Scientific Research Publishing

Genome-Wide SNPs Identification and Determination of Proteins Associated with Stress Response in Sorghum (*Sorghum bicolor* L. Monech) Accessions

Hind Emad Fadoul[1*], Marmar A. El Siddig[1], Abdel Wahab H. Abdalla[2], Adil A. El Hussein[1]

[1]Department of Botany, Faculty of Science, University of Khartoum, Sudan
[2]Department of Agronomy, Faculty of Agriculture, University of Khartoum, Sudan
Email: *nody_bot@hotmail.com, *adilelhussein@hotmail.com

Abstract

Current efforts in sorghum breeding programs are exploiting genotyping-by-sequencing (GBS) data to provide full-genome scans for desired traits. The aim of this study was to utilize GBS approach for the identification of genomic regions associated with stress response in sorghum (*Sorghum bicolor* L. Monech) accessions. DNA samples of twenty sorghum accessions, having different response to drought, were used to prepare GBS libraries for sequencing. SNPs were called using the TASSELGBS pipeline and the tags that present at least 10 times in the dataset were considered and aligned to the reference genome of *Sorghum bicolor*. The identified SNPs were all compared with the published sorghum transcript related to stress response gene activity. Overall; 94.40% tags were aligned and 69,736 putative SNPs positions were identified. Blast search revealed homology to annotated heat and drought–tolerance associated genes which code for ATPases, Peroxidase, Hydrophobic protein LTI6A, Aquaporin SIP2-1, Aconitate hydratase and phosphatidylinositol-4-phosphate-5-kinase. The phylogeny of the 20 accessions was constructed using the generated SNPs data**.** Phylogenetic analysis data showed that the phenotypically tolerant line **(**El9**)** makes a separate cluster and the same for the accessions HSD8653 and HSD5612 near to the cluster that includes most accessions with known post-flowering drought tolerance (HSD7410, HDS10033, HSD8552, GESHEISH and HSD8849). Post-flowering drought sensitive accessions (Tabat, Wadahmed, HSD6468 and HSD6478) formed a separate cluster while the sensitive accession HSD9959 and the tolerant accessions HSD8511 and HSD9566 were distributed between the two clusters. Thus, cluster analysis confirmed the variation among accessions in post-flowering

drought tolerance. With further validation, these markers may be used for marker assisted selection for breeding new sorghum genotypes with stress adaptation.

Keywords

SNPs, Sorghum, Stress Response, GBS

1. Introduction

Sorghum is the dietary staple food for about 500 million people over 30 countries and is especially important in semi-arid regions of Africa, along with pearl millet and cassava [1]. It is believed that sorghum yields and/or further resistance to biotic and abiotic factors could be improved upon using suitable improved varieties, and preferably developed from environmentally suited local landraces and/or wild/weedy varieties [2]. Important abiotic stresses in sorghum are broadly defined into three categories: water, temperature, and nutrition. Although sorghum possesses excellent drought resistance compared to most other crops, drought stress is the primary factor that reduces sorghum production worldwide [3]. Drought response in sorghum has been classified into two distinct stages; pre-flowering drought response that occurs prior to anthesis and post-flowering drought response that is observed when water limitation occurs during the grain-filling stage. Post-flowering drought stress tolerance is indicated when plants remain green and fill grain normally [4].

Marker-assisted selection was employed to improve the stay green trait involved in the drought tolerance of sorghum [5]. Recent advances in sorghum genomics, including availability of an aligned sorghum genome sequence [6], utilizing newer marker techniques such as DArT, and single-nucleotide polymorphisms [7], have strengthened the foundation for better integration of molecular marker technologies in applied sorghum breeding programs [8].

GBS is a simple highly multiplexed system for constructing reduced representation libraries for the Illumina NGS platform developed in the Buckler lab [7]. It generates large numbers of SNPs for use in genetic analyses and genotyping [9].

Current efforts in sorghum are exploiting genotyping-by-sequencing data to provide full-genome scans across >100,000 SNP (single-nucleotide polymorphism) loci in each member of a portion of a global reference collection of sorghum germplasm that has been phenotyped with a lysimetric system to explore the allelic variation available for key components of the drought resistance phenotype.

