Project Title: Global analysis of epigenetic regulation of gene expression in response to drought stress in Sorghum.

Project Director: A.S.N. Reddy, Professor, Dept. of Biology, Colorado State University, Fort Collins. Email: reddy@olostate.edu
Co-PD: Asa Ben-Hur, Associate Professor, Department of Computer Sciences, CSU, Fort Collins. Email: asa@cs.colostate.edu

Abiotic stresses including drought are major limiting factors of crop yields and cause significant crop losses. Acquisition of stress tolerance to abiotic stresses requires coordinated regulation of a multitude of biochemical and physiological changes, and most of these changes depend on alterations in gene expression. The goal of this work is to perform global analysis of differential regulation of gene expression and alternative splicing, and their relationship with chromatin landscape in drought sensitive and tolerant cultivars.

Analysis of the sorghum transcriptome using single molecule long reads

In our initial analysis of sorghum database for alternative splicing (AS) and splice isoforms we found that it was not well annotated with very few known splice variants per gene. Effective analysis of AS in sorghum required improved gene models. To address this, we first sequenced the sorghum transcriptome using single molecule real time long-read isoform sequencing (Iso-Seq) that permits accurate prediction of splice variants. This is a new technology with very limited tools for the analysis of such data, especially due to the high error rate of the resulting reads. We developed a new pipeline called TAPIS (Transcriptome Analysis Pipeline for Isoform Sequencing) that predicts splice isoforms and uses the reference genome and Illumina reads (when available) for error correction. Our Iso-seq analysis revealed transcriptome-wide full-length isoforms at an unprecedented scale with over 11,000 novel splice isoforms. Additionally, we uncovered alternative polyadenylation sites in ~11,000 expressed genes and more than 2,100 novel genes. These results were published recently (Abdel-Ghany et al, Nature Communications, 2016). These sorghum gene and splice isoform models based on Iso-seq are being used in all our subsequent analysis.

Global analysis of gene expression and alternative splicing in sorghum cultivars To identify drought-tolerant and -susceptible cultivars of sorghum we screened seven sorghum varieties, which include the sequenced BTx623 cultivar. Based on the growth phenotype, we selected two drought-tolerant and two drought-susceptible cultivars. To investigate the differential regulation of gene expression and alternative splicing in response to drought we have performed a comprehensive analysis of differential expression and differential alternative splicing (DAS) in sorghum in drought-resistant and drought-sensitive cultivars. Two biological replicates across two time points (1h and 6h) were used for each line. Depending on the line, up to 1300 genes were differentially expressed in response to drought treatment. Each cultivar had a unique set of up- and down-regulated genes in response to drought treatment at both points. There was also a common set of genes that were differentially expressed in response to treatment in all four lines at both time points. We then performed differential alternative splicing analysis using Miso and iDiffIR (http://combi.cs.colostate.edu/idiffir; a manuscript is in preparation), a new tool that we developed to identify statistically significant differences in intron retention and exon skipping across conditions. This analysis identified differentially expressed genes and splicing events that are correlated with the drought-resistant phenotype (Abdel-Ghany et al., manuscript preparation)

Drought and other abiotic stresses regulate alternative polyadenylation in sorghum

A genome-wide alternative polyadenylation (APA) and gene expression analysis in sorghum seedlings subjected to several abiotic stresses has revealed that in response to different abiotic

stresses, the relative levels of mRNA isoforms derived from APA within protein-coding regions and 5´UTRs increased significantly, isoforms derived from polyadenylation within introns increased slightly, whereas, the levels of isoforms with poly (A) sites within 3´UTRs decreased significantly. A novel sequence element was found to be associated with stress-induced APA within introns. Poly (A) sites mapped to unannotated regions of the genome increased significantly in response to stresses and majority of these sites mapped to new unannotated transcriptional units. Interestingly, prolonged drought led to a significant down-regulation of several core histone genes, suggesting the importance of chromatin reorganization in abiotic stress responses in sorghum. Altogether, these results uncovered a widespread role for APA in regulating abiotic stress responses in this important cereal and bioenergy crop. A manuscript describing these results is in review (Chakrabarti et al., Genome Research, in review)

Genome-wide changes in histone modifications in response to drought stress

To identify DNA regions in the genome that are marked with activation and repression marks on histones in control and treated lines we performed chromatin immunoprecipitation (ChIP) using antibodies to histone modifications. Replicates from 4 selected lines were grown under control and drought conditions, fixed with formaldehyde and nuclei were isolated for chromatin preparation. We then performed chromatin immunoprecipitation using antibodies to activation (H3K9Ac and H3K4me3) and repressive (H3K27me3) marks of histones as well as histone 3 with each sample. A mock ChiP was performed with preimmune serum. Analysis of thess data showed a strong correlation between drought-induced genes and histone activation marks.

