 and sit-miR156d-2. These miRNAs would have a putative role in post-transcriptional gene silencing for regulation of lignocellulose pathway gene expression. Identification of previously reported molecular markers in the genic and regulatory regions of lignocellulose pathway genes revealed the presence of SSR and ILP markers in 34 genes (Supplementary Table S11). Of these, three genes have two and three markers each, and the remaining

28 genes possess single markers. Among the markers, SSRs were found to be predominant (~81%) and the rest are ILPs (~19%).

## Expression Profile of Lignocellulose Pathway Genes in Tissues and Dehydration Stress

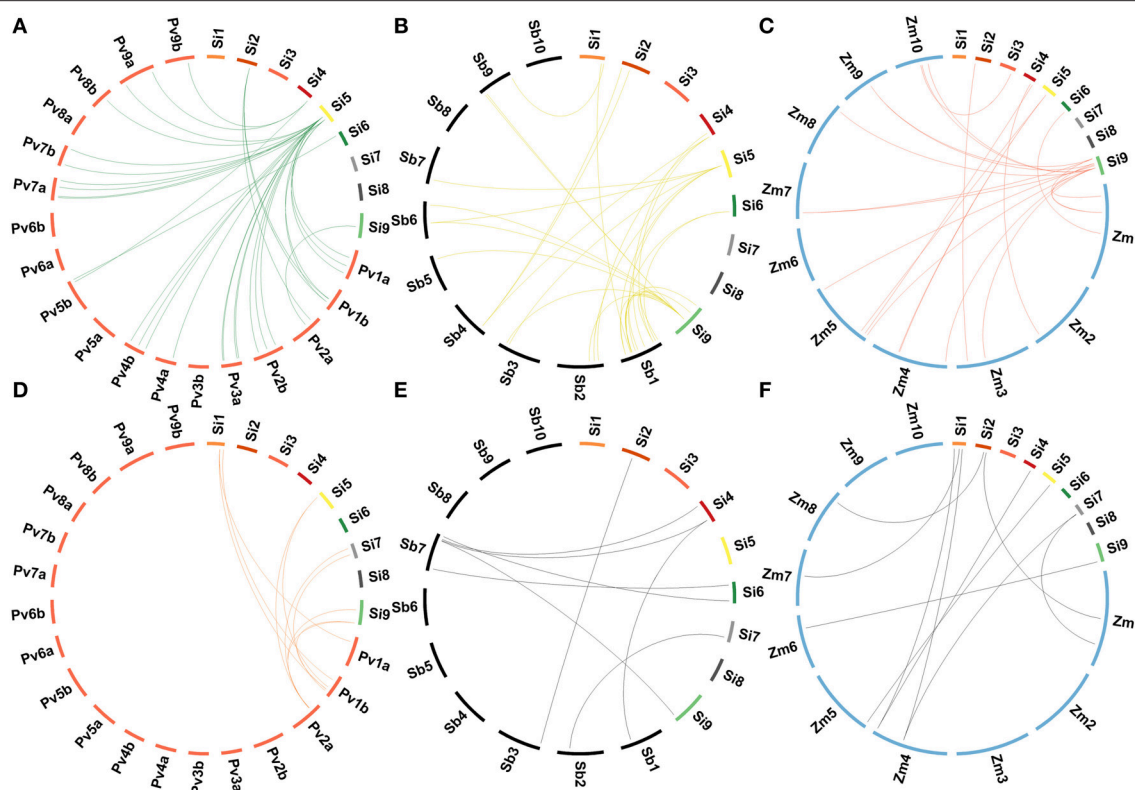
Expression of all the genes in four tissues and dehydration stress was calculated using RPKM values derived from RNA-seq data. Tissue-specific expression profile showed differential expression pattern of all the genes with relatively lower expression in leaf (Figure 5). In the case of *CesA/CsI* and *Gsl* superfamilies, higher expression of *SiCesA1*, *SiGsl2*, *SiGsl10*, and *SiGsl12* was evidenced in all the four tissues when compared to the other members of the same gene family. Tissue-specific higher expression of *SiCslD1* in spica, and *SiCslE4* and *SiCslJ2* in leaf was also

observed. Many genes including *SiCesA6*, *SiCesA8*, *SiCslA3*, *SiCslC3*, *SiGsl3*, and *SiGsl7* were not expressed in these tissues (Figure 5A). Tissue-specific expression profiling of monolignol genes showed higher expression of *SiPAL1*, *SiPAL2*, *SiPAL7*, *SiC4H2*, *Si4CL1*, *Si4CL3*, *Si4CL6*, *SiHCT2*, *SiCOMT2*, *SiCCR11*, *SiCAD1*, and *SiCAD5* in all the four studied tissues. Tissue-specific higher expression was evidenced with *SIPAL4*, *Si4CL10*, and *SiCAD3* in root, and *Si4CL9* and *SiCAD12* in spica. Similar to *CesA/CsI* and *Gsl*, monolignol genes also showed a relatively lower expression in leaf tissue (Figure 5B). Expression profiling of all the genes in response to dehydration stress showed almost a uniform expression in both control and stress samples (Figure 5). Comparison of expression patterns between tissues and stress library revealed that the expression of predominant lignocellulose pathway genes was unaltered. Only three genes, namely *SiCslA8*,



**FIGURE 5 |** Heat map showing the expression of (A) cellulose biosynthesis genes, and (B) monolignol biosynthesis genes in four different tissues and dehydration stress library. The Illumina RNA-seq data were re-analyzed and the heat map was generated. Bar at the top with the values 0.0, 5.0, and 10.0 represent low, intermediate and high expression, respectively.





**FIGURE 6 |** Comparative genome map showing homologous relationships between *CesA/CsI* and *Gsl* superfamilies of *Setaria italica* and (A) *Panicum virgatum*, (B) *Sorghum bicolor*, (C) *Zea mays*, and between monolignol biosynthesis genes of *Setaria italica* and (D) *Panicum virgatum*, (E) *Sorghum bicolor*, (F) *Zea mays*.

*SiCslA9*, and *Si4CL4* showed a higher expression in dehydration stress library compared to control, of which *SiCslA8* and *SiCslA9* were expressed only during stress and not in any of the tissue-specific RNA-Seq libraries. Few genes which were highly expressed in control were observed to be down regulated during stress and this includes *SiCslA5*, *SiCslA6*, *SiCslA7*, *SiCslF2*, and *SiCCR26* (Figure 5).

## Homologous Relationships of Lignocellulose Pathway Genes with Other Grasses

Homologs of foxtail millet lignocellulose pathway genes in sequenced *C<sub>4</sub>* panicoid genomes, namely switchgrass (*Panicum virgatum*), sorghum (*Sorghum bicolor*), and maize (*Zea mays*) were derived (Figure 6). A maximum lignocellulose pathway gene-based homology was observed between foxtail millet and switchgrass as 19 genes of foxtail millet showed homology with 60 genes of switchgrass (Supplementary Table S12). Of the 19 foxtail millet genes, six belonged to *SiGsl*, four to *SiCCR*, three each to *SiCsl* and *SiPAL*, and one each to *SiHCT*, *Si4CL* and *SiCAD*. Eighteen foxtail millet genes showed orthologous relationship with 41 sorghum genes, of which *SiGsl11* had a maximum of 11 homologs, followed by *SiGsl7* (7 homologs) and *SiGsl5* and *SiCCR17* (3 homologs

each) (Supplementary Table S13). In the case of foxtail millet-maize homology, 26 foxtail millet genes showed homologous relationship with 38 maize genes (Supplementary Table S14). Among the foxtail millet genes, *SiGsl* had a maximum of 7 homologs in maize, followed by *SiGsl7* (3 homologs).

Among the lignocellulose pathway proteins, CADs and COMTs were well characterized as they play key role in secondary cell wall lignification (Saballos et al., 2009; Saathoff et al., 2011a,b, 2012; Sattler et al., 2012; Trabucco et al., 2013). Sequence analysis of these proteins in several grasses identified the presence of conserved motifs in few members, which distinguish them as lignifying proteins from the rest of non-lignifying proteins. Lignifying CADs possess additional 12 amino acids T<sup>49</sup>, Q<sup>53</sup>, L<sup>58</sup>, M<sup>60</sup>, C<sup>95</sup>, W<sup>119</sup>, V<sup>276</sup>, P<sup>286</sup>, M<sup>289</sup>, L<sup>290</sup>, F<sup>299</sup>, and I<sup>300</sup>, which are involved in substrate recognition and binding (Youn et al., 2006). Of the 13 *SiCAD* proteins, *SiCAD11* contains 11 of 12 conserved amino acid residues. Of note, the active substrate-binding residues, W119 and F298, which determine specificity for aromatic alcohols and, the NADP(H) binding site, S212, were present in *SiCAD11*. Sequence-based homology analysis showed higher percentage of identity between *SiCAD11* and lignifying CADs of other grasses namely switchgrass (Pavir.J34526; 91%), sorghum (Sobic.006G211900; 89%) and maize (GRMZM5G844562; 85%). Similarly, the conserved amino acids M<sup>130</sup>, N<sup>131</sup>, L<sup>136</sup>, A<sup>162</sup>, H<sup>166</sup>, F<sup>176</sup>, M<sup>180</sup>, H<sup>183</sup>, I<sup>319</sup>, M<sup>320</sup>,

and N<sup>324</sup>, which function in substrate-binding and positioning in COMTs (Sattler et al., 2012; Trabucco et al., 2013) are found to be present in SiCOMT02 of foxtail millet. Sequence-based homology with SiCOMT02 showed high percent identity to sorghum (Sobic.007G047300; 94%), switchgrass (Pavir.Fa01907; 85%), and maize (AC196475.3; 89%).

## Duplication and Divergence of Lignocellulose Pathway Genes

The number of non-synonymous substitutions per non-synonymous site (Ka) and synonymous substitutions per synonymous site (Ks) was calculated for paralogous as well as homologous gene pairs and Ka/Ks ratio along with time of divergence (in million years ago; mya) were derived. The ratio of Ka to Ks for tandemly duplicated gene-pairs ranged from 0.09 to 0.18 with an average value of 0.13, which suggested that these genes were under strong positive purifying selection (Ka/Ks > 1) and the duplication event was predicted to occur around 25 mya (**Supplementary Table S15**). In the case of Ka/Ks ratio of homologous gene-pairs, it was maximum between foxtail millet-switchgrass (average Ka/Ks = 0.91; **Supplementary Table S12**), whereas foxtail millet-sorghum and foxtail millet-maize homologs showed an average ratio of 0.19 (**Supplementary Tables S13, S14**). Since these values were less than 1, it signifies the intense positive selective pressure acted on respective protein-coding genes. The time of divergence between foxtail millet and switchgrass was predicted to occur around 4.7 mya, whereas the divergence of foxtail millet-sorghum and foxtail millet-maize occurred around 27 mya. This demonstrates that duplication and divergence have played a key role in shaping the lignocellulose pathway multigene families in foxtail millet and other C<sub>4</sub> grass genomes.

## Expression Profile of Candidate Genes during Stress and Hormone Treatments

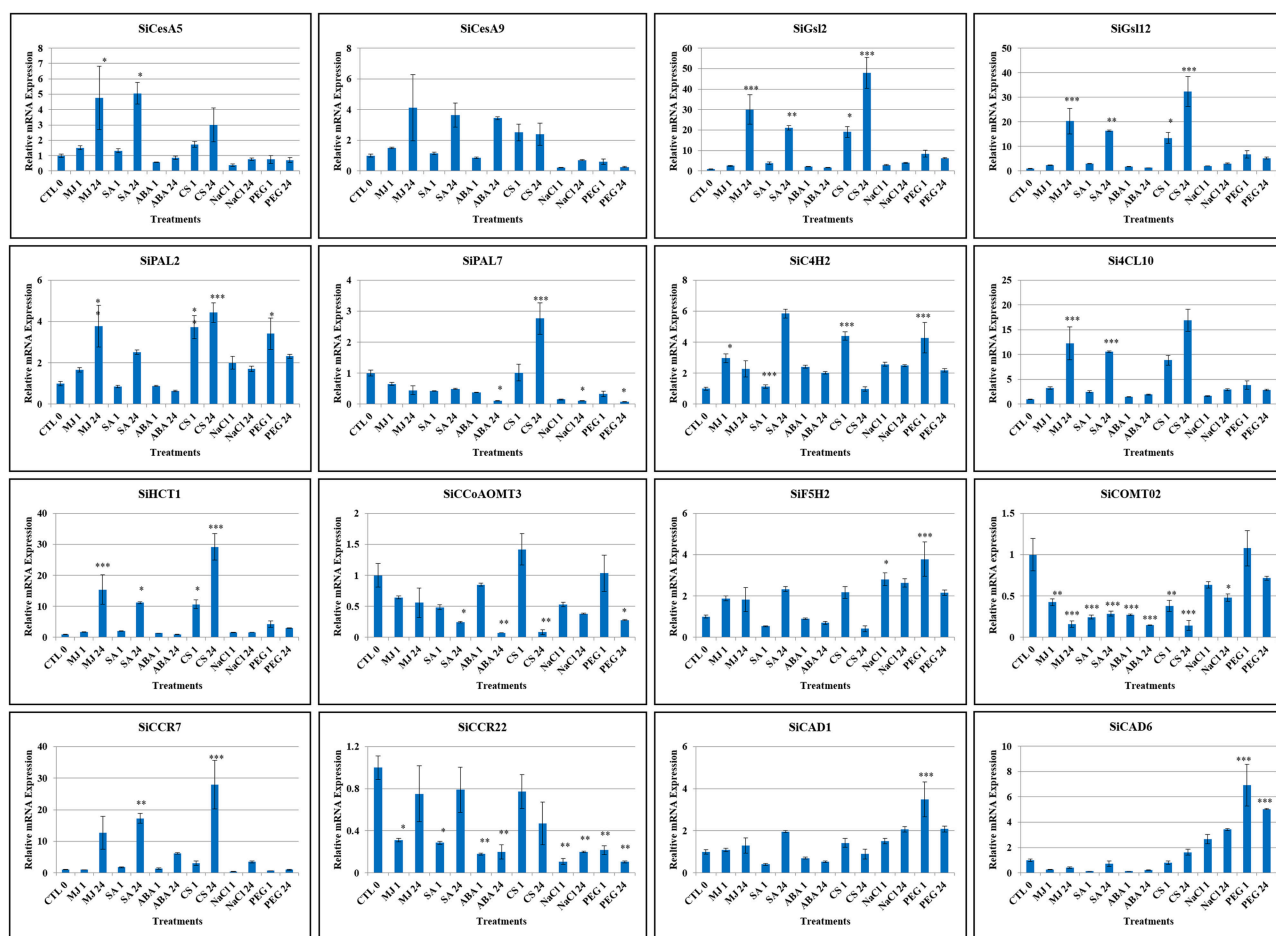
Expression patterns of sixteen candidate lignocellulose biosynthesis genes, namely *SiCesA5*, *SiCesA9*, *SiGsl2*, *SiGsl12*, *Si4CL10*, *SiPAL2*, *SiPAL7*, *SiC4H2*, *SiHCT1*, *SiCCoAOMT3*, *SiF5H2*, *SiCOMT2*, *SiCCR7*, *SiCCR22*, *SiCAD1*, and *SiCAD6* in response to stress (dehydration, salinity, cold) and hormone (abscisic acid, salicylic acid, methyl jasmonate) treatments was performed at two time points (1 h, early; 24 h, late). These candidates were chosen based on; (i) expression profiles deduced *in silico* using RNA-seq data, (ii) representing the nine chromosomes of foxtail millet, and (iii) their function in secondary cell wall formation such as *SiCOMT2* in lignification. Overall, the study demonstrated differential expression pattern of these genes during stress and hormone treatments except *SiCCR22* which was found to be down-regulated under all conditions (**Figure 7**). *SiGsl2* and *SiGsl12* were found to be highly expressed during all the three stress conditions, whereas *SiCAD6* was up-regulated during both salinity and dehydration stress. Dehydration stress has been observed to induce the expression of all the genes except *SiCCoAOMT3*, *SiCOMT2*, *SiCesA5*, *SiCCR22*, *SiPAL7*, *SiCCR7*, and *SiCesA9*, though the degree of expression varied between the genes. Salinity stress

showed an induction in expression of *SiC4H2*, *SiCAD6*, *SiF5H2*, *SiGsl12*, and *SiGsl2*, while *SiPAL2* was induced during early salt stress and *SiCAD01*, *Si4CL10*, and *SiCCR7* were found to be up-regulated in late phase salinity stress, thus suggesting a significant higher expression among the members of *SiGsl* and *SiCAD* family. Significant up-regulation of *SiGsl2*, *SiGsl12*, *Si4CL10*, *SiHCT1*, and *SiCCR7* during cold stress suggests the putative involvement of these genes in strengthening the cell wall for tolerance to low temperature. Higher expression of these genes was also found during both early and late phases of treatment with salicylic acid and methyl jasmonate. Differential expression of candidate genes was observed during the treatment of all the hormones except abscisic acid, which showed no effect on the expression of majority of candidate genes except *SiGsl2*, which was induced at early phase of ABA treatment, *SiCCR7* and *SiCes9*, which were induced at late phase of ABA treatment, and *SiC4H2*, which was induced at both the phases of ABA treatment. In addition, expression of *SiCCoAOMT3*, *SiCOMT2*, *SiCCR22*, *SiPAL7*, *SiCAD1*, and *SiCAD6* was found to be down-regulated during hormone treatments, while *SiF5H2* was up-regulated only under late phase of salicylic acid treatment.

## DISCUSSION

Cellulose, hemicelluloses and lignin constitute the complex polymeric structure of secondary cell wall and the lignocellulose biosynthesis pathway involves the action of *cellulose synthase* (*CesA*), *cellulose synthase-like* (*Csl*), *glucan synthase-like* (*Gsl*), *phenylalanine ammonia lyase* (*PAL*), *trans-cinnamate 4-hydroxylase* (*C4H*), *4-coumarate CoA ligase* (*4CL*), *hydroxycinnamoyl CoA:shikimate/quinic acid hydroxycinnamoyl transferase* (*HCT*), *p-coumaroyl shikimate 3'-hydroxylase* (*C3H*), *caffeoyl CoA 3-O-methyltransferase* (*CCoAOMT*), *ferulate 5-hydroxylase* (*F5H*), *caffeic acid O-methyltransferase* (*COMT*), *cinnamoyl CoA reductase* (*CCR*), and *cinnamyl alcohol dehydrogenase* (*CAD*) genes, which are well studied in several crop plants as well as trees for understanding and improving biofuel traits (Zhong and Ye, 2015). In the present study, all these gene families in foxtail millet were systematically identified and characterized using *in silico* approaches, and expression profiling of chosen genes was performed in response to several stress as well as hormonal treatments for identifying target genes for functional characterization.

A total of 13 *CesA* and 36 *Csl* genes were identified in foxtail millet, and all the *SiCesA* proteins were found to possess the characteristic cellulose synthase (CS) domain and 12 *SiCesA* had an additional zinc finger (ZF) structure. Similarly, 11 *CesA* proteins have been reported in rice, of which 9 contained both CS and ZF domain, and 2 lacked ZF domain (Wang et al., 2010). Role of *CesA* proteins in cellulose biosynthesis in both primary and secondary cell walls has been well dissected in *Arabidopsis*. In this plant, 10 *CesA* genes have been identified (Richmond and Somerville, 2000), of which *AtCesA1*, *AtCesA3*, and *AtCesA6* were reported to be involved in primary cell wall cellulose synthesis (Persson et al., 2007), *AtCesA4*, *AtCesA7*, and *AtCesA8* in secondary cell wall development, and *AtCesA2*,



**FIGURE 7 | Relative expression of candidate lignocellulose biosynthesis genes analyzed using qRT-PCR under dehydration (PEG), salinity (NaCl) and cold stress (CS) as well as abscisic acid (ABA), salicylic acid (SA) and methyl jasmonate (MJ) treatments for 0 (Control: CTL), 1 and 24 h. *Act2* was used as an internal control to normalize the data.** The error bars representing standard deviation were calculated based on three technical replicates for biological triplicates. Statistical analysis between treatment and control using Tukey-Kramer multiple comparisons test has been performed and the differences in the effects of stress treatments in all the genes were considered statistically significant at \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

*AtCesA5*, *AtCesA9*, and *AtCesA10* in tissue-specific cellulose biosynthesis processes (Gardiner et al., 2003; Taylor et al., 2003). Recent functional characterization of *AtCesA* proteins led to the identification of unidirectional movement of these protein complexes in seed coat epidermal cells, which deposit cellulose that are involved in mucilage extrusion, adherence and ray formation (Griffiths et al., 2015). In flax (*Linum usitatissimum*), 14 distinct *CesA* genes were identified and were targeted for silencing using virus-induced gene silencing (VIGS) approach, which showed impacts on outer-stem tissue organization and secondary cell wall formation (Chantreau et al., 2015). A genome-wide association study of single nucleotide polymorphisms (SNPs) developed through re-sequencing of diverse chickpea accessions revealed a superior haplotype and favorable natural allelic variants in the upstream regulatory region of a *CesA* gene, denoted as *Ca\_Kabuli\_CesA3* (Kujur et al., 2015). Interestingly, up-regulation of this superior gene haplotype resulted in higher transcript expression of *Ca\_Kabuli\_CesA3*

gene in pollen and pod of high pod/seed number chickpea accession, thus resulting in enhanced accumulation of cellulose (Kujur et al., 2015). The specific allelic variant caused cellulose changes specifically in pollen tubes of chickpea and therefore, investigating the homologous gene of foxtail millet identified in the present study will provide novel clues on its role, which could be manipulated for achieving greater biomass yield and bioconversion efficiency.