The objective of the present study was to identify significant SNPs of genomic regions associated with stress response in 20 sorghum accessions collected from Sudan, using GBS.

2. Materials and Methods

2.1. Plant Materials

Seeds of twenty sorghum accessions, for this study, were obtained from Agricultural Plant Genetic Resources Conservation and Research Center, Agricultural Research Corporation (ARC), Wad Medani, Sudan, as well as from Department of Agronomy, Faculty of Agriculture, University of Khartoum, Sudan. The accessions were classified by Agricultural Research Corporation into three groups (tolerant, intermediate and sensitive) depending on morphological characters previously measured in the field (Table 1).

2.2. DNA Extraction

Genomic DNA was isolated from leaf tissues using CTAB method [10]. The DNA was eluted in 50 μl Nuclease-Free Water. The samples were quantified using Qubit Fluorometer and Nanodrop. The quality of samples was checked in 0.8% Agarose gel electrophoresis.

2.3. Preparation of GBS Libraries

To 100 ng of gDNA, barcoded adapters were added and digested with *Ape*k1 enzyme for 2 h at 75°C. The barcoded and common adapters were ligated to the sticky ends of the digested DNA with T4 DNA Ligase enzyme at 22°C for 60 mins, followed by heat inactivation. The ligated products were pooled into two groups of 10 samples each. The pool of ligated products was size-selected to 700 bp - 1 KB on 2% gel. The pool was PCR amplified to generate the final library pool.

2.4. Quantity and Quality Check (QC) of Library on Bioanalyzer

The library pool was analyzed in Bioanalyzer 2100 (Agilent Technologies) using High Sensitivity (HS) DNA chip as per manufacturer's instructions. The library pool was sequenced on illumina NextSeq platform.

Table 1. Sorghum accession used in this study.

Sorghum accession	Post flowering drought tolerance	Sorghum accession	Post flowering drought tolerance
HSD5612	Intermediate tolerance	TABAT	Sensitive
HSD7410	Tolerant	E19	Tolerant
HSD8511	Tolerant	GESHEISH	Intermediate tolerance
HSD6478	Sensitive	WADAHMED	Sensitive
HSD9566	Tolerant	HSD8266	Tolerant
HSD9959	Sensitive	HSD10033	Tolerant
HSD8653	Tolerant	HSD8163	Tolerant
HSD6468	Sensitive	HSD8552	Intermediate tolerance
HSD8228	Tolerant	HSD8849	Tolerant
HSD8150	Intermediate tolerance	HSD7507	Tolerant

2.5. Data Analysis

SNPs were called from illumina FASTq files using the TASSELGBS pipeline. Only 75 bp tags present at least 10 times in the dataset were considered. Reads were aligned to the reference genome by Bowtie 2 tool with only the best hit alignment used. SNPs with >95% missing data were discarded. SNPs were filtered by minor allele frequency, as rare SNPs are especially useful for inferring differences between accessions after aligned with reference genome.

2.6. Identification of Stress Response-Associated SNPs

To identify stress response genes, the published stress response related genes from a reference sorghum plant were used. A total of 587 genes were obtained from the published sorghum transcriptome [11]. The 587 genes positions were compared with the obtained SNPs positions belonging to the 20 sorghum samples to identify SNPs positions related to stress response genes.

2.7. Phylogenetic Analysis

The accessions data were used for phylogenetic analysis using the cladogram function in TASSELGBS.

3. Results and Discussion

The gDNA extracted and purified from the 20 sorghum accessions was digested using ApeK1 enzyme to prepare the GBS library; the library pool was sequenced independently on Illumina Nextseq platform. The next generation sequencing for the accessions' samples was performed using 2×150 bp chemistry; the reads statistics for the generated data is shown in Table 2. A total of 61,656,784 raw reads were retained. These reads were not uniform among different accessions; the highest number of reads (4,486,288) was recorded for HSD8163 while the least number (965,338) was recorded for HSD10033 (Table 2).

Table 2. Summary of sequenced raw reads recorded for each Sorghum accession.