Association between intron retention and chromatin accessibility

Intron retention (IR) is the most prevalent form of alternative splicing in plants. IR, like other forms of alternative splicing, has an important role in increasing gene product diversity and regulating transcript functionality. Splicing is known to occur co-transcriptionally and is influenced by the speed of transcription, which in turn, is affected by chromatin structure. While it is well-established that promoter regions are highly accessible and are over-represented with DNase I Hypersensitive Sites (DHSs), not much is known about DHSs in the bodies of genes, and their relationship to splicing in general, and IR in particular. We use DNase I-seq and histone activation to investigate the relationship between IR and chromatin structure. We found that IR events are highly enriched in open chromatin. The more open chromatin in IR can also be the result of regulation mediated by DNA-binding proteins. To test this, we performed an exhaustive search for footprints left by DNA-binding proteins that are associated with IR. We identified several hundred short sequence elements that exhibit footprints in their DNase I-seq coverage, the telltale sign for binding events of a regulatory protein, protecting its binding site from cleavage by DNase I. A highly significant fraction of those sequence elements are conserved across monocots and dicots, a strong indication of their functional importance. Our studies have established an association between IR and chromatin accessibility and identified conserved sequence elements for DNA-binding proteins that affect splicing (Ullah et al., BMC Genomics, in revision).

In summary, our Iso-Seq study revealed transcriptome-wide full-length isoforms at an unprecedented scale with over 11000 novel splice isoforms. Additionally, we uncovered alternative polyadenylation sites of ~11000 expressed genes and many novel genes. Overall, Iso-Seq results greatly enhanced sorghum gene annotations that are not only useful in analyzing all our RNA-seq, ChIP-seq and ATAC-seq data but also serve as a great resource to the plant biology community. Our studies identified differentially expressed genes and splicing events that are correlated with the drought-resistant phenotype. An association between alternative splicing and chromatin accessibility was also revealed. Several computational tools developed here

(TAPIS and iDiffIR) have been made freely available to the research community in analyzing alternative splicing and differential alternative splicing.

Deliverables

Publications:

Salah E. Abdel-Ghany*, Michael Hamilton*, Jennifer L. Jacobi, Peter Ngam, Nicholas Devitt, Faye Schilkey, Asa Ben-Hur, Anireddy S.N. Reddy (2016) A survey of the sorghum transcriptome using single-molecule long reads. (*joint first authors) **Nature Communications** 7:11706 | DOI: 10.1038/ncomms11706.

D. Xing, Y.Wang, Michael Hamilton, Asa Ben-Hur and A.S.N. Reddy (2015) Transcriptome-wide identification of RNA targets of Arabidopsis serine/arginine protein 45 (SR45) uncovers the unexpected roles of this RNA binding protein in RNA processing. **Plant Cell** 27, 3294-3308.

Saiprasad G. Palusa and A.S.N. Reddy (2015) Differential Recruitment of Splice Variants from *SR* Pre-mRNAs to Polysomes During Development and in Response to Stresses. **Plant and Cell Physiology**. 56: 421-427.

Reddy, A.S.N., Marquez, Y., Kalyna, M. and Barta, A. (2013) Complexity of the alternative splicing in plants. **Plant Cell** 25: 3657-3683. **Featured in "IN BRIEF" section of Plant Cell.**

Manohar Chakrabarti, Laura de Lorenzo, Salah E. Abdel-Ghany, Anireddy S.N. Reddy, and Arthur G. (2017) **A genome-wide alternative** polyadenylation and gene expression analysis provides molecular insights into the abiotic stress response in sorghum. **(Chakrabarti et al., Genome Research, in review**)

Ullah, F., Hamilton, M., Reddy, A.S. Ben-Hur, A (2017) Exploring the relationship between intron retention and chromatin accessibility in plants. **BMC Genomics (In revision).**

Salah E. Abdel-Ghany, Michael Hamilton, Ullah, F., Hamilton, M., Reddy, A.S. Ben-Hur, A (2017) Transcriptome-wide analysis of gene expression and alternative splicing in drought-sensitive and drought-tolerant sorghum genotypes. **(manuscript in preparation).**

Oral/ Poster Presentations:

Michael Hamilton, Salah E. Abdel-Ghany, Anireddy S.N. Reddy, and Asa Ben-Hur (2016) Alternative and differential polyadenylation detection from single molecule long sequencing. Intelligent Systems for Molecular Biology (ISMB) Integrative RNA Biology, July 8-12th, Orlando, FL.