Physical map of *SiCesA* genes showed their distribution in chromosomes 2, 3, 4, 5, and 9, with a maximum of 4 genes in chromosome 4 and minimum of one gene in chromosome 3 (Figure 4). Extension of gene families is attributed to the occurrence of three major duplication mechanisms, namely segmental, tandem and retroposition (Cannon et al., 2004). However, none of these duplications were found to be involved in the expansion of *SiCesA* genes as revealed through MCScanX analysis though both tandem and segmental duplication events were reported in *OsCesA* family (Wang et al., 2010). Being a



member of glycosyltransferase 2 (GT2) family, Cesa proteins have the conserved “DXD, D, QXXRW” motif (Somerville et al., 2004) and conforming to this, all the *SiCesa* proteins except *SiCesa8* have a “DCD, D, QVLRW” consensus sequence, whereas *SiCesa8* had a unique “DYD, D” sequence and the motif “QXXRW” was absent. Similar sequence variations have also been reported by Wang et al. (2010) in rice. Studies on *Cesa* gene family in crop plants have revealed the presence of a large family of cellulose synthase-like (*Csl*) genes with sequence similarity to *Cesa* (Richmond and Somerville, 2000), and these genes are shown to be involved in biosynthesis of hemicelluloses (Yin et al., 2009). Similar to *Cesa*, *Csl* proteins also belong to GT2 family and possess the conserved “DXD, D, QXXRW” motif (Somerville et al., 2004). In foxtail millet, 36 *Csl* genes were identified and categorized as *CslA*, *CslC*, *CslD*, *CslE*, *CslF*, *CslH*, and *CslJ* in accordance to the classification followed by Wang et al. (2010) in rice. Interestingly, 2 *CslJ* genes were identified in foxtail millet, which were reported to be specific to cereals though they are not present in rice and *Brachypodium* (Fincher, 2009). Domain analysis has shown the presence of GT2 domains in all *SiCslA* and *SiCslC* proteins, whereas other *SiCsl* possess CS domain. Similar reports in *Arabidopsis* and rice have shown the presence of characteristic GT2 domain in *CslA* and *CslC* proteins (Yin et al., 2009; Wang et al., 2010). Studies have shown that *CslA* and *CslC* subgroups are the most divergent proteins, which have evolved through duplication and divergence from a common ancestral gene (Yin et al., 2009; Del Bem and Vincentz, 2010), and therefore share similar structural and physicochemical properties (Youngs et al., 2007). Nevertheless, membrane topology and enzymatic function of these proteins are contrastingly different (Davis et al., 2010; Liepman and Cavalier, 2012). In addition, predominant *SiCslD* family proteins have an additional RING/Ubox like zinc-binding domain, which contains a C3HC4 motif capable of binding to zinc cations.

Molecular processes and biological functions of *Csl* genes have been less explored when compared to *Cesa* genes though *Csl* proteins are equally important in cell structuring. Numerous reports have supported the involvement of *CslA* protein in the synthesis of 1,4- $\beta$ -mannan and glucomannan backbones (Dhugga et al., 2004; Liepman et al., 2005; Suzuki et al., 2006; Goubet et al., 2009; Gille et al., 2011) and heterologous expression of *CslA* genes has shown the activity of single enzyme in integrating mannose and glucose into glcomannan chains (Suzuki et al., 2006; Liepman et al., 2007; Gille et al., 2011). Similarly, *CslC* genes encode for xyloglucan glucan synthase, which are involved in xyloglucan biosynthesis (Cocuron et al., 2007). Heterologous expression of *AtCslC4* in *Pichia pastoris* produced soluble 1,4- $\beta$ -glucans with a low degree of polymerization, whereas expression of *AtCslC4* along with *AtXXT1* (xyloglucan xylosyltransferase) produced insoluble 1,4- $\beta$ -glucans with a higher degree of polymerization suggesting the cooperative action of both the enzymes in xyloglucan biosynthesis (Liepman and Cavalier, 2012). Though *CslD* proteins were speculated to be involved in xylan and homogalacturonan synthesis (Hamann et al., 2004; Bernal et al., 2008a,b; Li et al., 2009), *Arabidopsis cslD* mutants have been shown to possess severe phenotypic defects including deformed

root hairs (*cslD2*; Bernal et al., 2008b), root hairs burst (*cslD3*; Bernal et al., 2008b), defective growth of pollen tube (*cslD1* and *cslD4*; Bernal et al., 2008b; Wang et al., 2011) and reduced plant growth (*cslD5*; Bernal et al., 2008a). These reports suggest the role of *CslD* in normal growth and development of plants beyond their function in xylan and homogalacturonan synthesis. The present study identified 4 *SiCslE* genes, whose characterization has not been performed yet in any crop species. One *CslE* gene in *Arabidopsis* and two in rice were reported to date. *CslF* family of genes were considered to be present among grass species and they regulate the synthesis of mixed-linkage glucan ( $\beta$ -1,3; 1,4, glucan) (Hazen et al., 2002; Burton et al., 2006). Mutation of barley *CslF6* gene resulted in reduction of (1,3;1,4)- $\beta$ -D-Glucan and had an impact on chemical composition of barley grains (Hu et al., 2014), whereas overexpression of this gene in *Nicotiana benthamiana* led to accumulation of (1,3;1,4)- $\beta$ -D-Glucan (Wong et al., 2015). Recently, Jin et al. (2015) has demonstrated the role of *OsCslF6* in affecting phosphate accumulation altering the level of carbon metabolism in rice. Similar to *CslF*, *CslH* and *CslJ* are also grass-specific gene family involved in deposition of (1,3;1,4)- $\beta$ -D-Glucan (Doblin et al., 2009; Yin et al., 2009, 2014). In the present study, two genes each belonging to *CslH* and *CslJ* family were identified.

Similar to *Cesa/Csl*, glucan synthase-like protein (*Gsl*) family are also involved in polysaccharide biosynthesis, particularly in synthesis 1,3- $\beta$ -D-glucan callose (Li et al., 1999). Calloses are deposited in developing cell walls of fiber cells, seed hairs and plasmodesmatal canals. Moreover, deposition of callose is also reported in response to pathogen invasion (Muthamilarasan and Prasad, 2013) and abiotic stress including desiccation, wounding and metal toxicity (Stone and Clarke, 1992). In spite of the importance of *Gsl* genes, limited studies have been performed on elucidating the molecular role of these genes and their respective proteins. In *Arabidopsis*, 12 *Gsl* genes have been identified (<https://cellwall.genomics.purdue.edu/intro/index.html>) and mutating *AtGSL5* has been found to confer resistance to powdery mildew infection (Nishimura et al., 2003). A similar report by Jacobs et al. (2003) has also shown that silencing of *AtGsl5* enhances the resistance of silenced lines to *Sphaerotheca fusca*, *Golovinomyces orontii*, and *Blumeria graminis*. In contrast to the role of callose in acting as a physical barrier to prevent pathogen invasion, the reports by Nishimura et al. (2003) and Jacobs et al. (2003) have demonstrated the resistance of *Arabidopsis* to pathogens in the absence of callose. These reports have proved the importance to study the molecular and physiological roles of *Gsl* proteins in response to biotic as well as abiotic stress, and the present investigation has identified 12 *SiGsl* genes which could serve as interesting candidates for functional characterization as foxtail millet is tolerant to environmental stresses.

In the case of monolignol biosynthesis, ten key enzymes namely PAL, C4H, 4CL, HCT, C3H, CCoAOMT, F5H, COMT, CCR, and CAD have been identified and characterized in the present study. Through systematic analysis, 10, 3, 17, 2, 2, 5, 2, 3, 26, and 13 proteins belonging to PAL, C4H, 4CL, HCT, C3H, CCoAOMT, F5H, COMT, CCR, and CAD families,



respectively were identified (**Supplementary Table S6**). These numbers compared with the genes reported in *Arabidopsis*, poplar and eucalyptus has shown that foxtail millet has higher number of *PAL* genes (10) whereas other three organisms have 4, 5, and 9 genes, respectively (Raes et al., 2003; Shi et al., 2010; Carocha et al., 2015). Both foxtail millet and poplar have 2 *C4H* and 17 *4CL* genes, whereas *Arabidopsis* and eucalyptus have lesser number of *C4H* and *4CL* genes. Of note, foxtail millet has a maximum of 26 *CCR* genes, while *Arabidopsis* has 7 and eucalyptus has 2 genes (Raes et al., 2003; Shi et al., 2010; Carocha et al., 2015). The identified monolignol biosynthesis genes were distributed in all the nine chromosomes of foxtail millet, of which two gene-pairs each of *SiPAL* (*SiPAL4-SiPAL5*; *SiPAL8-SiPAL9*) and *SiCAD* (*SiCAD2-SiCAD3*; *SiCAD8-SiCAD9*), and one pair each of *Si4CL* (*Si4CL11-Si4CL12*) and *SiCCoAOMT* (*SiCCoAOMT3-SiCCoAOMT4*) were identified to be tandemly duplicated (**Figure 4**). Phylogenetic analysis of foxtail millet monolignol biosynthesis proteins with *bona fide* proteins of eucalyptus, tobacco, poplar, *Arabidopsis*, maize, medicago and grape revealed that predominant proteins of foxtail millet are highly divergent (**Figure 3**).

Furthermore, promoter analysis has been performed for foxtail millet lignocellulose biosynthesis genes, which revealed the presence of diverse *cis*-regulatory elements that fall under the following categories; (i) *cis*-elements which are universally present in all the gene family members, (ii) *cis*-elements which are present in all the gene family members except one gene, and (iii) *cis*-element which is unique to any one gene of its corresponding gene family (**Supplementary Tables S8, S9**). These data suggest the transcriptional control of cell wall genes by the action of network of transcription factors. This would assist in understanding gene regulatory mechanism controlling the expression of lignocellulose genes and fine tuning them to achieve the optimal pattern of secondary cell-wall deposition. Since gene expression is also regulated at post-transcriptional level through miRNAs, the present study also identified foxtail millet miRNAs which target the transcripts of lignocellulose biosynthesis genes (**Supplementary Table S10**). Moreover, different kinds of molecular markers including SSRs, eSSRs, and ILPs present in both upstream and genic region of lignocellulose biosynthesis genes have been identified (**Supplementary Table S11**), which could be useful for conducting genomics-assisted breeding for biofuel traits in foxtail millet. *In silico* expression profiles of all the lignocellulose biosynthesis genes in four tissues as well as dehydration library revealed the differential expression of these genes in these tissues and during stress, thus signifying their putative involvement in biological functions other than cell wall structuring. This is supported by the reports on mutants of studied genes in *Arabidopsis* and other plants in which severe phenotypic defects have been observed.

In addition to being potential targets for biofuel traits, the lignocellulose biosynthesis genes have also been reported to play vital role in abiotic stress responses. Chen et al. (2005) have shown that *Arabidopsis CesA8* mutants accumulate increased levels of ABA, proline and sugars, and express higher levels of stress-related genes, and thus possess enhanced tolerance

to drought and osmotic stress. Considering this, Guerriero et al. (2014) analyzed the expression of nine putative *CesA* genes in response to cold, heat and salt stress in *Medicago sativa* and identified a salt/heat-induced and a cold/heat-repressed group of genes, which suggest the putative involvement of cellulose synthases in conferring abiotic stress tolerance. Similar to *CesA* genes, *Csl* genes have also been shown to participate in stress responsive machinery. Characterization of the *salt-overly sensitive6* allele of *AtCslD5* has demonstrated reactive oxygen species-based signaling mechanism in response to osmotic stress in *Arabidopsis* (Zhu et al., 2010). Similarly, accumulation of callose in response to environmental stimuli through overexpression of *Gsl* genes has been extensively studied (Nedukha, 2015). Stass and Horst (2009) have reported the production of abiotic stress-induced callose in all the plants through a highly conserved signaling pathway. Lignification has also been reported to be induced during abiotic stresses (Moura et al., 2010). In view of these, expression profiling of candidate genes in response to dehydration, salinity and cold stress as well as ABA, SA, and MeJA treatments was performed, which showed significant higher expression of *SiGsl2* and *SiGsl12* in all the stress conditions. Few genes including *SiCAD6*, *SiC4H2*, *SiPAL2*, *SiF5H2*, *Si4CL10*, *SiHCT1*, and *SiCCR7* were evidenced to be up-regulated either at early or late or both the phases of stresses. Similarly, differential expression patterns were observed for all the genes during hormone treatments and of note, ABA treatment has no significant impact on the expression of the majority of genes.

Noteworthy, the expression profiles of candidate lignocellulose biosynthesis genes were in correlation with the *cis*-regulatory elements present in the promoter regions of respective genes. The genes which are up-regulated during dehydration and salinity stress including *SiGsl2*, *SiGsl12*, *SiPAL2*, *SiC4H2*, *Si4CL10*, *SiF5H2*, *SiHCT1*, *SiCAD1*, and *SiCAD6* have one or more “response to dehydration stress” *cis*-motifs ABRELATERD1, ACGTATERD1 and MYCATRD22 in their promoter regions (Vandepoele et al., 2009; Yan et al., 2014). Similarly, *SiGsl2*, *SiGsl12* and *Si4CL10* that showed higher expression under cold stress have CACGT motif, which was reported to be responsible for response to cold stress (Vandepoele et al., 2009). In case of hormonal treatments, methyl jasmonate responsive *cis*-element BOXLCOREDPCAL (Yan et al., 2014) was found in the promoter regions of *SiCesA5*, *SiGsl2*, *SiGsl12*, *Si4CL10*, *SiPAL2*, *SiC4H2*, and *SiCCR7*. These genes showed significant up-regulation at either early or late or both the phases of methyl jasmonate treatment. Similarly, ABA-responsive genes such as *SiC4H2*, *SiCCR7*, *SiGsl2*, and *SiCesA9* have both MYCCONSensusAT and MYCATRD22 *cis*-motifs, which have been reported to be MYC recognition site in the promoter of dehydration responsive *rd22* gene which in turn was ABA-dependent (Yan et al., 2014), suggesting that these genes were activated in response to ABA. Thus the present study demonstrates that the interaction of *cis*-elements and transcription factors has resulted in differential gene expression through activation or repression respective genes in response to various environmental stresses and hormone treatments (Lee et al., 2002; Benitez et al., 2013). The findings and potential

correlation between the cis-elements to response to a specific elicitor condition are indirect. It is possible that they are linked, but such primary evidence is not provided here. It is also not known if there were any changes to cell walls in the plants used for expression analyses. Altogether, the present investigation suggests the putative involvement of these genes in strengthening the cell wall for tolerance to abiotic stresses, and they could serve as potential candidates for further functional characterization.

## CONCLUSIONS

The present study has identified the genes belonging to *CesA/CsI*, *GsI*, *PAL*, *C4H*, *4CL*, *HCT*, *C3H*, *CCoAOMT*, *F5H*, *COMT*, *CCR*, and *CAD* superfamilies in foxtail millet and the genes were mapped onto nine chromosomes. *In silico* analyses of putative protein properties and gene structures revealed diverse characteristic features of these proteins and their gene duplication analysis showed that few gene family members underwent tandem duplication. Phylogenetic analysis of respective proteins demonstrated that except *CesA/CsI* and *GsI* superfamily, the monolignol biosynthesis proteins are highly diverse. Promoter analysis showed the presence of various unique and common *cis*-regulatory elements in the upstream of lignocellulose biosynthesis genes and potential miRNAs of foxtail millet were identified to target few genes for post-transcriptional gene silencing. In addition, three types of molecular markers were found in lignocellulose biosynthesis genes, which could be used in genomics-assisted breeding. Comparative genome mapping of foxtail millet lignocellulose biosynthesis genes with the sequenced *C<sub>4</sub>* panicoid genomes revealed higher homology with switchgrass, followed by sorghum and maize. Evolutionary analysis showed that both paralogous and homologous gene-pairs underwent intense positive purifying selection, and duplication occurred ~25 mya, whereas divergence of foxtail millet and switchgrass occurred ~4 mya. Similarly, divergence of foxtail millet from sorghum and maize was predicted to occur ~27 mya. *In silico* expression analysis of all the identified genes in four tissues and dehydration stress library of foxtail millet revealed their differential expression pattern, and also suggested the putative biological function of these genes in processes other than cell wall biosynthesis. Expression profiling of candidate genes in response to dehydration, salinity and cold stress along with ABA, SA and MeJA treatments supported the differential expression of these genes with significant higher expression of *SiGsI12*, *SiHCT1*, and *SiCAD6* genes. The results suggested that these genes could be used as potential candidates for functional characterization for biofuel traits. Though similar studies have already been completed in switchgrass, sorghum and maize, the present study conducted in biofuel model foxtail millet would facilitate improving the crop for efficient biofuel production.

## AUTHOR CONTRIBUTIONS

MP conceived and designed the experiments. MM, YK, JJ, SS, CL performed the experiments. MM, CL, MP analyzed the results.

MM, MP wrote the manuscript. MP approved the final version of the manuscript.

## FUNDING

Research on foxtail millet genomics at MP's laboratory is funded by the Core Grant of National Institute of Plant Genome Research, New Delhi, India.

## ACKNOWLEDGMENTS

MM acknowledges University Grants Commission, New Delhi, India for providing Research Fellowship. The authors thank Mr. Rohit Khandelwal for critically reading the manuscript.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2015.00965>

**Supplementary Figure S1 | Multiple sequence alignment of *SiCesA* proteins.**

**Supplementary Figure S2 | Multiple sequence alignment of *SiCsI* proteins.**

**Supplementary Figure S3 | Multiple sequence alignment of *SiGsI* proteins.**

**Supplementary Figure S4 | Prediction of transmembrane domains in the *SiGsI* proteins.** Red line represents transmembrane, blue line represents inside and pink line represents outside orientation.

