Research Plan

Dylan M. D'Agate

Half Hollow Hills East High School

A. Rationale

Global climate change and its effect on increasing drought conditions poses a major challenge to crop production and food security in developing countries. In fact it is estimated that the risk of drought affecting the global food supply could triple by 2040. Novel methods to help farmers combat drought conditions are urgently needed. Genetic engineering offers great potential for combating the effects of drought on crops. Research on gene expression in a single plant species during drought conditions has provided useful information. However, few studies have compared plant gene expression in different plant species during drought conditions using a transcriptomic profiling technology, called RNA-Seq. RNA-Seq is a novel method used to amplify an entire transcriptome to be surveyed in a fast and efficient manner.

The National Center for Biotechnology Information (NCBI) has created an international public library of RNA-Seq data called Gene Expression Omnibus (GEO). Using GEO we will try to obtain the RNA-seq data on two important crops, Sorghum bicolor (cereal crop) and Solanum tuberosum (potato) during drought conditions. Both drought tolerant cereal crop and drought sensitive potato are important crops in developing countries at risk for food insecurity due to drought. The cereal crop is an ideal drought tolerant model for evaluating important genes that may promote drought tolerance. By understanding how genes are differentially expressed in these two plant species during drought, scientists can better understand which genes to target through genetic engineering and help improve plant drought tolerance.

B. Hypothesis

I hypothesize that during drought conditions, drought tolerant plants will express and regulate genes differently as compared to drought sensitive plants.

C. Procedures

We will search studies using the search criteria (Sorghum bicolor, [Organism], Solanum tuberosum [Organism]) AND "high throughput sequencing" AND "Drought") on the leaves of Sorghum bicolor and Solanum tuberosum in an international public repository, National Center for Biotechnology Information's (NCBI) Gene Expression Omnibus (GEO). GEO archives gene expression and other functional genomics data sets creating a "public library" of genomic data. Once the studies are obtained, we will download the raw RNA-Seq data from the leaves of 4 different conditions: Sorghum bicolor (cereal crop) normal, Sorghum bicolor (cereal crop) drought, Solanum tuberosum (potato) normal, Solanum tuberosum (potato) drought. Next we will utilize an open, web-based platform for data intensive biomedical research called Galaxy Toolbox. Galaxy Toolbox consists of multiple programs used to help analyze RNA-seq information. The first program we will use is called Fast QC, a quality control program for RNA-seq reads. Fast QC works by importing raw RNA sequencing data and exporting a report detailing the quality of the data. Once the quality control of the data is completed we will use a program called Trimmomatic to trim off bases that fall below a specified quality threshold.
Next, we will perform a hierarchical indexing for spliced alignment of transcripts (HISAT2). This program efficiently aligns short sequencing reads. The program takes the trimmed sequencing reads from the RNA-seq samples and aligns them, producing a longer, more complete sequence of the original RNA transcripts. The HISAT2 output data will then be used by another program called StringTie which aligns the complete sequence to a control genome and completes the tallying process for each positively identified sequence to generate a gene expression profile for the sample.

Data Analysis

We will use a program called DESeq2 to determine if there are statistically significant differences in gene expression counts between the two samples (Sorghum bicolor (cereal crop) and Solanum tuberosum (potato)) during drought conditions. Next, we will examine if there are differentially expressed genes that are shared between the two plant species. Finally, we will examine the Gene Ontology of both plant species to better understand which gene processes are expressed by each plant species during drought conditions.

Risk and Safety

1. Human Participants in Research – N/A
2. Vertebrate Animals in Research – N/A
3. Potentially Hazardous Biological Agents – N/A
4. Potentially Hazardous Chemicals/Activities/Devices- N/A

D. Bibliography

1. Lobell, D. B. 2012. The influence of climate change on global crop productivity. *Plant Physiol.*160, 1686–1697.

2. Ashraf, M. 2010. Inducing drought tolerance in plants: recent advances. Biotechnology advances, 28(1), 169–83.

3. Azevedo H, Silva-Correia J, Oliveira J, et al. 2015. A strategy for the identification of new abiotic stress determinants in Arabidopsis using web-based data mining and reverse genetics. OMICS.2011;15(12):935–47.

4. Fracasso A, Trindade LM, Amaducci S. 2016. Drought stress tolerance strategies revealed by RNA-Seq in two sorghum genotypes with contrasting WUE. BMC Plant Biol. BMC Plant Biology; 16: 115.

5. Benny J, Pisciotta A, Tiziano,C et al. 2019. Identification of key genes and its chromosome regions linked to drought responses in leaves across different crops through meta-analysis of RNA-Seq data. BMC Plant Biology;194(19):1-18.

NO ADDENDUMS EVER