Fahad Ullah, ASN Reddy, and Asa Ben-Hur (2016) Exploring the relationship between intron retention and DNase I hypersensitivity in plants. Intelligent Systems for Molecular Biology (ISMB) Integrative RNA Biology, July 8-12th, Orlando, FL

Salah E. Abdel-Ghany, Michael Hamilton, Jennifer L Jacobi, Peter Ngam, Nicholas Devitt, Faye Schilkey, Asa Ben-Hur, Anireddy S.N. Reddy (2016) Detection of splice isoforms and alternative polyadenylation using single molecule long reads. Post-transcriptional gene regulation in plants. July 12- 14, Austin, Texas.

Michael Hamilton, A.S.N. Reddy and Asa Ben-Hur. Predicting differential intron retention with iDiffIR. Plant and Animal Genome Conference, 2016

Salah E. Abdel-Ghany, Michael Hamilton, Fahad Ullah, Asa Ben-Hur, Anireddy S.N. Reddy (2016) Global analysis of epigenetic regulation of gene expression in response to drought stress in Sorghum. National Institute of Food and Agriculture and Department of Energy meeting. San Diego, Jan 7, 2016.

Towards a plant splicing code. Presented by Asa Ben-Hur at the 2015 CSU symposium on computational and systems biology.

A survey of the sorghum transcriptome using the single-molecule long read. Invited talk presented by AS.N. Reddy at the Next Generation Sequencing USA congress 27-28th October 2015 at the Joseph B Martin Conference Center at Harvard Medical School, Boston MA

Differential splicing and long read transcriptome assembly using iDiffIR and TAPIS. Presented by Asa Ben-Hur at the 2015 Genomics Science PI meeting, Tysons Virginia.

Biology in the era of big data. Presented by Asa Ben-Hur at the 2015 Rocky Mountain High Performance Computing conference, Boulder Colorado.

Transcriptome-wide analyses of alternative splicing, splice isoforms and alternative polyadenylation in Sorghum using single-molecule long reads. Invited talk presented by AS.N. Reddy at the NextGen Genomics, Biology, Bioinformatics and Technologies (NGBT) Conference, Oct 1-3, India.

A survey of the sorghum transcriptome using the single-molecule long read. Invited talk presented by AS.N. Reddy at the Post-transcriptional gene regulation in plants, July 10-11, Paris, France.

Predicting differential intron retention with iDiffIR. Will be presented by Michael Hamilton at the 2016 PAG ARAPORT workshop.

Community Resources Generated:

Improved annotations of the sorghum genome obtained by analysis of Pacific Biosciences Iso-Seq reads.

Other products/ outcomes:

We developed a pipeline for analysis of Pacific Biosciences Iso-Seq reads called TAPIS, available freely at: https://bitbucket.org/comp_bio/tapis.

We continued the development of our iDiffIR software for prediction of differential intron retention (see http://combi.cs.colostate.edu/idiffir/ and https://bitbucket.org/comp_bio/idiffir).

Training:

Salah E. Abdel-Ghany (Special Assistant Professor, Biology): Performed RNA-seq, ChIP-seq and DNAse-FLASH experiments and validated Iso-seq results. He is also involved some of the data analysis.

Cory Culligan (MS student, Biology - Graduated) Analyzed drought resistance, performed some RNA-seq analysis and RT-PCR studies.

Shea Moore-Farrell (Rotating Ph.D. student, Biology) worked with Salah E. Abdel-Ghany on ChIP-seq experiments.

Kimberly Hallowell (Undergraduate student, Biology) worked with Salah in setting up experiments.

Fahad Ullah (Ph.D. candidate, computer science): analysis of RNA-seq, ChIP-seq, DNAse-seq and DNAse-FLASH data.

Michael Hamilton (Ph.D. candidate, computer science): development of software for analysis of Pacific Biosciences data; development of software for prediction of differential intron retention events.

Collaborations:

Arthur Hunt, Professor, University of Kentucky