**Supplementary Figure S5 | Multiple sequence alignment of monolignol biosynthesis proteins.**

**Supplementary Figure S6 | Gene structure of *SiCesA* genes.**

**Supplementary Figure S7 | Gene structure of *SiCsI* genes.**

**Supplementary Figure S8 | Gene structure of *SiGsI* genes.**

**Supplementary Figure S9 | Gene structure of monolignol biosynthesis genes.**

**Supplementary Table S1 | Details of primers used in qRT-PCR analysis.**

**Supplementary Table S2 | Details of *SiCesA/CsI* and *SiGsI* superfamily genes of foxtail millet.**

**Supplementary Table S3 | Details of various domains present in *SiCesA* proteins.**

**Supplementary Table S4 | Details of various domains present in *SiCsI* proteins.**

**Supplementary Table S5 | Details of various domains present in *SiGsI* proteins.**

**Supplementary Table S6 | Details of monolignol biosynthesis pathway genes of foxtail millet.**

**Supplementary Table S7 | Details of various domains present in monolignol biosynthesis pathway proteins.**

**Supplementary Table S8 | Summary of cis-regulatory elements present in *SiCesA/CsI* and *SiGsI* superfamily genes.**

**Supplementary Table S9 | Summary of cis-regulatory elements present in monolignol biosynthesis pathway genes.**

**Supplementary Table S10 | Details of foxtail millet miRNAs identified to target the transcripts of lignocellulose pathway genes.**

**Supplementary Table S11 | Summary of molecular markers present in lignocellulose pathway genes.**

**Supplementary Table S12 | The Ka/Ks ratios and estimated divergence time for homologous lignocellulose pathway proteins between *Setaria italica* and *Panicum virgatum*.**

**Supplementary Table S13 | The Ka/Ks ratios and estimated divergence time for homologous lignocellulose pathway proteins between *Setaria italica* and *Sorghum bicolor*.**

**Supplementary Table S14 | The Ka/Ks ratios and estimated divergence time for homologous lignocellulose pathway proteins between *Setaria italica* and *Zea mays*.**

**Supplementary Table S15 | The Ka/Ks ratios and estimated divergence time for tandemly duplicated lignocellulose pathway proteins.**

## REFERENCES

- Appenzeller, L., Doblin, M., Barreiro, R., Wang, H. Y., Niu, X. M., Kollipara, K., et al. (2004). Cellulose synthesis in maize: isolation and expression analysis of the cellulose synthase (CesA) gene family. *Cellulose* 11, 287–299. doi: 10.1023/B:CELL.0000046417.84715.27
- Benitez, L. C., da Maia, L. C., Ribiero, M. V., Pegoraro, C., Peters, J. A., de Oliveira, A. C., et al. (2013). Salt induced change of gene expression in salt sensitive and tolerant rice species. *J. Agri. Sci.* 5, 251–260. doi: 10.5539/jas.v5n10p251
- Bennetzen, J. L., Schmutz, J., Wang, H., Percifield, R., Hawkins, J., Pontaroli, A. C., et al. (2012). Reference genome sequence of the model plant *Setaria*. *Nat. Biotechnol.* 30, 555–561. doi: 10.1038/nbt.2196
- Bernal, A. J., Jensen, J. K., Harholt, J., Sørensen, S., Møller, I., Blaukopf, C., et al. (2008a). Disruption of ATCSLD5 results in reduced growth, reduced xylan and homogalacturonan synthase activity and altered xylan occurrence in *Arabidopsis*. *Plant J.* 52, 791–802. doi: 10.1111/j.1365-3113X.2007.03281.x
- Bernal, A. J., Yoo, C. M., Mutwil, M., Jensen, J. K., Hou, G., Blaukopf, C., et al. (2008b). Functional analysis of the cellulose synthase-like genes CSLD1, CSLD2, and CSLD4 in tip-growing *Arabidopsis* cells. *Plant Physiol.* 148, 1238–1253. doi: 10.1104/pp.108.121939
- Bonawit, N. D., and Chapple, C. (2010). The genetics of lignin biosynthesis: connecting genotype to phenotype. *Annu. Rev. Genet.* 44, 337–363. doi: 10.1146/annurev-genet-102209-163508
- Burton, R. A., Shirley, N. J., King, B. J., Harvey, A. J., and Fincher, G. B. (2004). The CesA gene family of barley. Quantitative analysis of transcripts reveals two groups of co-expressed genes. *Plant Physiol.* 134, 224–236. doi: 10.1104/pp.103.032904
- Burton, R. A., Wilson, S. M., Hrmova, M., Harvey, A. J., Shirley, N. J., Medhurst, A., et al. (2006). Cellulose synthase-like CslF genes mediate the synthesis of cell wall (1, 3;1, 4)-beta-D-glucans. *Science* 311, 1940–1942. doi: 10.1126/science.1122975
- Byrt, C. S., Grof, C. P. L., and Furbank, R. T. (2011). C<sub>4</sub> plants as biofuel feedstocks: optimising biomass production and feedstock quality from a lignocellulosic perspective. *J. Integ. Plant Biol.* 53, 120–135. doi: 10.1111/j.1744-7909.2010.01023.x
- Cannon, S. B., Mitra, A., Baumgarten, A., Young, N. D., and May, G. (2004). The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biol.* 4:10. doi: 10.1186/1471-2229-4-10
- Carocha, V., Soler, M., Hefer, C., Cassan-Wang, H., Fevèreiro, P., Myburg, A. A., et al. (2015). Genome-wide analysis of the lignin toolbox of *Eucalyptus grandis*. *New Phytol.* 206, 1297–1313. doi: 10.1111/nph.13313
- Chantreau, M., Chabbert, B., Billiard, S., Hawkins, S., and Neutelings, G. (2015). Functional analyses of cellulose synthase genes in flax (*Linum usitatissimum*) by virus-induced gene silencing. *Plant Biotechnol. J.* doi: 10.1111/pbi.12350. [Epub ahead of print].
- Chen, Z., Hong, X., Zhang, H., Wang, Y., Li, X., Zhu, J. K., et al. (2005). Disruption of the cellulose synthase gene, AtCesA8/IRX1, enhances drought and osmotic stress tolerance in *Arabidopsis*. *Plant J.* 43, 273–283. doi: 10.1111/j.1365-3113X.2005.02452.x
- Cocuron, J. C., Lerouxel, O., Drakakaki, G., Alonso, A. P., Liepman, A. H., Keegstra, K., et al. (2007). A gene from the cellulose synthase-like C family encodes a  $\beta$ -1, 4 glucan synthase. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8550–8555. doi: 10.1073/pnas.0703133104
- Cui, X. J., Shin, H. S., Song, C., Laosinchai, W., Amano, Y., and Brown, R. M. (2001). A putative plant homolog of the yeast  $\beta$ -1, 3-glucan synthase subunit FKS1 from cotton (*Gossypium hirsutum* L.) fibers. *Planta* 213, 223–230. doi: 10.1007/s004250000496
- Dai, X., and Zhao, P. X. (2011). psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res.* 39, W155–W159. doi: 10.1093/nar/gkr319
- Davis, J., Brandizzi, F., Liepman, A. H., and Keegstra, K. (2010). *Arabidopsis* mannan synthase CSLA9 and glucan synthase CSLC4 have opposite orientations in the Golgi membrane. *Plant J.* 64, 1028–1037. doi: 10.1111/j.1365-3113X.2010.04392.x
- Del Bem, L. E., and Vincentz, M. G. (2010). Evolution of xyloglucan-related genes in green plants. *BMC Evol. Biol.* 10:341. doi: 10.1186/1471-2148-10-341
- Dhugga, K. S., Barreiro, R., Whitten, B., Stecca, K., Hazebrook, J., Randhawa, G. S., et al. (2004). Guar seed beta-mannan synthase is a member of the cellulose synthase super gene family. *Science* 303, 363–366. doi: 10.1126/science.1090908
- Diao, X., Schnable, J., Bennetzen, J. L., and Li, J. (2014). Initiation of *Setaria* as a model plant. *Front. Agri. Sci. Eng.* 1, 16–20. doi: 10.15302/J-FASE-2014011
- Djerbi, S., Lindskog, M., Arvestad, L., Sterky, F., and Teeri, T. T. (2005). The genome sequence of black cottonwood (*Populus trichocarpa*) reveals 18 conserved cellulose synthase (CesA) genes. *Planta* 221, 739–746. doi: 10.1007/s00425-005-1498-4
- Doblin, M. S., Pettolino, F. A., Wilson, S. M., Campbell, R., Burton, R. A., Fincher, G. B., et al. (2009). A barley cellulose synthase-like CSLH gene mediates (1, 3;1, 4)-beta-D-glucan synthesis in transgenic *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 106, 5996–6001. doi: 10.1073/pnas.0902019106
- Doblin, M. S., Pettolino, F., and Bacic, A. (2010). Plant cell walls: the skeleton of the plant world. *Funct. Plant Biol.* 37, 357–381. doi: 10.1071/FP09279
- Farrokhi, N., Burton, R. A., Brownfield, L., Hrmova, M., Wilson, S. M., Bacic, A., et al. (2006). Plant cell wall biosynthesis: genetic, biochemical and functional genomics approaches to the identification of key genes. *Plant Biotech. J.* 4, 145–167. doi: 10.1111/j.1467-7652.2005.00169.x
- Fincher, G. B. (2009). Revolutionary times in our understanding of cell wall biosynthesis and remodeling in the grasses. *Plant Physiol.* 149, 27–37. doi: 10.1104/pp.108.130096
- Gardiner, J. C., Taylor, N. G., and Turner, S. R. (2003). Control of cellulose synthase complex localization in developing xylem. *Plant Cell* 15, 1740–1748. doi: 10.1105/tpc.012815
- Gille, S., Cheng, K., Skinner, M. E., Liepman, A. H., Wilkerson, C. G., and Pauly, M. (2011). Deep sequencing of voodoo lily (*Amorphophallus konjac*): an approach to identify relevant genes involved in the synthesis of the hemicellulose glucomannan. *Planta* 234, 515–526. doi: 10.1007/s00425-011-1422-z
- Goubet, F., Barton, C. J., Mortimer, J. C., Yu, X., Zhang, Z., Miles, G. P., et al. (2009). Cell wall glucomannan in *Arabidopsis* is synthesised by CSLA glycosyltransferases, and influences the progression of embryogenesis. *Plant J.* 60, 527–538. doi: 10.1111/j.1365-3113X.2009.03977.x
- Griffiths, J. S., Šola, K., Kushwaha, R., Lam, P., and Tateno, M., Young, R., et al. (2015). Unidirectional movement of cellulose synthase complexes in *Arabidopsis* seed coat epidermal cells deposit cellulose involved in mucilage extrusion, adherence, and ray formation. *Plant Physiol.* 168, 502–520. doi: 10.1104/pp.15.00478
- Guerriero, G., Legay, S., and Hausman, J.-F. (2014). Alfalfa cellulose synthase gene expression under abiotic stress: a hitchhiker's guide to RT-qPCR normalization. *PLoS ONE* 9:e103808. doi: 10.1371/journal.pone.0103808
- Hamann, T., Osborne, E., Youngs, H., Misson, J., Nussaume, L., and Somerville, C. (2004). Global expression analysis of CESA and CSL genes in *Arabidopsis*. *Cellulose* 11, 279–286. doi: 10.1023/B:CELL.0000046340.99925.57
- Harakava, R. (2005). Genes encoding enzymes of the lignin biosynthesis pathway in *Eucalyptus*. *Genet. Mol. Biol.* 28, 601–607. doi: 10.1590/s1415-47572005000400015
- Hazen, S. P., Scott-Craig, J. S., and Walton, J. D. (2002). Cellulose synthase-like genes of rice. *Plant Physiol.* 128, 336–340. doi: 10.1104/pp.010875



- Hu, G., Burton, C., Hong, Z., and Jackson, E. (2014). A mutation of the cellulose-synthase-like (*CsLF6*) gene in barley (*Hordeum vulgare* L.) partially affects the  $\beta$ -glucan content in grains. *J. Cereal Sci.* 59, 189–195. doi: 10.1016/j.jcs.2013.12.009
- Humphreys, J. M., Hemm, M. R., and Chapple, C. (1999). New routes for lignin biosynthesis defined by biochemical characterization of recombinant ferulate 5-hydroxylase, a multifunctional cytochrome P450-dependent monooxygenase. *Proc. Natl. Acad. Sci. U.S.A.* 96, 10045–10050. doi: 10.1073/pnas.96.18.10045
- Jacobs, A. K., Lipka, V., Burton, R. A., Panstruga, R., Strizhov, N., Schulze-Lefert, P., et al. (2003). An Arabidopsis callose synthase, GSL5, is required for wound and papillary callose formation. *Plant Cell* 15, 2503–2513. doi: 10.1105/tpc.016097
- Jin, C., Fang, C., Yuan, H., Wang, S., Wu, Y., Liu, X., et al. (2015). Interaction between carbon metabolism and phosphate accumulation is revealed by a mutation of a cellulose synthase-like protein, CSLF6. *J. Exp. Bot.* 66, 2557–2567. doi: 10.1093/jxb/erv050
- Khan, Y., Yadav, A., Suresh, B. V., Muthamilarasan, M., Yadav, C. B., and Prasad, M. (2014). Comprehensive genome-wide identification and expression profiling of foxtail millet [*Setaria italica* (L.)] miRNAs in response to abiotic stress and development of miRNA database. *Plant Cell Tiss. Organ Cult.* 118, 279–292. doi: 10.1007/s11240-014-0480-x
- Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42, D68–D73. doi: 10.1093/nar/gkt1181
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109
- Kujur, A., Bajaj, D., Upadhyaya, H. D., Das, S., Ranjan, R., Shree, T., et al. (2015). A genome-wide SNP scan accelerates trait-regulatory genomic loci identification in chickpea. *Sci. Rep.* 5, 11166. doi: 10.1038/srep11166
- Kumar, K., Muthamilarasan, M., Bonthala, V. S., Roy, R., and Prasad, M. (2015). Unraveling 14-3-3 proteins in *C4* panicoids with emphasis on model plant *Setaria italica* reveals phosphorylation-dependent subcellular localization of RS splicing factor. *PLoS ONE* 10:e0123236. doi: 10.1371/journal.pone.0123236
- Kumari, K., Muthamilarasan, M., Misra, G., Gupta, S., Subramanian, A., Parida, S. K., et al. (2013). Development of eSSR-markers in *Setaria italica* and their applicability in studying genetic diversity, cross-transferability and comparative mapping in millet and non-millet species. *PLoS ONE* 8:e67742. doi: 10.1371/journal.pone.0067742
- Larsen, K. (2004). Cloning and characterization of a ryegrass (*Lolium perenne*) gene encoding cinnamoyl-CoA reductase (CCR). *Plant Sci.* 166, 569–581. doi: 10.1016/j.plantsci.2003.09.026
- Lata, C., Gupta, S., and Prasad, M. (2013). Foxtail millet: a model crop for genetic and genomic studies in bioenergy grasses. *Crit. Rev. Biotechnol.* 33, 328–343. doi: 10.3109/07388551.2012.716809
- Lata, C., Mishra, A. K., Muthamilarasan, M., Bonthala, V. S., Khan, Y., and Prasad, M. (2014). Genome-wide investigation and expression profiling of AP2/ERF transcription factor superfamily in foxtail millet (*Setaria italica* L.). *PLoS ONE* 9:e113092. doi: 10.1371/journal.pone.0113092
- Lee, T. H., Tang, H., Wang, X., and Paterson, A. H. (2012). PGDD: a database of gene and genome duplication in plants. *Nucleic Acids Res.* 41, D1152–D1158. doi: 10.1093/nar/gks1104
- Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., et al. (2002). Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298, 799–804. doi: 10.1126/science.1075090
- Lerouxel, O., Cavalier, D. M., Liepman, A. H., and Keegstra, K. (2006). Biosynthesis of plant cell wall polysaccharides - a complex process. *Curr. Opin. Plant Biol.* 9, 621–630. doi: 10.1016/j.pbi.2006.09.009
- Li, H. J., Bacic, A., and Read, S. M. (1999). Role of a callose synthase zymogen in regulating wall deposition in pollen tubes of *Nicotiana glauca* Link et Otto. *Planta* 208, 528–538. doi: 10.1007/s004250050590
- Li, M., Xiong, G. Y., Cui, J. J., Tang, D., Zhang, B. C., Pauly, M., et al. (2009). Rice cellulose synthase-like D4 is essential for normal cell-wall biosynthesis and plant growth. *Plant J.* 60, 1055–1069. doi: 10.1111/j.1365-313X.2009.04022.x
- Li, P., and Brutnell, T. P. (2011). *Setaria viridis* and *Setaria italica*, model genetic systems for the Panicoid grasses. *J. Exp. Bot.* 62, 3031–3037. doi: 10.1093/jxb/err096
- Liepman, A. H., and Cavalier, D. M. (2012). The CELLULOSE SYNTHASE-LIKE A and CELLULOSE SYNTHASE-LIKE C families: recent advances and future perspectives. *Front. Plant Sci.* 3:109. doi: 10.3389/fpls.2012.00109
- Liepman, A. H., Nairn, C. J., Willats, W. G., Sørensen, I., Roberts, A. W., and Keegstra, K. (2007). Functional genomic analysis supports conservation of function among CELLULOSE SYNTHASE-LIKE A gene family members and suggests diverse roles of mannans in plants. *Plant Physiol.* 143, 1881–1893. doi: 10.1104/pp.106.093989
- Liepman, A. H., Wilkerson, C. G., and Keegstra, K. (2005). Expression of cellulose synthase-like (Csl) genes in insect cells reveals that CslA family members encode mannan synthases. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2221–2226. doi: 10.1073/pnas.0409179102
- McFarlane, H. E., Döring, A., and Persson, S. (2014). The cell biology of cellulose synthesis. *Annu. Rev. Plant Biol.* 65, 69–94. doi: 10.1146/annurev-arplant-050213-040240
- Mishra, A. K., Muthamilarasan, M., Khan, Y., Parida, S. K., and Prasad, M. (2013). Genome-wide investigation and expression analyses of WD40 protein family in the model plant foxtail millet (*Setaria italica* L.). *PLoS ONE* 9:e86852. doi: 10.1371/journal.pone.0086852
- Moura, J. C. M. S., Bonine, C. A. V., De Oliveira Fernandes Viana, J., Dornelas, M. C., and Mazzafera, P. (2010). Abiotic and biotic stresses and changes in the lignin content and composition in plants. *J. Integr. Plant Biol.* 52, 360–376. doi: 10.1111/j.1744-7909.2010.00892.x
- Muthamilarasan, M., Bonthala, V. S., Mishra, A. K., Khandelwal, R., Khan, Y., Roy, R., et al. (2014b). C<sub>2</sub>H<sub>2</sub>-type of zinc finger transcription factors in foxtail millet define response to abiotic stresses. *Funct. Integr. Genomics* 14, 531–543. doi: 10.1007/s10142-014-0383-2
- Muthamilarasan, M., Dhaka, A., Yadav, R., and Prasad, M. (2015). Exploration of millet models for developing nutrient rich graminaceous crops. *Plant Sci.* doi: 10.1016/j.plantsci.2015.08.023. [Epub ahead of print].
- Muthamilarasan, M., Khandelwal, R., Yadav, C. B., Bonthala, V. S., Khan, Y., and Prasad, M. (2014a). Identification and molecular characterization of MYB Transcription Factor Superfamily in *C4* model plant foxtail millet (*Setaria italica* L.). *PLoS ONE* 9:e109920. doi: 10.1371/journal.pone.0109920
- Muthamilarasan, M., and Prasad, M. (2013). Plant innate immunity: an updated insight into defense mechanism. *J. Biosci.* 38, 433–449. doi: 10.1007/s12038-013-9302-2
- Muthamilarasan, M., and Prasad, M. (2015). Advances in *Setaria* genomics for genetic improvement of cereals and bioenergy grasses. *Theor. Appl. Genet.* 128, 1–14. doi: 10.1007/s00122-014-2399-3
- Muthamilarasan, M., Venkata Suresh, B., Pandey, G., Kumari, K., Parida, S. K., and Prasad, M. (2014c). Development of 5123 intron-length polymorphic markers for large-scale genotyping applications in foxtail millet. *DNA Res.* 21, 41–52. doi: 10.1093/dnares/dst039
- Nedukha, O. M. (2015). Callose: localization, functions, and synthesis in plant cells. *Cytol. Genet.* 49, 49–57. doi: 10.3103/S0095452715010090
- Nishimura, M. T., Stein, M., Hou, B. H., Vogel, J. P., Edwards, H., and Somerville, S. C. (2003). Loss of a callose synthase results in salicylic acid-dependent disease resistance. *Science* 301, 969–972. doi: 10.1126/science.1086716
- Osakabe, K., Tsao, C. C., Li, L., Popko, J. L., Umezawa, T., Carraway, D. T., et al. (1999). Coniferyl aldehyde 5-hydroxylation and methylation direct syringyl lignin biosynthesis in angiosperms. *Proc. Natl. Acad. Sci. U.S.A.* 96, 8955–8960. doi: 10.1073/pnas.96.16.8955
- Pandey, G., Misra, G., Kumari, K., Gupta, S., Parida, S. K., Chattopadhyay, D., et al. (2013). Genome-wide development and use of microsatellite markers for large-scale genotyping applications in foxtail millet [*Setaria italica* (L.)]. *DNA Res.* 20, 197–207. doi: 10.1093/dnares/dst002
- Parvathi, K., Chen, F., Guo, D., Blount, J. W., and Dixon, R. A. (2001). Substrate preferences of O-methyltransferases in alfalfa suggest new pathways for 3-O-methylation of monolignols. *Plant J.* 25, 193–202. doi: 10.1046/j.1365-313X.2001.00956.x
- Pear, J. R., Kawagoe, Y., Schreckengost, W. E., Delmer, D. P., and Stalker, D. M. (1996). Higher plants contain homologs of the bacterial celA genes encoding the catalytic subunit of cellulose synthase. *Proc. Natl. Acad. Sci. U.S.A.* 93, 12637–12642. doi: 10.1073/pnas.93.22.12637
- Persson, S., Paredes, A., Carroll, A., Palsdottir, H., Doblin, M., Poindexter, P., et al. (2007). Genetic evidence for three unique components in primary cell-wall