Accession	Raw Reads	Accession	Raw Reads
HSD5612	4,198,996	TABAT	3,436,844
HSD7410	1,687,323	E19	3,191,451
HSD8511	4,320,193	GESHEISH	1,777,488
HSD6478	1,870,329	WADAHMED	2,571,771
HSD9566	3,605,439	HSD8266	4,342,633
HSD9959	3,851,744	HSD10033	965,338
HSD8653	3,522,793	HSD8163	4,065,312
HSD6468	4,264,344	HSD8552	1,564,602
HSD8228	3,966,240	HSD8849	1,607,044
HSD8150	4,486,288	HSD7507	2,360,612
Total reads	35,773,689	Total reads	25,883,095

Three millions quality reads data per sample were generated; the TASSEL-GBS pipeline was used for identifying and calling SNPs from next generation sequenced accessions with a reference genome, as a result; a total of 814859 distinct tags were identified for all 20 sorghum accessions (Table 3). Overall; 94.40% tags were aligned to the reference genome of *Sorghum bicolor* (http://www.ncbi.nlm.nih.gov/genome/?term=sorghum) using Bowtie 2 tool, 716,829 cut positions were determined. SNPs were called from the aligned tags, resulted in 69,736 SNPs positions in all 20 accessions. A total of 587 genes associated with stress response were obtained from the published sorghum transcriptome [11].

Stress Response Gene and Associated SNPs

Blast search using the flanking sequence against NCBI databases against the reference genome B35 (BT × 623) was considered significant and revealed homology to annotated genes with function in stress adaptation. This genomic regions is found to be coding for the AAA+ ATPases, Peroxidase, Hydrophobic protein LTI6A, Aquaporin SIP2-1, Aconitate hydratase and phosphatidylinositol-4-phosphate-5-kinase.

According to [12] such a spatial arrangement suggests a significant role for linkage group A in both drought stress tolerance and yield in sorghum. Linkage group A harbored many important genes encoding key photosynthetic enzymes, heat shock proteins, cell membrane ATPase and ABA-responsive genes. The SNPs identified in this study refers to a gene has a similar sequence to the AT-Pases, Peroxidase, Hydrophobic protein LTI6A, Aquaporin SIP2-1, Aconitate hydratase and phosphatidylinositol-4-phosphate-5-kinase. The plasma membrane proton pumps is an enzymes whose activity is altered significantly in response to a number of factors in the environment. Besides regulation of growth and development processes, the plasma membrane proton pump also plays a role in plastic adaptation of plants to changing conditions, especially conditions of stress. Adaptation is a complex process. Some of the modifications in plants subjected to abiotic stress are indicated to be adaptive. Physiological modifications caused by environmental stress and allowing continued plant functions are ascribed by plant physiologists as being adaptive [13]. GBS approach has also been used by Thurber *et al.* (2015) to detect three genomic regions necessary for temperate adaptation across 1160 sorghum conversion lines, containing the Dw1, Dw2, and Dw3 loci on chromosomes 9, 6, and 7, respectively. Also, [14]

Table 3. Summary of tags alignment against *Sorghum bicolor* (sorghum) Genome.

Total Tags	814,859 (100%)
Aligned tags > 1 times	194,203 (23.83%)
Aligned exactly 1 time	566,877 (69.57%)
Aligned 0 times	53,779 (6.60%)
Overall alignment rate	93.40%

have identified genomic regions involved in tolerance to drought stress and drought stress induced leaf senescence in juvenile barley. They found that the major QTLs for drought stress and leaf senescence were located on chromosome 5H and 2H. Blast X search for associated marker sequences revealed that respective SNPs are in some cases located in proteins related to drought stress or leaf senescence.

To investigate genetic relationships among these accessions, a phylogenic analysis using the marker data generated and developed by GBS was performed (Figure 1). Generally, the tolerant line El9 makes a separate cluster. HSD8653 and HSD5612 were found to be near to the cluster which includes most accessions with post-flowering drought tolerance (HSD7410, HDS10033, HSD8552, GESHEISH and HSD8849). Post-flowering drought sensitive accessions (Tabat, Wadahmed, HSD6468 and HSD6478) formed a separate cluster while the sensitive accession HSD9959 and the tolerant accessions HSD8511 and HSD9566 were distributed between the two clusters. Thus, cluster analysis suggested the presence of great variation among accessions in post-flowering drought tolerance.