- cellulose synthase complexes in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15566–15571. doi: 10.1073/pnas.0706592104
- Petti, C., Shearer, A., Tatenno, M., Ruwaya, M., Nokes, S., Brutnell, T., et al. (2013). Comparative feedstock analysis in *Setaria viridis* L. as a model for C<sub>4</sub> bioenergy grasses and panicle crop species. *Front. Plant Sci.* 4:181. doi: 10.3389/fpls.2013.00181
- Pichon, M., Courbou, I., Beckert, M., Boudet, A. M., and Grima-Pettenati, J. (1998). Cloning and characterization of two maize cDNAs encoding Cinnamoyl-CoA Reductase (CCR) and differential expression of the corresponding genes. *Plant Mol. Biol.* 38, 671–676. doi: 10.1023/A:1006060101866
- Puranik, S., Sahu, P. P., Mandal, S. N., Venkata Suresh, B., Parida, S. K., and Prasad, M. (2013). Comprehensive genome-wide survey, genomic constitution and expression profiling of the NAC transcription factor family in foxtail millet (*Setaria italica* L.). *PLoS ONE* 8:e64594. doi: 10.1371/journal.pone.0064594
- Qi, X., Xie, S., Liu, Y., Yi, F., and Yu, J. (2013). Genome-wide annotation of genes and noncoding RNAs of foxtail millet in response to simulated drought stress by deep sequencing. *Plant Mol. Biol.* 83, 459–473. doi: 10.1007/s11103-013-0104-6
- Raes, J., Rohde, A., Christensen, J. H., Van de Peer, Y., and Boerjan, W. (2003). Genome-wide characterization of the lignification toolbox in *Arabidopsis*. *Plant Physiol.* 133, 1051–1071. doi: 10.1104/pp.103.026484
- Richmond, T. A., and Somerville, C. R. (2000). The cellulose synthase superfamily. *Plant Physiol.* 124, 495–498. doi: 10.1104/pp.124.2.495
- Saathoff, A. J., Hargrove, M. S., Haas, E. J., Tobias, C. M., Twigg, P., Sattler, S., et al. (2012). Switchgrass PviCAD1: understanding Residues Important for Substrate Preferences and Activity. *Appl. Biochem. Biotechnol.* 168, 1086–1100. doi: 10.1007/s12010-012-9843-0
- Saathoff, A. J., Sarath, G., Chow, E. K., Dien, B. S., and Tobias, C. M. (2011b). Downregulation of cinnamyl-alcohol dehydrogenase in switchgrass by RNA silencing results in enhanced glucose release after cellulase treatment. *PLoS ONE* 6:e16416. doi: 10.1371/journal.pone.0016416
- Saathoff, A. J., Tobias, C. M., Sattler, S. E., Haas, E. E., Twigg, P., and Sarath, G. (2011a). Switchgrass contains two cinnamyl alcohol dehydrogenase involved in lignin formation. *Bioener. Res.* 4, 120–133. doi: 10.1007/s12155-010-9106-2
- Saballos, A., Ejeta, G., Sanchez, E., Kang, C., and Vermerris, W. (2009). A genomewide analysis of the cinnamyl alcohol dehydrogenase family in sorghum [*Sorghum bicolor* (L.) Moench] identifies *SbCAD2* as the brown midrib6 gene. *Genetics* 181, 783–795. doi: 10.1534/genetics.108.098996
- Saeed, A. I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., et al. (2003). TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 34, 374–378.
- Sattler, S., Palmer, N., Saballos, A., Greene, A., Xin, Z., Sarath, G., et al. (2012). Identification and characterization of four missense mutations in brown midrib 12 (Bmr12), the caffeic O-methyltransferase (COMT) of sorghum. *Bioenergy Res.* 5, 855–865. doi: 10.1007/s12155-012-9197-z
- Saxena, I. M., and Brown, R. M. (2000). Cellulose synthases and related enzymes. *Curr. Opin. Plant Biol.* 3, 523–531. doi: 10.1016/S1369-5266(00)00125-4
- Schmer, M. R., Vogel, K. P., Mitchel, R. B., and Perrin, R. K. (2008). Net energy of cellulosic ethanol from switchgrass. *Proc. Natl. Acad. Sci. U.S.A.* 105, 464–469. doi: 10.1073/pnas.0704767105
- Shen, H., Mazarei, M., Hisano, H., Escamilla-Trevino, L., Fu, C., Pu, Y., et al. (2013). A genomics approach to deciphering lignin biosynthesis in switchgrass. *Plant Cell* 25, 4342–4361. doi: 10.1105/tpc.113.118828
- Shi, R., Sun, Y. H., Li, Q., Heber, S., Sederoff, R., and Chiang, V. L. (2010). Towards a systems approach for lignin biosynthesis in *Populus trichocarpa*: transcript abundance and specificity of the monolignol biosynthetic genes. *Plant Cell Physiol.* 51, 144–163. doi: 10.1093/pcp/pcp175
- Somerville, C., Bauer, S., Brininstool, G., Facette, M., Hamann, T., Milne, J., et al. (2004). Toward a systems approach to understanding plant cell walls. *Science* 306, 2206–2211. doi: 10.1126/science.1102765
- Stass, A., and Horst, W. J. (2009). “Callose in abiotic stress,” in *Chemistry, Biochemistry and Biology of (1→3)-β-Glucans and Related Polysaccharides*, eds A. Bacic, G. B. Fincher, and B. A. Stone, (New York, NY: Academic), 499–524.
- Stone, B. A., and Clarke, A. E. (1992). *Chemistry and Biology of (1, 3)-β-Glucans*. Victoria: La Trobe University Press.
- Suresh, B. V., Muthamilarasan, M., Misra, G., and Prasad, M. (2013). FmMDb: a versatile database of foxtail millet markers for millets and bioenergy grasses research. *PLoS One* 8:e71418. doi: 10.1371/journal.pone.0071418
- Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34, W609–W612. doi: 10.1093/nar/gkl315
- Suzuki, S., Li, L., Sun, Y. H., and Chiang, V. L. (2006). The cellulose synthase gene superfamily and biochemical functions of xylem-specific cellulose synthase-like genes in *Populus trichocarpa*. *Plant Physiol.* 142, 1233–1245. doi: 10.1104/pp.106.086678
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Taylor, N. G., Howells, R. M., Huttly, A. K., Vickers, K., and Turner, S. R. (2003). Interactions among three distinct CesA proteins essential for cellulose synthesis. *Proc. Natl. Acad. Sci. U.S.A.* 100, 1450–1455. doi: 10.1073/pnas.0337628100
- Trabucco, G. M., Matos, D. A., Lee, S. J., Saathoff, A. J., Priest, H. D., Mockler, T. C., et al. (2013). Functional characterization of cinnamyl alcohol dehydrogenase and caffeic acid O-methyltransferase in *Brachypodium distachyon*. *BMC Biotechnol.* 13:61. doi: 10.1186/1472-6750-13-61
- Vandepoele, K., Quimbaya, M., Casneuf, T., Veylder, D. L., and de Peer, Y. V. (2009). Unraveling transcriptional control in Arabidopsis using cis-regulatory elements and coexpression networks. *Plant Physiol.* 150, 535–546. doi: 10.1104/pp.109.136028
- van der Weijde, T., Alvim Kamei, C. L., Torres, A. F., Vermerris, W., Dolstra, O., Visser, R. G., et al. (2013). The potential of C<sub>4</sub> grasses for cellulosic biofuel production. *Front. Plant Sci.* 4:107. doi: 10.3389/fpls.2013.00107
- Vanholme, R., Storme, V., Vanholme, B., Sundin, L., Christensen, J. H., Goeminne, G., et al. (2012). A systems biology view of responses to lignin biosynthesis perturbations in *Arabidopsis*. *Plant Cell* 24, 3506–3529. doi: 10.1105/tpc.112.102574
- van Parijs, F. R., Rutting, T., Boerjan, W., Haesaert, G., Byrne, S. L., Asp, T., et al. (2015). Clade classification of monolignol biosynthesis gene family members reveals target genes to decrease lignin in *Lolium perenne*. *Plant Biol.* 17, 877–892. doi: 10.1111/plb.12316
- Voorrips, R. E. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Hered.* 93, 77–78. doi: 10.1093/jhered/93.1.77
- Wang, L., Guo, K., Li, Y., Tu, Y., Hu, H., Wang, B., et al. (2010). Expression profiling and integrative analysis of the CESA/CSL superfamily in rice. *BMC Plant Biol.* 10:282. doi: 10.1186/1471-2229-10-282
- Wang, W., Wang, L., Chen, C., Xiong, G., Tan, X. Y., Yang, K. Z., et al. (2011). *Arabidopsis* CSLD1 and CSLD4 are required for cellulose deposition and normal growth of pollen tubes. *J. Exp. Bot.* 62, 5161–5177. doi: 10.1093/jxb/err221
- Wang, Y., Chantreau, M., Sibout, R., and Hawkins, S. (2013). Plant cell wall lignification and monolignol metabolism. *Front. Plant Sci.* 4:220. doi: 10.3389/fpls.2013.00220
- Wang, Y., Tang, H., DeBarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCSanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40, e49. doi: 10.1093/nar/gkr1293
- Warnasooriya, S. N., and Brutnell, T. P. (2014). Enhancing the productivity of grasses under high-density planting by engineering light responses: from model systems to feedstocks. *J. Exp. Bot.* 65, 2825–2834. doi: 10.1093/jxb/eru221
- Wong, S. C., Shirley, N. J., Little, A., Khoo, K. H., Schwerdt, J., Fincher, G. B., et al. (2015). Differential expression of the HvCSLF6 gene late in grain development may explain quantitative differences in (1, 3;1, 4)-β-glucan concentration in barley. *Mol. Breed.* 35, 20. doi: 10.1007/s11032-015-0208-6
- Yadav, C. B., Muthamilarasan, M., Pandey, G., and Prasad, M. (2015). Identification, characterization and expression profiling of Dicer-like, Argonaute and RNA-dependent RNA polymerase gene families in foxtail millet. *Plant Mol. Biol. Rep.* 33, 43–55. doi: 10.1007/s11105-014-0736-y
- Yan, J., Wang, B., Jiang, Y., Cheng, L., and Wu, T. (2014). GmFNSII-controlled soybean flavone metabolism responds to abiotic stresses and regulates plant salt tolerance. *Plant Cell Physiol.* 55, 74–86. doi: 10.1093/pcp/pct159

- Yin, Y., Huang, J., and Xu, Y. (2009). The cellulose synthase superfamily in fully sequenced plants and algae. *BMC Plant Biol.* 9:99. doi: 10.1186/1471-2229-9-99
- Yin, Y., Johns, M. A., Cao, H., and Rupani, M. (2014). A survey of plant and algal genomes and transcriptomes reveals new insights into the evolution and function of the cellulose synthase superfamily. *BMC Genomics* 15:260. doi: 10.1186/1471-2164-15-260
- Youn, B., Camacho, R., Moinuddin, S. G. A., Lee, C., Davin, L. B., Lewis, N. G., et al. (2006). Crystal structures and catalytic mechanism of the Arabidopsis cinnamyl alcohol dehydrogenases AtCAD5 and AtCAD4. *Org. Biomol. Chem.* 4, 1687–1697. doi: 10.1039/b601672c
- Youngs, H. L., Hamann, T., Osborne, E., and Somerville, C. R. (2007). “The cellulose synthase superfamily,” in *Cellulose: Molecular and Structural Biology*, eds M. Brown and I. M. Saxena (Berlin: Springer), 35–49.
- Zhang, G., Liu, X., Quan, Z., Cheng, S., Xu, X., Pan, S., et al. (2012). Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. *Nat. Biotechnol.* 30, 549–554. doi: 10.1038/nbt.2195
- Zhong, R., and Ye, Z. H. (2015). Secondary cell walls: biosynthesis, patterned deposition and transcriptional regulation. *Plant Cell Physiol.* 56, 195–214. doi: 10.1093/pcp/pcu140
- Zhu, J., Lee, B. H., Dellinger, M., Cui, X., Zhang, C., Wu, S., et al. (2010). A cellulose synthase-like protein is required for osmotic stress tolerance in *Arabidopsis*. *Plant J.* 63, 128–140. doi: 10.1111/j.1365-3113.2010.04227.x
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Muthamilarasan, Khan, Jaishankar, Shweta, Lata and Prasad. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Proteomic Responses of Switchgrass and Prairie Cordgrass to Senescence

Bimal Paudel<sup>1†</sup>, Aayudh Das<sup>1†</sup>, Michael Long Tran<sup>1</sup>, Arvid Boe<sup>2</sup>, Nathan A. Palmer<sup>3</sup>, Gautam Sarath<sup>3</sup>, Jose L. Gonzalez-Hernandez<sup>2</sup>, Paul J. Rushton<sup>4</sup> and Jai S. Rohila<sup>1,2\*</sup>

## OPEN ACCESS

### Edited by:

Soren K. Rasmussen,  
University of Copenhagen, Denmark

### Reviewed by:

Xiyin Wang,  
North China University of Science and  
Technology, China  
Abu Hena Mostafa Kamal,  
University of Texas at Arlington, USA  
Angela Mehta,  
Embrapa Recursos Genéticos e  
Biotecnologia, Brazil

### \*Correspondence:

Jai S. Rohila  
jai.rohila@sdstate.edu

### †Present Address:

Aayudh Das,  
Department of Biochemistry and  
Biophysics, Texas A&M University,  
College Station, TX, USA

<sup>†</sup>These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Crop Science and Horticulture,  
a section of the journal  
Frontiers in Plant Science

**Received:** 23 November 2015

**Accepted:** 24 February 2016

**Published:** 14 March 2016

### Citation:

Paudel B, Das A, Tran M, Boe A,  
Palmer NA, Sarath G,  
Gonzalez-Hernandez JL, Rushton PJ  
and Rohila JS (2016) Proteomic  
Responses of Switchgrass and Prairie  
Cordgrass to Senescence.  
Front. Plant Sci. 7:293.  
doi: 10.3389/fpls.2016.00293

<sup>1</sup> Department of Biology and Microbiology, South Dakota State University, Brookings, SD, USA, <sup>2</sup> Department of Plant Science, South Dakota State University, Brookings, SD, USA, <sup>3</sup> Grain, Forage and Bioenergy Research Unit, United States Department of Agriculture - Agricultural Research Service, Lincoln, NE, USA, <sup>4</sup> 22nd Century Group Inc., Clarence, NY, USA

Senescence in biofuel grasses is a critical issue because early senescence decreases potential biomass production by limiting aerial growth and development. 2-Dimensional, differential in-gel electrophoresis (2D-DIGE) followed by mass spectrometry of selected protein spots was used to evaluate differences between leaf proteomes of early (ES)- and late- senescing (LS) genotypes of Prairie cordgrass (ES/LS PCG) and switchgrass (ES/LS SG), just before and after senescence was initiated. Analysis of the manually filtered and statistically evaluated data indicated that 69 proteins were significantly differentially abundant across all comparisons, and a majority (41%) were associated with photosynthetic processes as determined by gene ontology analysis. Ten proteins were found in common between PCG and SG, and nine and 18 proteins were unique to PCG and SG respectively. Five of the 10 differentially abundant spots common to both species were increased in abundance, and five were decreased in abundance. Leaf proteomes of the LS genotypes of both grasses analyzed before senescence contained significantly higher abundances of a 14-3-3 like protein and a glutathione-S-transferase protein when compared to the ES genotypes, suggesting differential cellular metabolism in the LS vs. the ES genotypes. The higher abundance of 14-3-3 like proteins may be one factor that impacts the senescence process in both LS PCG and LS SG. Aconitase dehydratase was found in greater abundance in all four genotypes after the onset of senescence, consistent with literature reports from genetic and transcriptomic studies. A Rab protein of the Ras family of G proteins and an s-adenosylmethionine synthase were more abundant in ES PCG when compared with the LS PCG. In contrast, several proteins associated with photosynthesis and carbon assimilation were detected in greater abundance in LS PCG when compared to ES PCG, suggesting that a loss of these proteins potentially contributed to the ES phenotype in PCG. Overall, this study provides important data that can be utilized toward delaying senescence in both PCG and SG, and sets a foundational base for future improvement of perennial grass germplasm for greater aerial biomass productivity.

**Keywords:** biofuel grasses, cordgrass, switchgrass, proteomics, senescence

## INTRODUCTION

Prairie cordgrass (PCG) and switchgrass (SG) are two warm-season  $C_4$  grasses that are widely adapted to North American climatic conditions and have great potential as feedstock for the lignocellulosic-based biofuels industry (Lee and Boe, 2005; Sarath et al., 2008; Gonzalez-Hernandez et al., 2009). Senescence is a crucial plant development process that enables the remobilization of nutrients within a plant (Noodén et al., 1997). Leaf senescence is marked by the degradation of subcellular compartments, such as chloroplasts, and the remobilization of nutrients to other parts of the plants, such as seeds and underground rhizomes in the case of PCG and SG. The balance between oxidative stress and antioxidant activity plays a crucial role during senescence (Prochazkova et al., 2001). Senescence can be initiated by complex signals of age-specific factors in the genome and by temperature or day length in the case of seasonal senescence, and is often accompanied with increased ROS (reactive oxygen species) and oxidative stress (Jones et al., 2012; Palmer et al., 2015). Although ROS are generated continuously during normal growth, they are balanced by various antioxidant pathways, thus maintaining an optimal cellular redox state for growth. However, during stress related and age specific senescence, these antioxidant pathways cannot overcome oxidative stress, leading to senescence rather than normal growth. Mitochondria are among the major sources of ROS and ROS-related stress signals during programmed cell death (Zhao and Xu, 2000; Fleury et al., 2002).

The onset of early senescence, which is different from aging (Lim et al., 2003), is known to affect the amount of accumulated biomass by a perennial grass (Sarath et al., 2014). During senescence, leaves lose their photosynthetic capability sharply and, as a result, much less carbon is assimilated by the plant. Because leaves are the primary organ in plants that fixes carbon and directly contributes to plant biomass yields, untimely leaf senescence can cause losses in the total potential biomass yields of perennial grasses (Rinerson et al., 2015). Moreover, early senescing plants become more vulnerable to pathogen attacks, especially fungal pathogens, further reducing the biomass of the biofuel crops (Ahonsi et al., 2013). From an evolutionary point of view, senescence evolved to maximize the reutilization of nutrients that were accumulated by leaves during the growing season (Bleecker, 1998). Senescence in plants is a genetically programmed sequence of biochemical and physiological changes, but little is known about it at the protein level in bioenergy crops. Early and recent studies (Gan and Amasino, 1997; Biswal and Biswal, 1999; Palmer et al., 2014, 2015) suggested that senescence is highly correlated with the differential expression of senescence-associated-genes (SAGs). By controlling these signature genes, which regulate the juvenile to adult phase transition in plants, it may be possible to modify or enhance the biomass properties of a wide range of bioenergy feedstock (Chuck et al., 2011). Identifying and using new SAGs is therefore necessary in breeding programs for bioenergy feedstock crops.

A transcriptomic study of switchgrass flag leaf development from elongation through the onset of senescence was performed recently by Palmer et al. (2015). Many candidate genes

were identified that were presumably involved in regulating the expression of senescence-related pathways. The authors found that during the onset of senescence, leaf chlorophyll content decline was associated with a significant upregulation in transcripts coding for enzymes involved in chlorophyll degradation and a large number of SAGs. Moreover, genes such as ureide, ammonium, nitrate, and molybdenum transporters that code for nitrogen and mineral utilization shared expression profiles that were significantly co-regulated with the expression profiles of NAC transcription factors. Similarly, Rinerson et al. (2015) identified 240 *WRKY* genes in the switchgrass genome and studied their expression during flag leaf development. Twenty-eight of these *WRKY* genes were identified as possible targets for increasing biomass yields in SG by delaying senescence. During senescence, protein activation by post-translational modification may be an additional mechanism of regulation. Several studies have shown that mRNA abundance for some genes may not necessarily be a predictor of protein abundance (Wang et al., 2006; Carp and Gepstein, 2007).

Similar to transcriptomics, high-throughput proteomic studies are a powerful tool to analyze changes in protein accumulation levels and post-translational modifications (Liu et al., 2013; Robbins et al., 2013; Wang et al., 2014), but no detailed proteomic study has been conducted on leaf senescence in either switchgrass or cordgrass.

Here, the power of global proteomics was utilized with an aim to profile and identify SG and PCG proteins that were associated with the senescence process using two contrasting genotypes for each species. Differentially abundant proteins in leaves from an early-senescence (ES) genotype were compared with those in a late-senescence (LS) genotype before and after the onset of senescence. Our results provide insights into the molecular basis of the differential responses of the two economically important bioenergy feedstock crops. These results may be instrumental in the rational engineering of senescence in SG and PCG with longer growing periods for increased biomass production.

## MATERIALS AND METHODS

### Plant Materials and Treatments

Leaf tissues were collected in triplicates from field-established clones of two different genotypes of SG [Genotype # 5 (ES SG), Genotype # 4 (LS SG)]. There was a 10-day difference between the two genotypes for date of onset of anthesis (10 August for the ES SG genotype compared 20 August for the LS SG genotype on average at Brookings, SD). These plants were selected from clonally maintained field nurseries established from random seedlings obtained from cultivar Sunburst; similarly, phenotypic selection from a larger PCG population was used to identify the ES PCG and LS PCG genotypes (Boe, unpublished). Similar to the phenological difference observed between the two switchgrass genotypes, there was about a 10-day difference in the onset of anthesis between the two PCG genotypes at the Brookings SD location. Leaves were collected from three clonal replicates of all four genotypes at two different times, before senescence and after senescence, for a total of 24 samples. Samples were assigned to groups A through H based on source population and harvest



time (Table 1). The timing of senescence, for the sake of before and after senescence, was determined by measuring chlorophyll content (Palmer et al., 2015) using a hand-held chlorophyll meter (Supplementary Figure 1). After harvest, all leaf tissues were snap frozen under liquid nitrogen and were stored at  $-80^{\circ}\text{C}$  until further use.

Sample Preparation, 2-D Differential In-Gel Electrophoresis (2D-DIGE), Gel Staining, Image Analysis, and Protein Identification by LC-MS/MS

The fold change in abundance of different proteins was analyzed in the form of ratios for eight different sets of comparisons based on group assignment (Table 1). Four are ratios of LS to ES PCG and SG: (i) the ratio of LS to ES PCG before senescence (G/E), (ii) the ratio of LS to ES PCG after senescence (H/F), (iii) the ratio for LS to ES SG before senescence (C/A), and (iv) the ratio of LS to ES SG after senescence (D/B). Similarly, the other four sets were as follows: (v) the ratio for after to before senescence in ES PCG (F/E), (vi) the ratio for after to before senescence in LS PCG (H/G), (vii) the ratio for after to before senescence in ES SG (B/A), and (viii) the ratio for after to before senescence in LS SG (D/C).

TABLE 1 | Plants, treatments, and groups for sample collection.

Genotype	Collection time	Group	Sample #
Switchgrass clone # 5 (early senescence) ES SG	Before senescence	A	Sample #1
			Sample #2
			Sample #3
	After senescence	B	Sample #4
			Sample #5
			Sample #6
Switchgrass clone # 4 (late senescence) LS SG	Before senescence	C	Sample #7
			Sample #8
			Sample #9
	After senescence	D	Sample #10
			Sample #11
			Sample #12
Prairie cordgrass-ND (early senescence) ES PCG	Before senescence	E	Sample #13
			Sample #14
			Sample #15
	After senescence	F	Sample #16
			Sample #17
			Sample #18
Prairie cordgrass-SD (late senescence) LS PCG	Before senescence	G	Sample #19
			Sample #20
			Sample #21
	After senescence	H	Sample #22
			Sample #23
			Sample #24

Two genotypes of SG and two genotypes of PCG were selected with two treatments for each for sample collection. Samples were collected in triplicate.

Protein Lysate Preparation

One gram of leaf tissue samples was ground to powder under liquid nitrogen using a mortar and pestle. Three hundred microliters of 2-D cell lysis buffer (30 mM Tris-HCl, pH 8.8, containing 7 M urea, 2 M thiourea, and 4% CHAPS) was added to this ground tissue and subjected to sonication on ice (Hurkman and Tanaka, 1986). Next, the tubes were shaken for 30 min on a shaker at room temperature. Then, samples were centrifuged for 30 min at  $4^{\circ}\text{C}$  at  $25,000 \times g$ , and the supernatant was collected. The concentration of protein was measured in the supernatant by Bio-Rad protein assay buffer with BSA as a standard following the standard manufacturer’s guidelines (Peterson, 1983). Lysate samples were diluted with 2-D lysis buffer to a concentration of 5 mg/ml.

Minimal Cy Dye Labeling

The 2D-DIGE were performed by Applied Biomics (Hayward, CA) following the protocol described in Robbins et al. (2013) and Das et al. (in press). Briefly, 1.0  $\mu\text{l}$  of diluted Cy Dye was added to 30  $\mu\text{g}$  of protein lysate (1:5 diluted with DMF from 1 nM/ $\mu\text{l}$  stock), followed by a short vortexing. The tubes were kept under dark on ice for 30 min followed by the addition of 1.0  $\mu\text{l}$  of 10 mM lysine to each of the samples and vortexing; then, the reaction was kept in the dark on ice for additional 15 min. A pooled protein sample was prepared containing equal amounts of all 24 samples and was labeled with Cy2 using the same protocol. The Cy2 labeled sample was used as an internal control to compare gel-to-gel variations. For one gel, 3 samples consisting of Cy2, Cy3, and Cy5 labeled samples were mixed with 2X 2-D sample buffer (8 M urea, 4% CHAPS, 20 mg/ml DTT, 2% pharmalytes, and a trace amount of bromophenol blue). Then, 100  $\mu\text{l}$  of destreak solution and rehydration buffer (7 M urea, 2 M thiourea, 4% CHAPS, 20 mg/ml DTT, 1% pharmalytes, and a trace amount of bromophenol blue) was also added to make a final volume of 350  $\mu\text{l}$  for the 18 cm IPG strip (pH 3-10). This was mixed well and spun before loading the labeled samples into the strip holder.

IEF, SDS-PAGE, Image Scan and Data Analysis

After loading the labeled samples into the IPG strip holder, 18 cm strips were put facedown, and 1.5 ml of mineral oil was added on the top of the strips. This was followed by the protocol provided (Amersham BioSciences) and isoelectric focusing (IEF) was carried out in the dark at  $20^{\circ}\text{C}$ . After IEF, the IPG strips were incubated in freshly made equilibration buffer I (50 mM Tris-HCl, pH 8.8, containing 6 M urea, 30% glycerol, 2% SDS, a trace amount of bromophenol blue, and 10 mg/ml DTT) for 15 min with slow shaking. Then, the strips were rinsed in freshly prepared equilibration buffer II (50 mM Tris-HCl, pH 8.8, containing 6 M urea, 30% glycerol, 2% SDS, a trace amount of bromophenol blue, and 45 mg/ml iodacetamide) for 10 min with slow shaking. The strips were then rinsed in the SDS gel running buffer once, followed by their transfer into the SDS gel (12.5% acrylamide SDS gel prepared using low florescent glass plates). They were then sealed with 0.5% (w/v) agarose solution (in SDS gel running buffer). Running of the SDS gels was carried out at  $15^{\circ}\text{C}$  and stopped when the dye front ran out of the gels.

After the SDS-PAGE images were scanned using Typhoon TRIO (GE Healthcare Bioscience, Pittsburgh, PA, USA), and analysis of the scanned images was conducted using Image Quant software (version 6.0, GE Healthcare). An in-gel analysis of the images was conducted using DeCyder software, version 6.5 (GE Healthcare Bioscience). For this step, a difference in-gel analysis (DIA) tool made by DeCyder software was used. The value of the estimated number of spots was 3000. For the sake of low experimental variation, an automated tool made by DeCyder software was chosen for background subtraction and normalization of visible protein spots. The DIA datasets, along with the images, were put into the biological variation analysis (BVA) module made by DeCyder software. The spot intensity data were normalized with the internal standard sample that was labeled with Cy2 dye. The protein spots ratios that had increased or decreased abundance by 1.5-fold and had a Student's *t*-test value of  $p \leq 0.05$  were considered to be differentially abundant, along with the extra condition that the spot should be present and analyzed in all 2D gels. These ratios were calculated by the DeCyder image analysis software from spot volumes. As per the manufacturer's instructions the spot ratios were calculated as follows: (volume of secondary image spot/volume of primary image spot). This ratio indicated the change in spot volume between the two images. These ratio values were normalized, so that the modal peak of volume ratios was zero (since the majority of proteins are not up or down regulated). This ratio parameter is referred to as the volume ratio. In all DeCyder 2D Software DIA tables (**Supplementary Tables 1, 2**) the volume ratio is expressed in the range of 1–1,000,000 for increases in spot volumes and  $-1$  to  $-1,000,000$  for decreases in spot volumes. Values between  $-1$  and  $1$  are not represented, hence a two-fold increase and decrease is represented by  $2$  and  $-2$ , respectively (and not  $2$  and  $0.5$  as might have been expected) (GE Healthcare Bio-Sciences, Pittsburgh, USA).

### Spot Picking and Trypsin Digestion and Mass Spectrometry

Protein spots that were statistically significant (see Section IEF, SDS-PAGE, Image Scan and Data Analysis) with a  $p \leq 0.1$  and cut off value of 1.5-fold were picked up using the Ettan Spot Picker (GE Healthcare). The picked gel spots were washed a few times, digested with modified porcine trypsin protease (Trypsin Gold, Promega), and then desalted using Zip-tip C18 (Millipore, Billerica, MA, USA) (Robbins et al., 2013; Gupta et al., 2014; Hayashi et al., 2015; Das et al., in press). Peptides were eluted from the Zip-tip with  $0.5 \mu\text{l}$  of matrix solution ( $\alpha$ -cyano-4-hydroxycinnamic acid, 5 mg/ml in 50% acetonitrile, 0.1% trifluoroacetic acid, and 25 mM ammonium bicarbonate) and spotted onto a MALDI plate (Robbins et al., 2013; Hayashi et al., 2015; Das et al., in press).

MALDI-TOF (MS) and TOF/TOF (tandem MS/MS) were performed on a 5800 mass spectrometer (AB Sciex). MALDI-TOF mass spectra were acquired in reflectron positive ion mode, averaging 2000 laser shots per spectrum. TOF/TOF tandem MS fragmentation spectra were acquired for each sample, averaging 2000 laser shots per fragmentation spectrum on each of the 10 most abundant ions present in each sample (excluding trypsin

autolytic peptides and other known background ions) (Gupta et al., 2014; Hayashi et al., 2015; Das et al., in press).

### Database Search

Analysis of the MS/MS results were performed using GPS Explorer, version 3.5, which was equipped with a MASCOT search engine (Matrix science). A search in the database of the National Center for Biotechnology Information non-redundant (NCBI nr) and in Phytozome v.10.3 ([http://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org\\_Pvirgatum](http://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Pvirgatum)) was performed without constraining the protein molecular weight or isoelectric point, with variable carbamidomethylation of cysteine, oxidation of methionine residues, and one missed cleavage allowed in the search parameters (**Table 2** along with **Supplementary Tables 5, 6**). Candidates with either protein score C.I.% (Confidence Interval) or Ion C.I.% greater than 95 were considered significant.

### Bioinformatic Analyses

The best hit proteins from MS/MS were blasted to the NCBI (<http://www.ncbi.nlm.nih.gov/>) and Phytozome (Goodstein et al., 2012) databases. NCBI, Arabidopsis Information Resource (<http://arabidopsis.org>) (Rhee et al., 2003), and Uniprot (<http://uniprot.org>) (The UniProt Consortium, 2008) were used to retrieve further information on protein functions. The Kyoto Encyclopedia of Genes and Genomes website (KEGG; <http://www.genome.jp/kegg/>) (Kanehisa and Goto, 2000) was used to retrieve information on proteins regarding their involvement in metabolic pathways. STRING database, version 9.1 (<http://string-db.org/>) (Szklarczyk et al., 2011), was used to predict protein-protein interactions. We retrieved primary interaction data from the STRING database and portrayed the interaction in Cytoscape 3.1.1 software, along with a categorization of their functions (Shannon et al., 2003).

## RESULTS AND DISCUSSION

### Genotypes of SG and PCG with Contrasting Senescence Phenotypes

The two contrasting genotypes of switchgrass selected from "Sunburst" differed in morphological and phenological traits. The LS genotype headed approximately 10 days later than the ES genotype and reached full senescence 2–3 weeks later in autumn. The LS genotype was more disease resistant, taller, and produced more average biomass yields ( $12.2 \text{ Mg dry matter ha}^{-1}$  compared with  $8.5 \text{ Mg dry matter ha}^{-1}$ ) than the ES genotype.

Similar differences were found between the LS PCG (origin in southeastern South Dakota) and the ES PCG (origin in southeastern North Dakota). The LS PCG headed approximately 2 weeks later and reached full senescence approximately 2–3 weeks later in autumn than the populations PCG plants. The LS PCG genotype was taller ( $2.5 \text{ m}$ ) than the ES PCG plants ( $1.25 \text{ m}$ ) at its maximum height during anthesis. At peak standing in mid-summer, ES PCG plants of PCG produced an average of  $7.1 \text{ Mg dry matter ha}^{-1}$  compared with  $15.4 \text{ Mg dry matter ha}^{-1}$  for the LS PCG plants.

**TABLE 2 | Proteins identified by 2D-DIGE followed by Mass Spectrometry.**

Spot number	Top ranked protein name	Accession No.	Protein MW (kDa)	Protein pI
1	Phosphoenolpyruvate carboxylase 1	CAPP1_MAIZE	109.22	5.8
2	Putative aconitate hydratase	ACOC_ORYSJ	98.02	5.7
3	**Phosphoenolpyruvate carboxylase	gi 154146924	49.67	0
4	Polyphosphate kinase	gi 116513638	83.23	5.8
5	Uncharacterized protein	gi 293332375	90.40	5.7
6	**Phosphoenolpyruvate carboxylase	gi 11991545	108.44	0
11	**Phosphoenolpyruvate carboxykinase	gi 220938712	55.61	0
13	PREDICTED: fibulin-2	gi 125829574	92.63	5.4
14, 264	*Phosphoenolpyruvate carboxylase	gi 154146806	49.30	0
15, 16, 19	*Transketolase, chloroplastic	TKTC_MAIZE	72.94	5.5
23	Pyruvate orthophosphate dikinase	gi 380691814	74.79	0
24	**V-type proton ATPase catalytic subunit A	VATA_MAIZE	61.91	5.9
25	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	PMGI_MAIZE	60.58	5.3
31	Succinate dehydrogenase [ubiquinone] flavoprotein subunit	DHSA_ORYSJ	68.81	6.6
33, 34, 35, 36	*Phosphoenolpyruvate carboxykinase	gi 220938712	55.61	7.6
38	ATP synthase subunit alpha	ATPA_SACOF	55.65	5.9
43	**Phosphoenolpyruvate carboxylase	gi 154146744	49.43	0.0
44	ATP synthase subunit alpha, mitochondrial	ATPAM_PHAVU	55.31	6.5
45, 46, 48	*ATP synthase subunit alpha, chloroplastic	ATPA_SACOF	55.65	5.9
49	ATP synthase subunit alpha, mitochondrial	ATPAM_ORYSI	55.33	5.9
53, 55	*ATP synthase subunit beta, chloroplastic	ATPB_SORBI	53.97	5.3
56, 127	*Ribulose biphosphate carboxylase large chain	RBL_SECDI	51.60	6.2
58, 59, 70	*Ribulose biphosphate carboxylase large chain	RBL_SETIT	52.64	6.4
60, 69, 286	*Ribulose biphosphate carboxylase large chain	RBL_AVEA	52.90	5.9
65	#Probable sucrose-phosphate synthase 1	SPSA1_CRAPL	118.95	6.1
97, 120	*Cysteine protease 1 precursor	gi 226496089	50.73	5.0
98	Sedoheptulose-1,7-bisphosphatase, chloroplastic	S17P_WHEAT	42.03	6.0
115	Fructose-bisphosphate aldolase, chloroplastic	ALFC_ORYSJ	41.98	6.4
128	Unknown	gi 223974857	39.04	7.9
129	**Chain A, Leaf Ferredoxin-Nadp+ Reductase I	gi 427930675	34.80	0
133	Oxygen-evolving enhancer protein 1, chloroplastic	PSBO_HELAN	34.20	5.4
134	Oxygen-evolving enhancer protein 1, chloroplastic	PSBO_SOLLC	34.92	5.9
140	**Glutathione S-transferase GSTF14	gi 46276327	30.71	0
142	**NAD-dependent epimerase/dehydratase	gi 226499246	27.75	0
144	**PREDICTED-proteasome subunit alpha type-7-B isoform 1	gi 357159679	27.28	0
146	**Predicted protein	gi 255082634	69.94	0
147	Beta-ketoacyl CoA thiolase	gi 171057883	41.78	7.6
208	Hypothetical protein azo3784	gi 119900073	12.28	8.2
209	Ribulose biphosphate carboxylase large chain	RBL_APHSI	51.55	6.0
221, 279	* <sup>S</sup> Putative cytochrome c oxidase subunit II PS17	PS17_PINST	1.70	9.6
225	Oxysterol-binding protein-related protein 1D	ORP1D_ARATH	92.27	6.1
226	cytochrome b6-f complex iron-sulfur subunit	gi 195612712	23.97	8.5
230	S-adenosylmethionine synthase	METK1_BRAJU	43.19	5.5
237	Protein CYP4	CYP4_CYNCA	55.61	7.6
273, 278	*Ras-related protein RAB1b	RAB1B_ARATH	23.16	6.5
275	14-3-3-like protein	1433_PEA	29.31	4.7
284	Ribulose biphosphate carboxylase large chain	RBL_LIQST	52.62	6.0
287	**Predicted protein	gi 326529055	44.22	0
291	Cysteine proteinase Mir3 precursor	gi 162463464	51.75	5.8
298	**Receptor-like protein kinase	gi 2832304	46.55	0

(Continued)

TABLE 2 | Continued

Spot number	Top ranked protein name	Accession No.	Protein MW (kDa)	Protein pI
301	**Auxin-binding protein ABP19a-like precursor	gil359806878	22.23	0
311	**Hypothetical protein ZEAMMB73_006777	gil413933720	6.12	0
313	**P1B-ATPase heavy-metal transporter	gil315623028	115.92	0
314	**Chloroplast PSI type III chlorophyll a/b-binding protein	gil159138869	13.85	0
316	**Hypothetical protein VITISV_010874	gil147843505	37.03	0
317	**Glutathione S-transferase F8	gil357114170	25.81	0
321	**Polyamine oxidase	gil14485487	56.47	0

\*Different spot number corresponds to the same protein.

\*\*pI value 0 corresponds to the proteins with isoelectric point either below 3.0 or above 10.0 or it's mixed up with a nearby spot, thus pI value is unspecific.

#Largest protein based on molecular weight; \*Smallest protein based on MW.

## Proteomic Response during Pre- and Post-Senescence in PCG and SG

A representative 2D-DIGE gel image of sample A1/C7 is shown in **Figure 1**. 2D-DIGE along with MS/MS revealed the differential abundance of 74 protein spots (**Figure 1**). Among those, 69 different protein spots were selected based on statistical analyses (threshold of significance of  $p \leq 0.1$ ) and a cut off value of 1.5-fold increase/decrease in protein abundance, which is shown in the heat map (**Figure 2**).

To achieve an improved understanding of the proteomic responses due to senescence, differentially abundant proteins that were identified in both PCG and SG were categorized according to their GO annotation function (**Figure 3**; Camon et al., 2003). Our analysis indicated that 69 differentially abundant proteins were found to be involved in various biological processes, including photosynthesis (41%), amino acid metabolism (13%), carbohydrate metabolism (12%), ATP synthesis (6%), protein metabolism (5%), kinase activity (4%), ATP signal transduction (3%), cell division (3%), sterol metabolism (1%), Auxin binding (1%), and GTPase signal transduction (1%).

When we analyzed the fold change in protein abundance for after-to-before senescence in LS and ES PCG, nine unique protein spots were found to be differentially abundant that were absent in SG (**Supplementary Table 1**), whereas ten protein spots were found to have common abundance patterns between SG and PCG (**Figure 4**). Among those 19 protein spots of interest, 9 had higher abundance, and 10 had lower abundance. Our proteomic profiling of both ES and LS SG genotypes revealed that 28 major protein spots were differentially abundant (**Supplementary Table 2**). Among these 28 protein spots, 18 proteins were found exclusively in SG, whereas the other 10 protein spots were found in both SG and PCG (**Figure 3**). Overall, 10 protein spots had increased and 18 protein spots had decreased abundance in SG.