The population structure and genome-wide linkage disequilibrium in 478 spring wheat cultivars from 17 populations across the United States and Mexico, have been studied by [15] using 1536 SNPs, 9 clusters were identified, concluding that the previously inferred populations share a common genetic identity. Also, [16] analyzed 194 accessions of sorghum using Diversity Array Technology (DArT) markers, the clustering of the segregating populations reflects the genetic relationships among the parental lines with regard to the variations panned by the diversity panel.
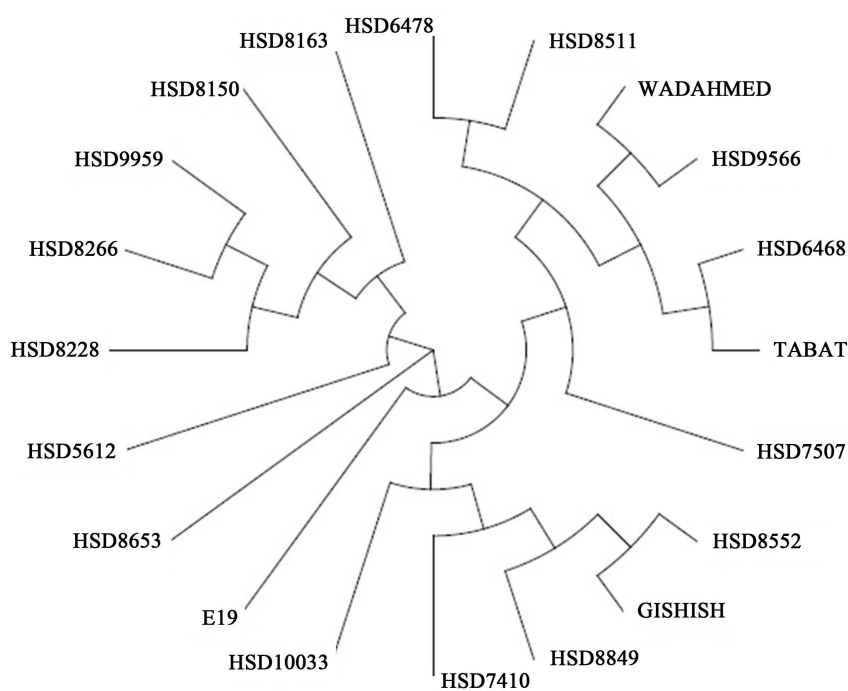

Figure 1. Phylogenetic analysis of sorghum accessions based on GBS results.

Blast search using the flanking sequence revealed homology to annotated gene with function in stress adaptation. With further validation, these markers may be used for marker assisted selection for breeding new sorghum genotypes with stress adaptation.

Acknowledgements

References

[1] ICRISAT (International Crops Research Institute for the Semi-Arid Tropics) (2010). http://www.icrisat.org/crop-sorghum.htm

[2] Makanda, I., Tongoona, P., Derera, J., Sibiya, J. and Fato, P. (2010) Combining Ability and Cultivar Superiority of Sorghum Germplasm for Grain Yield across Tropical Low- and Mid-Altitude Environments. *Field Crops Research*, **116**, 75-85. https://doi.org/10.1016/j.fcr.2009.11.015

[3] Blum, A. (2005) Drought Resistance, Water-Use Efficiency, and Yield Potential—Are They Compatible, Dissonant, or Mutually Exclusive? *Crop Pasture Science*, **56**, 1159-1168. https://doi.org/10.1071/AR05069

[4] Rosenow, D.T. and Clark, L.E. (1995) Drought and Lodging Resistance for a Quality Sorghum Crop. *Proceedings of the* 50*th Annual Corn and Sorghum Industry Research Conference*, Chicago, IL, 82-97.