To investigate the changes in the protein profiles of before—and after the onset of senescence (four different comparisons: B/A, D/C, F/E, and H/G) in PCG and SG (**Table 1**), we identified 10 major differentially abundant protein spots based on statistical significance ( $p \leq 0.05$ ) (seven different proteins: putative aconitate hydratase, ribulose biphosphate carboxylase large chain, oxygen-evolving enhancer protein 1, glutathione S-transferase, hypothetical protein ZEAMMB73, and hypothetical

protein VITISV) (**Figure 2**). Five protein spots were found to have higher abundance and five spots had lower abundance, most likely in response to senescence processes.

## Biological Implications of Selected Proteins

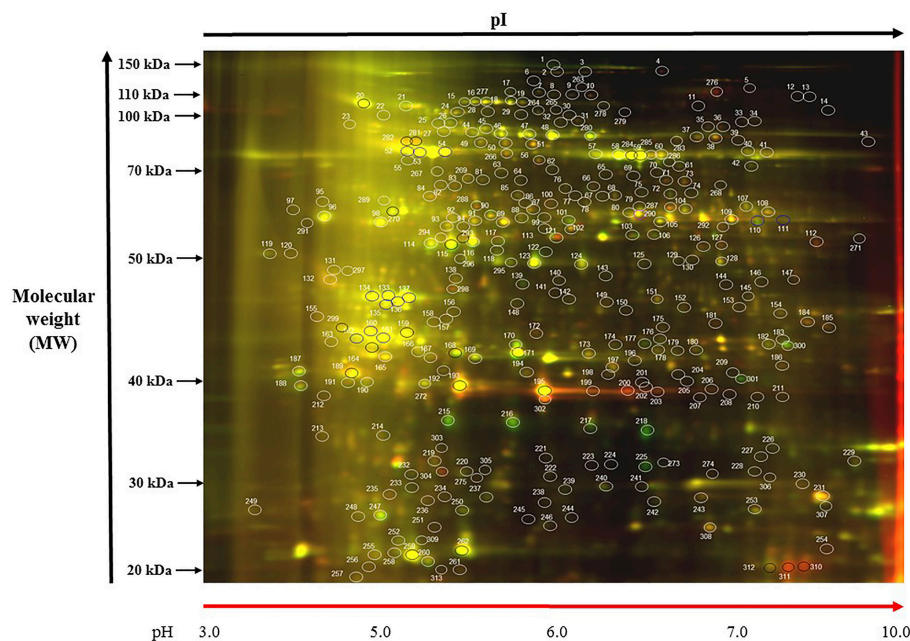
### Role of Putative Aconitate Hydratase during Senescence

Putative aconitate hydratase was found to have increased abundance during senescence in all four comparisons (**Figure 5**). Aconitate hydratase catalyzes steps in the TCA cycle and glyoxylate cycle that isomerizes citrate to isocitrate (Evans et al., 1966). Sugar metabolism toward gluconeogenesis, hexose formation, and conversion to sucrose seems to be an important phenomenon to signal the source-sink translocation in perennial grasses (**Supplementary Figure 2**). Aconitate hydratase in *Arabidopsis* is reported to also bind mRNA of CSD2 (CuZn superoxide dismutase 2), which was found to play a key role in antioxidant defense mechanisms (Gregersen and Holm, 2007; Moeder et al., 2007). Moreover, reduced aconitate hydratase levels in cells reportedly enhanced resistance to oxidative stress. It was also reported that upregulation of the cytoplasmic aconitate hydratase gene in wheat flag leaf plays a crucial role in the mitochondria with respect to oxidative cell damage (Gregersen and Holm, 2007). Therefore, it is possible that aconitate hydratase in both SG and PCG helps to mediate oxidative stress and regulate plant growth or senescence by influencing the expression of genes for redox balance (Moeder et al., 2007). Two genes encoding aconitate hydratases were significantly upregulated in senescing SG flag leaves (Palmer et al., 2015), corroborating the findings at the protein level.

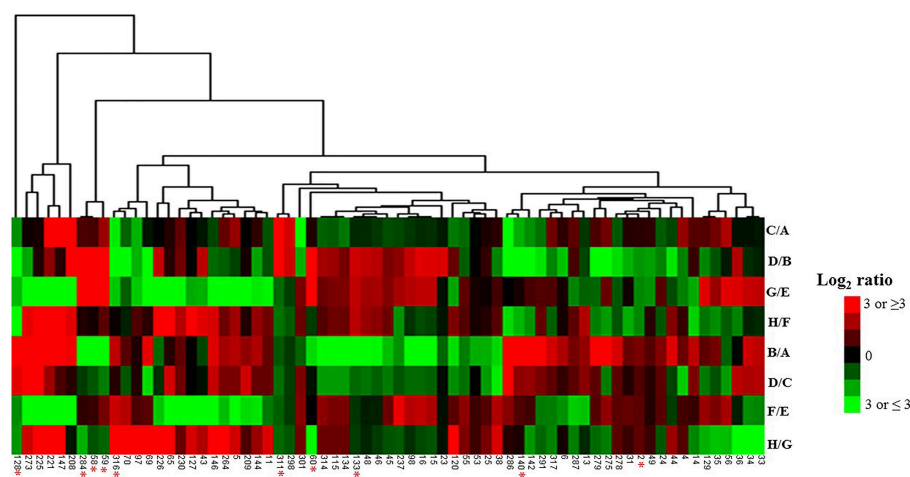
### Role of Glutathione S Transferase in Anti-Oxidation and Its Early Senescing and Late Senescing Response in Prairie Cordgrass

It has been reported that GST is a marker gene for oxidative stress and that it participates in cellular protection when plants are subjected to wounding, pathogen attack, and lipid peroxidation (Bilang and Sturm, 1995; Grant et al., 2000). Consistent with previous findings, our proteomic analysis also showed significantly higher levels (five-fold) of a GST protein





**FIGURE 1 | Representative gel image: (A1/C7).** Protein samples of SG and PCG leaves, control and treatment samples, were labeled with Cy3 (green) and Cy5 (red), respectively, and mixed in equal ratios. The first-dimension isoelectric points (pI, ranges from 3 to 10) and second-dimension molecular mass (in kDa and ranges from 20 to 150 kDa) are noted. Color coding: green spots indicate protein abundance is high in Cy3, red spots indicate protein abundance is high in Cy5, yellow spots indicates where protein abundance is similar in both the cases.

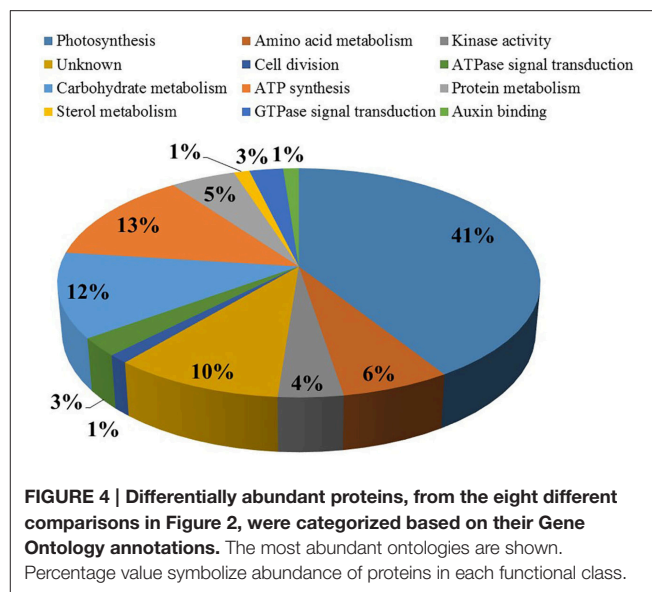
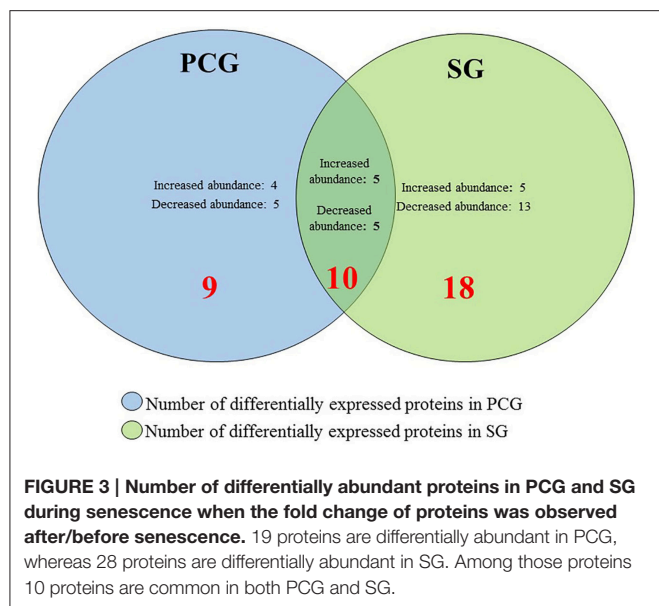


**FIGURE 2 | Heat map of 69 protein identified as differentially abundant during senescence.** Heat map of proteins associated with senescence in eight different experiment groups. C/A, Before senescence (late/early SG); D/B, After senescence (late/early SG); G/E, Before senescence (late/early PCG); H/F, After senescence (late/early PCG); B/A, Early SG (after/before senescence); D/C, Late SG (after/before senescence); F/E, Early PCG (after/before senescence); H/G, Late PCG (after/before senescence). Hierarchical clustering was done as described in the method Section Bioinformatic Analyses. \*Means selected 10 protein spots (7 different proteins) showing common abundance pattern between SG and PCG when analyzing the fold change of protein abundance for after to before senescence.

when its fold change was analyzed before and after senescence in the ES SG genotype (Figure 5).

Our results also showed that a GST was found to be significantly higher in LS PCG before senescence than ES PCG (Figure 6). However, after senescence, the levels were higher in ES PCG than LS PCG (Supplementary Table 1), similar to what

was observed for SG. This observation suggests that consistent higher abundance of GST levels in PCG before the onset of senescence may play a key role in delaying senescence signals by potentially delaying ROS mediated signaling. On the other hand, we found increased levels of this GST during senescence when we compared GST levels from after-to-before senescence in four



different sets of comparisons (B/A, D/C, F/E, and H/G). These results imply that the levels of this specific GST increased during senescence. ROS are continuously generated in plants in different organelles, such as mitochondria, chloroplasts, lysosomes, and peroxisomes, but are eliminated by different mechanisms to keep a balance between generation and removal. GST plays a crucial role in the removal of reactive oxygen species. Major pathways involving GST retrieved from KEGG indicate that GST seems to be a crucial enzyme in balancing cellular oxidative stress.

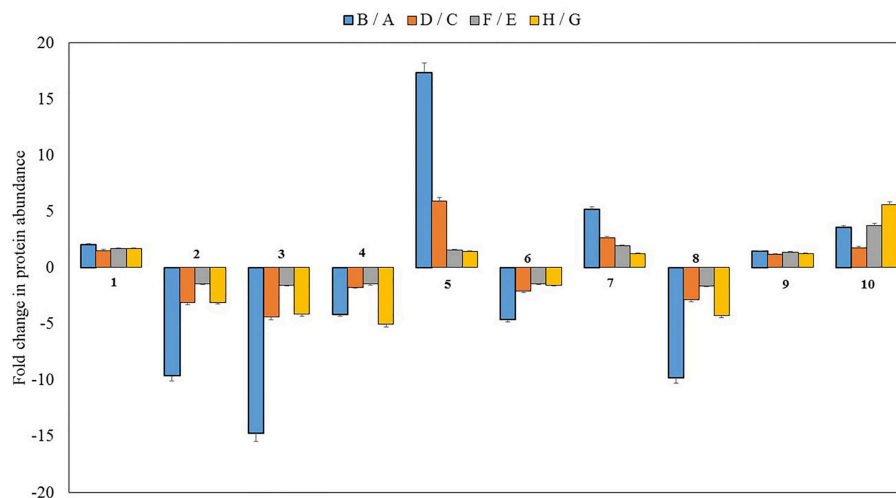
### Proteins Involved in the Regulation of Photosynthesis During Pre- and Post-Senescence

It is known that photosynthetic activity declines rapidly during leaf senescence (Kura-Hotta et al., 1987). Previous studies in rice showed that RuBisCO is highly active during leaf expansion but drastically reduced during senescence (Suzuki et al., 2001). Based on our 2-D DIGE analysis in SG and PCG, we also found significantly reduced levels of RuBisCO large subunit when the fold changes for this protein were compared before and after senescence in both PCG and SG genotypes (Figure 5). RuBisCO is the major enzyme responsible for assimilating carbon dioxide (Spreitzer and Salvucci, 2002). Reduced RuBisCO content can be expected to reduce CO<sub>2</sub> fixation and net photosynthesis. It has been reported that RuBisCO is the prime target for degradation as a consequence of increased levels of ROS in chloroplasts (Khanna-Chopra, 2012).

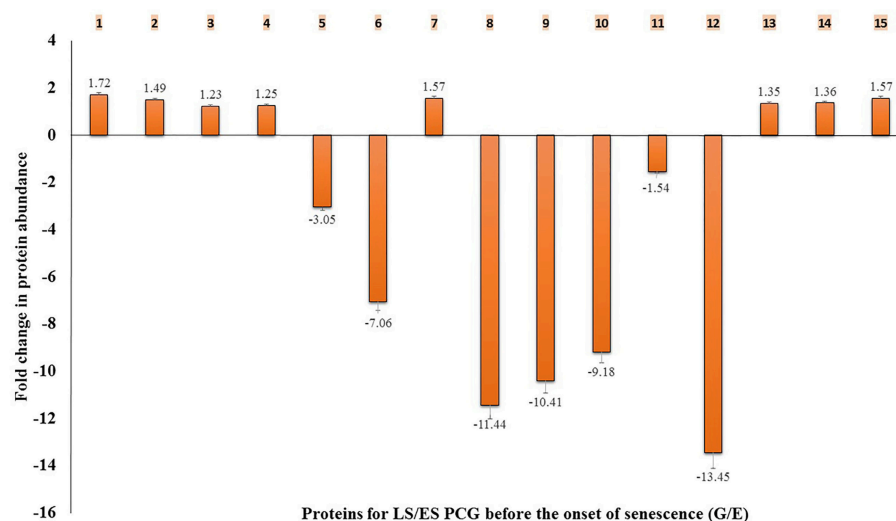
Furthermore, 2D-DIGE analysis in the current investigation revealed that other proteins involved in photosynthesis, such as oxygen evolving enhancer (OEC) protein in photosystem II, ferredoxin-NADP-reductase, and the  $\alpha$ - and  $\beta$ -subunits of ATP synthase, had higher abundance in the LS PCG compared with the ES PCG genotype, critically when protein fold changes were compared before senescence. All of these proteins are involved in electron transfers or in ATP synthesis in chloroplasts, (Haehnel, 1984). Consistent higher abundance of these proteins

in the LS PCG compared with the ES PCG before senescence suggests that an overall upregulated machinery of photosynthesis in plants may contribute to delaying the senescence process (Supplementary Figure 3). In line with our findings in PCG, downregulation of OEC has also been found to decrease photosynthesis in mutant *Arabidopsis* plants where one of the *psbO* genes was defective (Hager et al., 2002; Ifuku et al., 2005).

Higher abundance of the enzymes involved in photosynthetic electron transfer and ATP synthesis, along with the increased abundance of enzymes involved in the Calvin cycle, seems to contribute to the accumulation of higher biomass in LS PCG compared with ES PCG. However, consistent increased abundance of the photosynthetic machinery itself may have contributed to delay senescence in the LS genotypes. Photosystem II and its corresponding OEC proteins are tightly regulated during senescence (Lim et al., 2007). Downregulation of *OEC* gene has been known to reduce photosynthetic activity in mutant *Arabidopsis* plants, indicating that OEC protein plays a vital role in photosynthesis by oxidizing water in photosystem II on thylakoid membranes (Lundin et al., 2007; Dwyer et al., 2012). Interestingly, our proteomic evaluation showed that the levels of OEC proteins were reduced after the onset senescence compared with before senescence in PCG leaves and similar results were observed in SG as well (Figure 5). Although why *OEC* is downregulated during senescence remains unknown, it is clear that downregulation of *OEC* leads to photo-damage in photosystem II, with the malfunctioning of acceptor and donor electron carriers (Allahverdiyeva et al., 2009). This photo-damage may play a role in producing more ROS, thus destroying chlorophylls and RuBisCO large subunit and signaling the initiation of senescence. Interestingly, higher abundance of ferredoxin NADP-reductase and the chloroplastic  $\alpha$ - and  $\beta$ -subunits of ATP synthase in the LS PCG compared with the ES PCG also supports the idea that increased abundance of photosynthesis-related proteins may contribute to delay the senescence process in this plant.



**FIGURE 5 | Fold change of 10 differentially abundant proteins in the after to before senescence comparisons among the four different genotypes ( $p < 0.01$ ).** B/A, Early SG (after/before senescence); D/C, Late SG (after/before senescence); F/E, Early PCG (after/before senescence); H/G, Late PCG (after/before senescence). 1- putative aconitate hydratase, 2, 3, 4, 8- ribulose biphosphate carboxylase large chain, 5- unknown protein of Zea mays, 6- oxygen-evolving enhancer protein 1, 7- a glutathione S-transferase, 9- hypothetical protein, 10- hypothetical protein VITISV. Standard error was calculated and displayed with error bar.



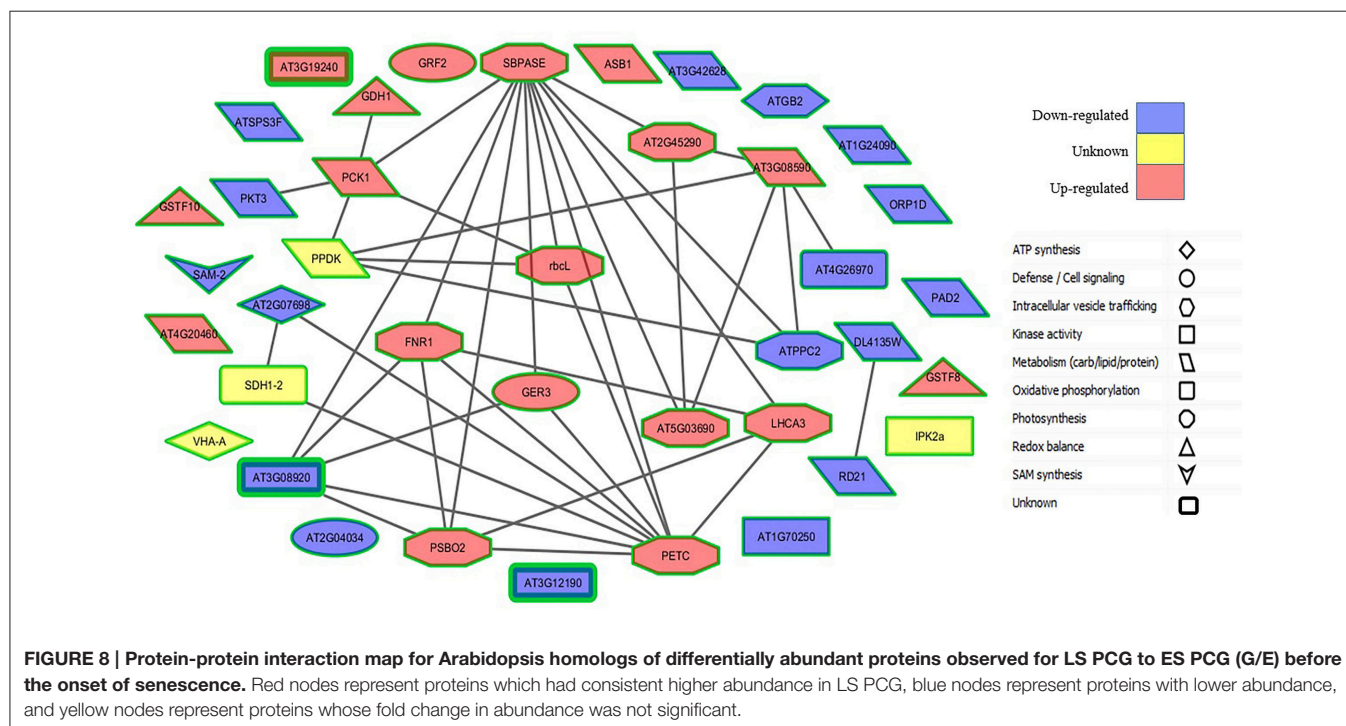
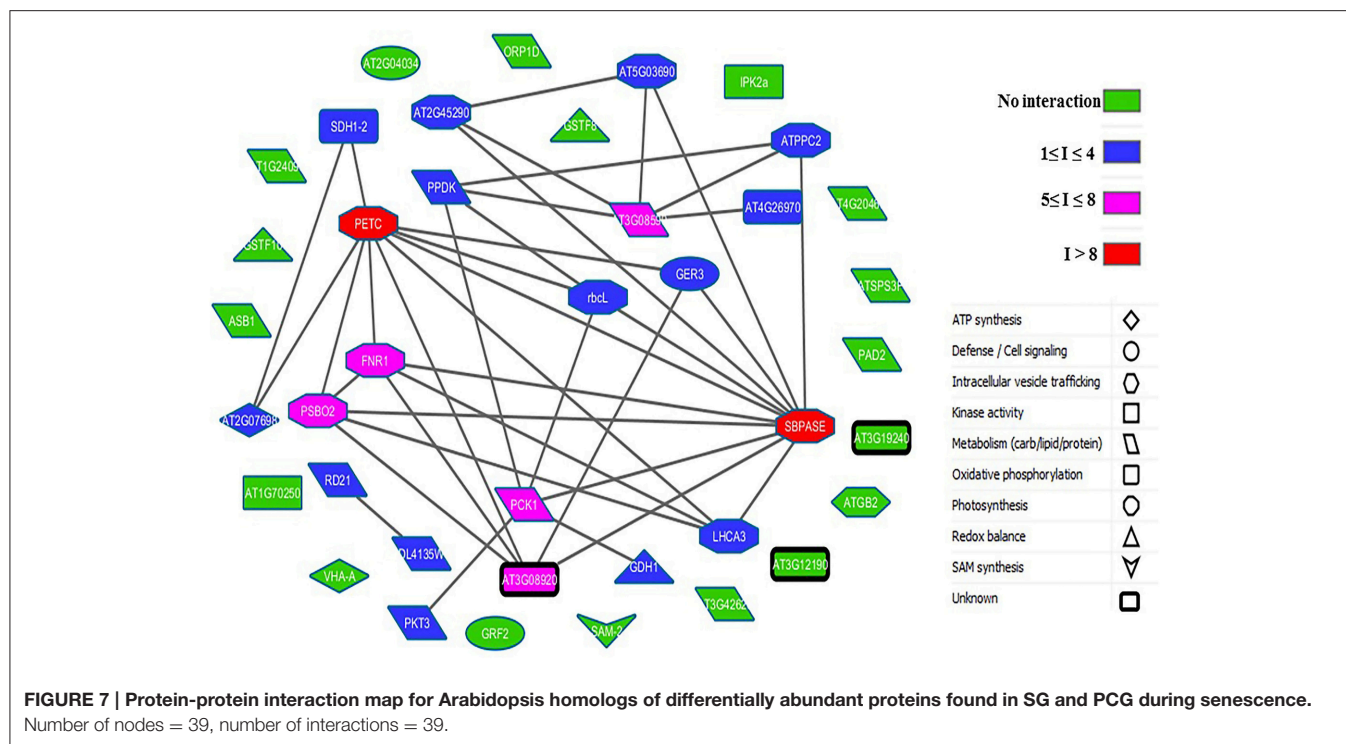
**FIGURE 6 | Fold change of abundance in 15 different proteins for LS/ES PCG before the onset of senescence (G/E).** 1- transketolase, 2- sedoheptulose-1,7-bisphosphatase, 3- fructose-bisphosphate aldolase, 4- glutathione S-transferase GSTF14, 5- S-adenosylmethionine synthase, 6- Ras-related protein, 7-14-3-3-like protein, 8- oxysterol-binding protein-related protein, 9- probable sucrose-phosphate synthase, 10- beta-ketoacyl CoA thiolase, 11- cysteine protease 1 precursor, 12- putative cytochrome c oxidase subunit II PS17, 13- ATP synthase subunit alpha, 14- ATP synthase subunit beta, 15- oxygen-evolving enhancer protein 1. Standard error was calculated and displayed with error bar.

### Role of 14-3-3 Like Protein, S-adenosylmethionine Synthase, and Ras-Related RAB Protein in Senescence Signaling

The 14-3-3 like proteins are well studied and are involved in apoptotic and cell survival signaling in *Arabidopsis* (van Hemert et al., 2001). Similarly, S-adenosylmethionine (SAM) synthase generates SAM which is involved in the ethylene synthesis pathway and in many methylation-dependent pathways. The

involvement of RAB protein in vesicular transport mechanisms of cells is well documented (Grbić and Bleecker, 1995; Fu et al., 2000; Woollard and Moore, 2008). We found differential abundance of these proteins in LS PCG to ES PCG (G/E) before senescence (Figure 6).

We observed that a 14-3-3 like protein was highly abundant in LS PCG compared with ES PCG before senescence, whereas SAM synthase and RAB protein had lower abundance. Earlier



studies showed that a 14-3-3 protein antagonizes pro-apoptotic activity by playing a key role in integrating the signals for cell death and survival (Fu et al., 2000). It has been reported that the overexpression of the Arabidopsis 14-3-3 gene *GF14λ* in cotton plants resulted in maintaining green plants and significantly delaying senescence and improving drought stress tolerance (Yan

et al., 2004; Gregersen et al., 2013). Similarly, the overexpression of a 14-3-3 gene in potato plants delayed senescence and elevated antioxidant activity, whereas the downregulation of this gene led to early senescence (Wilczyński et al., 1998; Łukasiewicz et al., 2002). Thus, the overexpression of genes encoding 14-3-3 proteins with appropriate promoters in PCG may be a good



strategy to delay senescence. Studies on RAB genes related to senescence are very limited, but their overexpression is known to accelerate leaf senescence in mutant *Arabidopsis* plants (Kwon et al., 2009). Consistent lower abundance of RAB proteins in the LS PCG compared with the ES PCG also highlights the probable role of RAB genes in the senescence processes of this plant.

### Higher Abundance of the Calvin Cycle and Pentose Phosphate Pathway Derived Proteins in Late Senescing PCG

Transketolase, sedoheptulose-1,7-bisphosphatase (SBPase), and fructose-bisphosphate aldolase, all of chloroplastic origin, had increased abundance in LS PCG compared with ES PCG when the fold changes were evaluated before senescence. All three enzymes are part of the Calvin cycle and play a role in important rate limiting steps (**Supplementary Figure 4**).

As one of the crucial pentose phosphate pathway (PPP) enzymes, transketolase catalyzes the reversible reaction in the formation of ribulose-5-phosphate, and this reaction increases the regeneration rate of RuBisCO (Jones et al., 2012). Transketolase also facilitates isoprenoids synthesis and ultimately reduces oxidative stress (Bouvier et al., 1998, 2005; Lichtenthaler, 1999; Loreto et al., 2001; Vickers et al., 2009). Therefore, the regulation of the PPP and the Calvin cycle seems to be a factor during senescence in plants. Increased abundance of transketolase in LS PCG compared with ES PCG potentially signifies its role in PPP signaling required for cell survival or senescence.

Similarly, SBPase, which produces sedoheptulose-7-phosphate during the Calvin cycle, speeds up the regeneration of RuBisCO with RuBP, thereby speeding up carbon fixation (Lefebvre et al., 2005). An SBPase loss of function mutant in *Arabidopsis*, *sbp*, showed severe growth retardation, poor SBPase dependent carbon assimilation, and starch biosynthesis (Moore et al., 2003). Moreover, overexpression of *SBPase* in tobacco plants resulted in enhanced photosynthesis and growth rates in the early stage of development (Lefebvre et al., 2005; Feng et al., 2007; Liu et al., 2012), and increased growth rate and accumulation of biomass (Tamoi et al., 2006). Similarly, the higher abundance of SBPase and fructose-bisphosphate aldolase in LS PCG also corroborates the importance of the Calvin cycle during senescence. The overexpression of SBPase, with suitable promoters in tobacco and rice, has been reported to enhance photosynthetic rates and growth rates, whereas the downregulation of this gene resulted in retarding growth (Lefebvre et al., 2005; Tamoi et al., 2006; Feng et al., 2007; Liu et al., 2012). Therefore, the overexpression of the genes responsible for this enzyme may be a good strategy for increasing biomass in biofuel crops, such as PCG and SG. Data from the study of natural senescence of switchgrass flag leaves would support a positive role for SBPase in delaying senescence. SBPase transcript levels decreased significantly in flag leaves with the onset of senescence in field-grown switchgrass plants (Palmer et al., 2015).

Another potential Calvin cycle enzyme, fructose bisphosphate aldolase, regulates photosynthetic carbon flux. Importantly, downregulation of fructose bisphosphate aldolase causes reduced

photosynthesis, altered levels of sugar and starch, and retarded growth in potato plants (Haake et al., 1998). Studies on young *Brassica napus* leaves showed that fructose bisphosphate aldolase expression is lower in mature green leaves but is then enhanced again during senescence (Buchanan-Wollaston, 1997). Consistent with earlier reports, our proteomic profiling shows increased abundance of fructose-bisphosphate aldolase in LS PCG, suggesting the significance of Calvin cycle activities in this late senescing genotype. It has also been reported that the overexpression of the *Arabidopsis* plastid fructose bisphosphate aldolase in tobacco plants resulted in augmented biomass by up to 2.2-fold (Uematsu et al., 2012). Producing plants overexpressing the genes encoding this enzyme may be another strategy to produce more biomass in biofuel crops. Biomass accumulation in LS PCG is higher compared with ES PCG (Boe, unpublished). Thus, the accumulation of higher amounts of biomass in LS PCG seems to be the result of upregulation of the Calvin cycle, increased photosynthesis, and increased carbon assimilation.

In  $C_4$  plants, such as SG and PCG,  $CO_2$  is generally not the rate-limiting factor as the oxygenase activity of RuBisCO is mostly limited due to high internal  $CO_2$  concentrations in these plants. As a result, the regeneration rate of RuBisCO plays a critical role in enhancing the photosynthesis rate and increasing biomass. Increased abundance of Calvin cycle enzymes (transketolase, sedoheptulose-1,7-bisphosphatase, and fructose-bisphosphate aldolase) along with increased abundance of photosynthesis machinery (oxygen evolving enhancer proteins, ferredoxin-NADP-reductase, and the  $\alpha$ - and  $\beta$ -subunits of ATP synthase) likely increases the regeneration rate of RuBisCO and the overall carbon assimilation in the LS PCG genotype.

### Oxidative Phosphorylation and Senescence

Two key proteins of oxidative phosphorylation, succinate dehydrogenase (Complex II) and cytochrome c oxidase (Complex IV), were found in differential abundance during senescence. Complex IV was found to be more abundant in ES PCG compared with LS PCG before senescence, whereas complex II had increased abundance during senescence in PCG when the protein fold change was analyzed for “after senescence” to “before senescence” (**Supplementary Figure 5**). Oxidative phosphorylation is well shown to generate ROS in both animal and plant cells (Fleury et al., 2002; Turrens, 2003).

Consistent lower abundance of complex IV of the oxidative phosphorylation cycle (i.e., cytochrome c-oxidase) in LS PCG compared with ES PCG, along with higher abundance of complex II (i.e., succinate dehydrogenase) during senescence in PCG, suggests the probable role played by oxidative phosphorylation pathways in the senescence mechanism of PCG. Upregulation of oxidative phosphorylation may contribute to the generation of ROS, thereby signaling for senescence; thus, its downregulation may play a role in delaying the senescence process, as observed in LS PCG. We also found that during senescence, there was higher abundance of proteins involved in fatty acid break down, proteolysis, starch conversion, and gluconeogenesis. Upregulation of fatty acid breakdown also

mediates mitochondrial dysfunction ROS generation, ultimately increasing oxidative stress in plants. Upregulation of genes involved in fatty acid break down, proteolysis, starch conversion, and gluconeogenesis also results in hexose accumulation, which signals for developmental senescence (Dai et al., 1999). Similarly, upregulation of sucrose phosphate synthase during senescence may have played a role in sucrose formation from hexoses and translocation to the sinks, which are rhizomes in the case of these perennial grasses (Lemoine et al., 2013). Many of these changes at the protein level mirror data observed at the transcript level in senescing switchgrass flag leaves (Palmer et al., 2015).

## Prediction of Protein-Protein Interactions

To better understand cellular and molecular functions, it is crucial to address various functional interactions between proteins. Using computational methods (see Section Bioinformatic Analyses), we tried to model a variety of functional protein-protein interactions between the differentially abundant proteins during senescence in PCG and SG to elucidate the complex underlying regulatory processes. From the database search, we found 39 homologous proteins in *Arabidopsis* (Supplementary Table 3) and analyzed their interactions (Supplementary Table 4). Our computational analysis of protein-protein interaction networks revealed two major proteins (PETC and SBPase) that have more than eight interactions considered to be at the central body of the network system (Figure 7). Most of the proteins central to this network were related to photosynthesis, whereas other connected proteins were involved in metabolism, ATP synthesis, and cell signaling (Supplementary Table 4). The majority of the proteins involved in photosynthesis were more abundant in LS PCG compared with ES PCG before senescence (Figure 8). This suggests that the upregulation of the photosynthesis machinery is a key factor in delaying senescence. Senescence mediated differential expression of genes encoding these central body proteins of the network may ultimately affect the specific pathways of related predicted partner proteins during growth.

## CONCLUSION

This study identified 69 statistically significant, differentially abundant proteins in PCG and SG genotypes contrasting in senescence, namely ES (early senescence) and LS (late senescence), using 2-Dimensional, differential in-gel electrophoresis (2D-DIGE) followed by mass spectrometry of leaf samples collected just before and after onset of senescence. The goal was to begin a catalog of leaf proteins that could be indicators of either phenotype as a starting point for more proteomic and physiological studies to understand mechanisms impacting senescence in these two biofuel grasses. LS genotypes of both grasses analyzed before senescence contained significantly higher abundances of a 14-3-3 like protein and a glutathione-S-transferase protein when compared to the ES genotypes. The higher abundance of 14-3-3 like proteins may be one factor that impacts the senescence process in both LS PCG and LS SG and provides a target at the protein and genetic level to evaluate in field nurseries containing PCG and SG germplasm.

Species specific differences in proteins between the ES and LS genotypes were also observed, indicating that subtle variations in the accumulation of specific proteins could influence longevity. As an example, the maintenance of proteins associated with the photosynthetic machinery in LS PCG compared to ES PCG suggests that factors that control proteolysis within plastids could be important in delaying senescence and boosting biomass yields, especially toward the end of the rowing season. Overall this study opens the door toward a more vibrant understanding of various differentially abundant senescence-related proteins and should prove valuable to the future improvement of perennial grasses for higher biomass yields (Kim et al., 2014; Das et al., 2015). Application of such gel-based quantitative proteomic approaches together with mapping of post-translational modifications will provide comprehensive insights into the regulation of various senescence-related proteins, corresponding to their biological function (Komatsu et al., 2013).

## AUTHOR CONTRIBUTIONS

AB, GS, JG, PR, and JR designed the experiments. BP, AD, MT performed the experiments. BP, AD, GS, JG, PR, NP, and JR analyzed the results. BP, AD, AB, NP, GS, JG, and JR wrote the manuscript.

## FUNDING

This research was supported by funding from the North Central Regional Sun Grant Center at South Dakota State University through a grant provided by the US Department of Agriculture under award number 2010-38502-21861. This work was supported in part by the USDA-ARS CRIS project 3042-21000-030-00D and CRIS project SD00H541-15. The U.S. Department of Agriculture, Agricultural Research Service, is an equal opportunity/affirmative action employer and all agency services are available without discrimination. Mention of commercial products and organizations in this manuscript is solely to provide specific information. It does not constitute endorsement by USDA-ARS over other products and organizations not mentioned.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.00293>

**Supplementary Figure 1 | Pre- and post-senescence periods were determined by measuring the chlorophyll content.** Red arrow indicates the mid line between pre- and post-senescence period that was determined on the basis of chlorophyll measurement. Sample interval indicates number of weeks and CCI indicates relative amount of chlorophyll content in that particular sample.

**Supplementary Figure 2 | Proposed model of sugar metabolism toward hexoses and sucrose formation during senescence which signals for early floral development, translocation of sugars from source to sink, and early senescence.** Analysis revealed higher abundance of five proteins shown above during senescence, whereas three proteins ( $\beta$ -ketoacyl CoA thiolase, cysteine protease, and sucrose phosphate synthase) had consistent lower abundance in LS PCG compared to ES PCG when analyzed before senescence onset.

**Supplementary Figure 3 | Three proteins (oxygen evolving enhancer protein of photosystem II, ferredoxin-NADP-reductase, and  $\alpha$ - and  $\beta$ -subunit of ATP synthase) found with consistently higher abundance in LS PCG compared to ES PCG that are involved in key steps of photosynthesis.** Figure of photosynthesis reproduced from KEGG reference pathway (<http://www.genome.jp/kegg/pathway/map/map00195.html>).

**Supplementary Figure 4 | Proteins of the Calvin cycle found in greater abundance in LS PCG compared to ES PCG that catalyze the vital rate limiting steps are marked with red arrows.** These enzymes are 1- fructose biphosphate aldolase, 2- transketolase, 3- sedoheptulose-1,7-bisphosphatase.

**Supplementary Figure 5 | Increased abundance of complex II, and complex IV of the oxidative phosphorylation pathway as observed in our study.** Higher abundance of these complexes may contribute to increased ROS levels, thus signaling the expression of senescence associated genes. Consistent higher abundance of complex IV was observed in ES PCG compared LS PCG. Figure of oxidative phosphorylation reproduced from KEGG pathway (<http://www.genome.jp/kegg/pathway/map/map00190.html>).

**Supplementary Table 1 | Fold change in differential abundance of protein spots for after to before senescence, in Early and Late PCG.** F/E represents

ratio for after to before senescence in Early PCG, whereas H/G represents that ratio for late PCG. Positive values represent increased abundance; negative values represent decreased abundance. All ratios included are statistically significant ( $p < 0.01$ ).

**Supplementary Table 2 | Fold change in differential abundance of protein spots for after to before senescence, in Early and Late SG.** B/A represents ratio for after to before senescence in Early SG, whereas D/C represents that ratio for late SG. Positive values represent increased abundance; negative values represent decreased abundance. All ratios included are statistically significant ( $p < 0.01$ ). Match quality H means high confidence, L means low confidence and N means no confidence. \*NA, not available.