[5] Harris, K., Subudhi, P.K., Borrell, A., Jordan, D., Rosenow, D., Nguyen, H., Klein, P. and Mullet, J. (2007) Sorghum Stay-Green QTL Individually Reduce Post-Flowering Drought-Induced Leaf Senescence. *Journal of Experimental Botany*, **58**, 327-338. https://doi.org/10.1093/jxb/erl225

[6] Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T. and Poliakov, A. (2009) The Sorghum Bicolor Genome and the Diversification of Grasses. *Nature*, **457**, 551-556. https://doi.org/10.1038/nature07723

[7] Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.C., Kawamoto, K., Buckler, E.S., *et al.* (2010) A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE*, **6**, e19379. https://doi.org/10.1371/journal.pone.0019379

[8] Mace, E.S. and Jordan, D.R. (2010) Location of Major Effect Genes in Sorghum (*Sorghum bicolor* (L.) Moench). *Theoretical and Applied Genetics*, **121**, 1339-1356. https://doi.org/10.1007/s00122-010-1392-8

[9] Beissinger, T.M., Hirsch, C.N., Sekhon, R.S., Foerster, J.M., Johnson, M. and Muttoni, G., *et al.* (2013) Marker Density and Read Depth for Genotyping Populations Using Genotyping-by-Sequencing. *Genetics*, **193**, 1073-1081. https://doi.org/10.1534/genetics.112.147710

[10] Hoisington, D., Khirallah, M. and Gonzalez-de-Leon, D. (1994) Laboratory Protocols: CIMMYT Applied Molecular Genetics Laboratory. 2nd Edition, CIMMYT, Mexico.

[11] Johnson, M., Lim, F., Finkler, A., Fromm, H., Slabas, R.A. and Knight, M.R. (2014) Transcriptomic Analysis of *Sorghum bicolor* Responding to Combined Heat and Drought Stress. *BMC Genomics*, **15**, 456. https://doi.org/10.1186/1471-2164-15-456

[12] Lebreton, L., Lazic-Jancic, V., Steed, A., Pekic, S. and Quarrie, A.S. (1995) Identification of QTL for Drought Response in Maize and Their Use in Testing Casual Re-

lationships between Traits. *Journal of Experimental Botany*, **46**, 853-865.
https://doi.org/10.1093/jxb/46.7.853

[13] Janicka-Russak, M. (2011) Plant Plasma Membrane H$^+$-ATPase in Adaptation of Plants to Abiotic Stresses. In: Shanker, A., Ed., *Abiotic Stress Response in Plants—Physiological, Biochemical and Genetic Perspectives*, InTech, 197-218.
http://www.intechopen.com/books/abioticstress-response-in-plants-physiological-biochemi-
cal-and-genetic-perspectives/plant-plasma-membrane-hatpase-in-adaptation-of-pla
nts-to-abiotic-stresses

[14] Wehner, G., Balko, C., Enders, M., Humbeck, K. and Ordon, F. (2015) Identification of Genomic Regions Involved in Tolerance to Drought Stress and Drought Stress Induced Leaf Senescence in Juvenile Barley. *Plant Biology*, **15**, 125.

[15] Chao, S., Dubcovsky, J., Dvorak, J., Luo, M., Baenziger, S.P., Matnyazov, R., Clark, D., Talbert, L.E., Anderson, A., Dreisigacker, S., Glover, K., Chen, J., Campbell, K., Bruckner, P.L., Rudd, J.C., Haley, S., Carver, F.B., Perry, S., Sorrells, M.E. and Akhunov, E.D. (2010) Population- and Genome-Specific Patterns of Linkage Disequilibrium and SNP Variation in Spring and Winter Wheat (*Triticum aestivum* L.). *BMC Genomics*, **11**, 727. https://doi.org/10.1186/1471-2164-11-727

[16] Fiedler, K., Bekele, W.A., Friedt, W., Snowdon, R., Steutzel, H., Zacharias, A. and Uptmoor, R. (2012) Genetic Dissection of the Temperature Dependent Emergence Processes in Sorghum Using Accumulative Emergence Model and Stability Parameters. *Theoretical Applied Genetics*, **125**, 1647-1661.
https://doi.org/10.1007/s00122-012-1941-4