**Supplementary Table 3 | Proteins, and their Arabidopsis homologs and percentage similarity.**

**Supplementary Table 4 | Protein function in protein-protein interaction table.**

**Supplementary Table 5 | Protein identifications with NCBI nr database search.**

**Supplementary Table 6 | Protein identifications with Phytosome v10.3 database search.**

## REFERENCES

- Ahonsi, M., Ames, K., Gray, M., and Bradley, C. (2013). Biomass reducing potential and prospective fungicide control of a new leaf blight of miscanthus  $\times$  giganteus caused by leptosphaerulina chartarum. *Bioenergy Res.* 6, 737–745. doi: 10.1007/s12155-012-9293-0
- Allahverdiyeva, Y., Mamedov, F., Holmström, M., Nurmi, M., Lundin, B., Styring, S., et al. (2009). Comparison of the electron transport properties of the *psbo1* and *psbo2* mutants of *Arabidopsis thaliana*. *Biochim. Biophys. Acta* 1787, 1230–1237. doi: 10.1016/j.bbabo.2009.05.013
- Bilang, J., and Sturm, A. (1995). Cloning and characterization of a glutathione S-transferase that can be photolabeled with 5-azido-indole-3-acetic acid. *Plant Physiol.* 109, 253–260. doi: 10.1104/pp.109.1.253
- Biswal, B., and Biswal, U. C. (1999). Leaf senescence: physiology and molecular biology. *Curr. Sci.* 77, 775–782.
- Bleecker, A. B. (1998). The evolutionary basis of leaf senescence: method to the madness? *Curr. Opin. Plant Biol.* 1, 73–78. doi: 10.1016/S1369-5266(98)80131-3
- Bouvier, F., D'Harlingue, A., Suire, C., Backhaus, R. A., and Camara, B. (1998). Dedicated roles of plastid transketolases during the early onset of isoprenoid biogenesis in pepper fruits. *Plant Physiol.* 117, 1423–1431. doi: 10.1104/pp.117.4.1423
- Bouvier, F., Rahier, A., and Camara, B. (2005). Biogenesis, molecular regulation and function of plant isoprenoids. *Prog. Lipid Res.* 44, 357–429. doi: 10.1016/j.plipres.2005.09.003
- Buchanan-Wollaston, V. (1997). The molecular biology of leaf senescence. *J. Exp. Bot.* 48, 181–199. doi: 10.1093/jxb/48.2.181
- Camon, E., Magrane, M., Barrell, D., Binns, D., Fleischmann, W., Kersey, P., et al. (2003). The Gene Ontology Annotation (GOA) project: implementation of GO in SWISS-PROT, TrEMBL, and InterPro. *Genome Res.* 13, 662–672. doi: 10.1101/gr.461403
- Carp, M.-J., and Gepstein, S. (2007). “Genomics and proteomics of leaf senescence,” in *Senescence Processes in Plants*, ed S. Gan (Oxford, UK: Blackwell Publishing Ltd.), 202–230.
- Chuck, G. S., Tobias, C., Sun, L., Kraemer, F., Li, C., Dibble, D., et al. (2011). Overexpression of the maize *Corngrass1* microRNA prevents flowering, improves digestibility, and increases starch content of switchgrass. *Proc. Natl. Acad. Sci. U.S.A.* 108, 17550–17555. doi: 10.1073/pnas.1113971108
- Dai, N., Schaffer, A., Petreikov, M., Shahak, Y., Giller, Y., Ratner, K., et al. (1999). Overexpression of Arabidopsis hexokinase in tomato plants inhibits growth, reduces photosynthesis, and induces rapid senescence. *Plant Cell Online* 11, 1253–1266. doi: 10.1105/tpc.11.7.1253
- Das, A., Eldakak, M., Paudel, B., Kim, D.-W., Hemmati, H., Basu, C., et al. (in press). Leaf proteome analysis reveals prospective drought and heat stress response mechanisms in soybean. *Biomed. Res. Int.*
- Das, A., Paudel, B., and Rohila, J. (2015). “Potentials of proteomics in crop breeding,” in *Advances in Plant Breeding Strategies: Breeding, Biotechnology and Molecular Tools, 1st Edn.*, eds J. M. Al-Khayri, S. M. Jain, and D. V. Johnson (Cham: Springer International Publishing), 513–537.
- Dwyer, S. A., Chow, W. S., Yamori, W., Evans, J. R., Kaines, S., Badger, M. R., et al. (2012). Antisense reductions in the PsbO protein of photosystem II leads to decreased quantum yield but similar maximal photosynthetic rates. *J. Exp. Bot.* 63, 4781–4795. doi: 10.1093/jxb/ers156
- Evans, M., Buchanan, B. B., and Arnon, D. I. (1966). A new ferredoxin-dependent carbon reduction cycle in a photosynthetic bacterium. *Proc. Natl. Acad. Sci. U.S.A.* 55, 928. doi: 10.1073/pnas.55.4.928
- Feng, L., Han, Y., Liu, G., An, B., Yang, J., Yang, G., et al. (2007). Overexpression of sedoheptulose-1, 7-bisphosphatase enhances photosynthesis and growth under salt stress in transgenic rice plants. *Funct. Plant Biol.* 34, 822–834. doi: 10.1071/FP07074
- Fleury, C., Mignotte, B., and Vayssières, J.-L. (2002). Mitochondrial reactive oxygen species in cell death signaling. *Biochimie* 84, 131–141. doi: 10.1016/S0300-9084(02)01369-X
- Fu, H., Subramanian, R. R., and Masters, S. C. (2000). 14-3-3 proteins: structure, function, and regulation. *Annu. Rev. Pharmacol. Toxicol.* 40, 617–647. doi: 10.1146/annurev.pharmtox.40.1.617
- Gan, S., and Amasino, R. M. (1997). Making sense of senescence (molecular genetic regulation and manipulation of leaf senescence). *Plant Physiol.* 113, 313–319.
- Gonzalez-Hernandez, J., Sarath, G., Stein, J., Owens, V., Gedy, K., and Boe, A. (2009). A multiple species approach to biomass production from native herbaceous perennial feedstocks. *In Vitro Cell. Dev. Biol. Plant* 45, 267–281. doi: 10.1007/s11627-009-9215-9
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytosome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186. doi: 10.1093/nar/gkr944
- Grant, J. J., Yun, B. W., and Loake, G. J. (2000). Oxidative burst and cognate redox signalling reported by luciferase imaging: identification of a signal network that functions independently of ethylene, SA and Me-JA but is dependent on MAPKK activity. *Plant J.* 24, 569–582. doi: 10.1046/j.1365-313x.2000.00902.x
- Grbić, V., and Bleecker, A. B. (1995). Ethylene regulates the timing of leaf senescence in Arabidopsis. *Plant J.* 8, 595–602. doi: 10.1046/j.1365-313X.1995.8040595.x
- Gregersen, P. L., Culetic, A., Boschian, L., and Krupinska, K. (2013). Plant senescence and crop productivity. *Plant Mol. Biol.* 82, 603–622. doi: 10.1007/s11103-013-0013-8
- Gregersen, P. L., and Holm, P. B. (2007). Transcriptome analysis of senescence in the flag leaf of wheat (*Triticum aestivum* L.). *Plant Biotechnol. J.* 5, 192–206. doi: 10.1111/j.1467-7652.2006.00232.x



- Gupta, D., Eldakak, M., Rohila, J. S., and Basu, C. (2014). Biochemical analysis of 'kerosene tree' *Hymenaea courbaril* L. under heat stress. *Plant Signal. Behav.* 9:e972851. doi: 10.4161/15592316.2014.972851
- Haake, V., Zrenner, R., Sonnewald, U., and Stitt, M. (1998). A moderate decrease of plastid aldolase activity inhibits photosynthesis, alters the levels of sugars and starch, and inhibits growth of potato plants. *Plant J.* 14, 147–157. doi: 10.1046/j.1365-313X.1998.00089.x
- Haehnel, W. (1984). Photosynthetic electron transport in higher plants. *Annu. Rev. Plant Physiol.* 35, 659–693. doi: 10.1146/annurev.pp.35.060184.03303
- Hager, M., Hermann, M., Biehler, K., Krieger-Liszka, A., and Bock, R. (2002). Lack of the small plastid-encoded PsbJ polypeptide results in a defective water-splitting apparatus of photosystem II, reduced photosystem I levels, and hypersensitivity to light. *J. Biol. Chem.* 277, 14031–14039. doi: 10.1074/jbc.M112053200
- Hayashi, G., Moro, C. F., Rohila, J. S., Shibato, J., Kubo, A., Imanaka, T., et al. (2015). 2D-DIGE-based proteome expression changes in leaves of rice seedlings exposed to low-level gamma radiation at Iitate village, Fukushima. *Plant Signal. Behav.* 10:e1103406. doi: 10.1080/15592324.2015.1103406
- Hurkman, W. J., and Tanaka, C. K. (1986). Solubilization of plant membrane proteins for analysis by two-dimensional gel electrophoresis. *Plant Physiol.* 81, 802–806. doi: 10.1104/pp.81.3.802
- Ifuku, K., Yamamoto, Y., Ono, T.-A., Ishihara, S., and Sato, F. (2005). PsbP protein, but not PsbQ protein, is essential for the regulation and stabilization of photosystem II in higher plants. *Plant Physiol.* 139, 1175–1184. doi: 10.1104/pp.105.068643
- Jones, R., Ougham, H., Thomas, H., and Waaland, S. (2012). *The Molecular Life of Plants*. Hoboken, NJ: John Wiley & Sons.
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27
- Khanna-Chopra, R. (2012). Leaf senescence and abiotic stresses share reactive oxygen species-mediated chloroplast degradation. *Protoplasma* 249, 469–481. doi: 10.1007/s00709-011-0308-z
- Kim, S. T., Kim, S. G., Agrawal, G. K., Kikuchi, S., and Rakwal, R. (2014). Rice proteomics: a model system for crop improvement and food security. *Proteomics* 14, 593–610. doi: 10.1002/pmic.201300388
- Komatsu, S., Mock, H.-P., Yang, P., and Svensson, B. (2013). Application of proteomics for improving crop protection/artificial regulation. *Front. Plant Sci.* 4:522. doi: 10.3389/fpls.2013.00522
- Kura-Hotta, M., Satoh, K., and Katoh, S. (1987). Relationship between photosynthesis and chlorophyll content during leaf senescence of rice seedlings. *Plant Cell Physiol.* 28, 1321–1329.
- Kwon, S. I., Cho, H. J., Bae, K., Jung, J. H., Jin, H. C., and Park, O. K. (2009). Role of an Arabidopsis Rab GTPase RabG3b in pathogen response and leaf senescence. *J. Plant Biol.* 52, 79–87. doi: 10.1007/s12374-009-9011-4
- Lee, D., and Boe, A. (2005). Biomass production of switchgrass in central South Dakota. *Crop Sci.* 45, 2583–2590. doi: 10.2135/cropsci2005.04-0003
- Lefebvre, S., Lawson, T., Fryer, M., Zakhleniuk, O. V., Lloyd, J. C., and Raines, C. A. (2005). Increased sedoheptulose-1, 7-bisphosphatase activity in transgenic tobacco plants stimulates photosynthesis and growth from an early stage in development. *Plant Physiol.* 138, 451–460. doi: 10.1104/pp.104.055046
- Lemoine, R., La Camera, S., Atanassova, R., Dédaldéchamp, F., Allario, T., Pourtau, N., et al. (2013). Source-to-sink transport of sugar and regulation by environmental factors. *Front. Plant Sci.* 4:272. doi: 10.3389/fpls.2013.00272
- Lichtenthaler, H. K. (1999). The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. *Annu. Rev. Plant Biol.* 50, 47–65. doi: 10.1146/annurev.arplant.50.1.47
- Lim, P. O., Kim, H. J., and Gil Nam, H. (2007). Leaf senescence. *Annu. Rev. Plant Biol.* 58, 115–136. doi: 10.1146/annurev.arplant.57.032905.105316
- Lim, P. O., Woo, H. R., and Nam, H. G. (2003). Molecular genetics of leaf senescence in Arabidopsis. *Trends Plant Sci.* 8, 272–278. doi: 10.1016/S1360-1385(03)00103-1
- Liu, M.-S., Li, H.-C., Lai, Y.-M., Lo, H.-F., and Chen, L.-F. O. (2013). Proteomics and transcriptomics of broccoli subjected to exogenously supplied and transgenic senescence-induced cytokinin for amelioration of postharvest yellowing. *J. Proteomics* 93, 133–144. doi: 10.1016/j.jprot.2013.05.014
- Liu, X.-L., Yu, H.-D., Guan, Y., Li, J.-K., and Guo, F.-Q. (2012). Carbonylation and loss-of-function analyses of SBPase reveal its metabolic interface role in oxidative stress, carbon assimilation, and multiple aspects of growth and development in Arabidopsis. *Mol. Plant* 5, 1082–1099. doi: 10.1093/mp/sss012
- Loreto, F., Mannozi, M., Maris, C., Nascetti, P., Ferranti, F., and Pasqualini, S. (2001). Ozone quenching properties of isoprene and its antioxidant role in leaves. *Plant Physiol.* 126, 993–1000. doi: 10.1104/pp.126.3.993
- Lundin, B., Hansson, M., Schoefs, B., Vener, A. V., and Spetea, C. (2007). The Arabidopsis PsbO2 protein regulates dephosphorylation and turnover of the photosystem II reaction center D1 protein. *Plant J.* 49, 528–539. doi: 10.1111/j.1365-313X.2006.02976.x
- Lukaszewicz, M., Matysiak-Kata, I., Aksamit, A., Oszmiański, J., and Szopa, J. (2002). 14-3-3 Protein regulation of the antioxidant capacity of transgenic potato tubers. *Plant Sci.* 163, 125–130. doi: 10.1016/S0168-9452(02)00081-X
- Moeder, W., Del Pozo, O., Navarre, D. A., Martin, G. B., and Klessig, D. F. (2007). Aconitase plays a role in regulating resistance to oxidative stress and cell death in Arabidopsis and Nicotiana benthamiana. *Plant Mol. Biol.* 63, 273–287. doi: 10.1007/s11103-006-9087-x
- Moore, B., Zhou, L., Rolland, F., Hall, Q., Cheng, W.-H., Liu, Y.-X., et al. (2003). Role of the Arabidopsis glucose sensor HXK1 in nutrient, light, and hormonal signaling. *Science* 300, 332–336. doi: 10.1126/science.1080585
- Noodén, L. D., Guimét, J. J., and John, I. (1997). Senescence mechanisms. *Physiol. Plant.* 101, 746–753. doi: 10.1111/j.1399-3054.1997.tb01059.x
- Palmer, N. A., Donze-Reiner, T., Horvath, D., Heng-Moss, T., Waters, B., Tobias, C., et al. (2015). Switchgrass (*Panicum virgatum* L.) flag leaf transcriptomes reveal molecular signatures of leaf development, senescence, and mineral dynamics. *Funct. Integr. Genomics* 15, 1–16. doi: 10.1007/s10142-014-0393-0
- Palmer, N. A., Saathoff, A. J., Waters, B. M., Donze, T., Heng-Moss, T. M., Twigg, P., et al. (2014). Global changes in mineral transporters in tetraploid switchgrasses (*Panicum virgatum* L.). *Front. Plant Sci.* 4:549. doi: 10.3389/fpls.2013.00549
- Peterson, G. L. (1983). Determination of total protein. *Methods Enzymol.* 91, 95–119. doi: 10.1016/S0076-6879(83)91014-5
- Prochazkova, D., Sairam, R., Srivastava, G., and Singh, D. (2001). Oxidative stress and antioxidant activity as the basis of senescence in maize leaves. *Plant Sci.* 161, 765–771. doi: 10.1016/S0168-9452(01)00462-9
- Rhee, S. Y., Beavis, W., Berardini, T. Z., Chen, G., Dixon, D., Doyle, A., et al. (2003). The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. *Nucleic Acids Res.* 31, 224–228. doi: 10.1093/nar/gkg076
- Rinerson, C., Scully, E., Palmer, N., Donze-Reiner, T., Rabara, R., Tripathi, P., et al. (2015). The WRKY transcription factor family and senescence in switchgrass. *BMC Genomics* 16:912. doi: 10.1186/s12864-015-2057-4
- Robbins, M. L., Roy, A., Wang, P.-H., Gaffoor, I., Sekhon, R. S., De O Buanafina, M. M., et al. (2013). Comparative proteomics analysis by DIGE and iTRAQ provides insight into the regulation of phenylpropanoids in maize. *J. Proteomics* 93, 254–275. doi: 10.1016/j.jprot.2013.06.018
- Sarath, G., Baird, L. M., and Mitchell, R. B. (2014). Senescence, dormancy and tillering in perennial C4 grasses. *Plant Sci.* 217–218, 140–151. doi: 10.1016/j.plantsci.2013.12.012
- Sarath, G., Mitchell, R. B., Sattler, S. E., Funnell, D., Pedersen, J. F., Graybosch, R. A., et al. (2008). Opportunities and roadblocks in utilizing forages and small grains for liquid fuels. *J. Ind. Microbiol. Biotechnol.* 35, 343–354. doi: 10.1007/s10295-007-0296-3
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Spreitzer, R. J., and Salvucci, M. E. (2002). Rubisco: structure, regulatory interactions, and possibilities for a better enzyme. *Annu. Rev. Plant Biol.* 53, 449–475. doi: 10.1146/annurev.arplant.53.100301.135233
- Suzuki, Y., Makino, A., and Mae, T. (2001). Changes in the turnover of Rubisco and levels of mRNAs of rbcL and rbcS in rice leaves from emergence to senescence. *Plant Cell Environ.* 24, 1353–1360. doi: 10.1046/j.0016-8025.2001.00789.x
- Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguéz, P., et al. (2011). The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* 39, D561–D568. doi: 10.1093/nar/gkq973



- Tamoi, M., Nagaoka, M., Miyagawa, Y., and Shigeoka, S. (2006). Contribution of fructose-1,6-bisphosphatase and sedoheptulose-1,7-bisphosphatase to the photosynthetic rate and carbon flow in the Calvin cycle in transgenic plants. *Plant Cell Physiol.* 47, 380–390. doi: 10.1093/pcp/pcj004
- The UniProt Consortium (2008). The universal protein resource (UniProt). *Nucleic Acids Res.* 36(Suppl. 1), D190–D195. doi: 10.1093/nar/gkm895
- Turrens, J. F. (2003). Mitochondrial formation of reactive oxygen species. *J. Physiol.* 552, 335–344. doi: 10.1113/jphysiol.2003.049478
- Uematsu, K., Suzuki, N., Iwamae, T., Inui, M., and Yukawa, H. (2012). Increased fructose 1, 6-bisphosphate aldolase in plastids enhances growth and photosynthesis of tobacco plants. *J. Exp. Bot.* 63, 3001–3009. doi: 10.1093/jxb/ers004
- van Hemert, M. J., Steensma, H. Y., and van Heusden, G. P. H. (2001). 14–3–3 proteins: key regulators of cell division, signalling and apoptosis. *Bioessays* 23, 936–946. doi: 10.1002/bies.1134
- Vickers, C. E., Gershenzon, J., Lerdau, M. T., and Loreto, F. (2009). A unified mechanism of action for volatile isoprenoids in plant abiotic stress. *Nat. Chem. Biol.* 5, 283–291. doi: 10.1038/nchembio.158
- Wang, P., Sun, X., Xie, Y., Li, M., Chen, W., Zhang, S., et al. (2014). Melatonin regulates proteomic changes during leaf senescence in *Malus hupehensis*. *J. Pineal Res.* 57, 291–307. doi: 10.1111/jpi.12169
- Wang, Y., Yun, B.-W., Kwon, E., Hong, J. K., Yoon, J., and Loake, G. J. (2006). S-nitrosylation: an emerging redox-based post-translational modification in plants. *J. Exp. Bot.* 57, 1777–1784. doi: 10.1093/jxb/erj211
- Wilczyński, G., Kulma, A., and Szopa, J. (1998). The expression of 14-3-3 isoforms in potato is developmentally regulated. *J. Plant Physiol.* 153, 118–126. doi: 10.1016/S0176-1617(98)80054-0
- Woollard, A. A., and Moore, I. (2008). The functions of Rab GTPases in plant membrane traffic. *Curr. Opin. Plant Biol.* 11, 610–619. doi: 10.1016/j.pbi.2008.09.010
- Yan, J. Q., He, C. X., Wang, J., Mao, Z. H., Holaday, S. A., Allen, R. D., et al. (2004). Overexpression of the Arabidopsis 14-3-3 protein GF14 lambda in cotton leads to a “Stay-Green” phenotype and improves stress tolerance under moderate drought conditions. *Plant Cell Physiol.* 45, 1007–1014. doi: 10.1093/pcp/pch115
- Zhao, Y., and Xu, J. (2000). Mitochondria, reactive oxygen species and apoptosis. *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Jin Zhan.* 28, 168–171.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Paudel, Das, Tran, Boe, Palmer, Sarath, Gonzalez-Hernandez, Rushton and Rohila. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Advantages of publishing in Frontiers



## OPEN ACCESS

Articles are free to read,  
for greatest visibility



## COLLABORATIVE PEER-REVIEW

Designed to be rigorous  
– yet also collaborative,  
fair and constructive



## FAST PUBLICATION

Average 85 days from  
submission to publication  
(across all journals)



## COPYRIGHT TO AUTHORS

No limit to article  
distribution and re-use



## TRANSPARENT

Editors and reviewers  
acknowledged by name  
on published articles



## SUPPORT

By our Swiss-based  
editorial team



## IMPACT METRICS

Advanced metrics  
track your article's impact



## GLOBAL SPREAD

5'100'000+ monthly  
article views  
and downloads



## LOOP RESEARCH NETWORK

Our network  
increases readership  
for your article

## Frontiers

EPFL Innovation Park, Building I • 1015 Lausanne • Switzerland  
Tel +41 21 510 17 00 • Fax +41 21 510 17 01 • [info@frontiersin.org](mailto:info@frontiersin.org)  
[www.frontiersin.org](http://www.frontiersin.org)

## Find us on

