Bioinformatics Analysis of Switchgrass (Panicum virgatum): An Examination of GEO Datasets to Determine Which Factors Affect Growth and Yield

Gagan Namburi, Ash Stuart (Faculty), & Yan Xu (Faculty)
Ramapo College of New Jersey, Mahwah, NJ, 07430

Introduction

Figure 1

Biofuels are utilized in transportation. They consist of fuels like ethanol and biodiesel which can be made from biomass materials (Biofuels explained). With the increasing need for higher biofuel production, scientists are working to determine which crops are the most suitable to grow in high quantities in order to maximize the potential biofuel synthesis to meet transport needs. A majority of the ethanol used in the United States is extracted from corn (Biofuels explained), but the problem with growing corn is that it requires a great amount of land. Another plant that is being proposed as a biofuel crop is switchgrass, also known as *Panicum virgatum*.

Panicum virgatum (Figure 1) is a versatile plant that can grow in various climates across North America (Fike et al.). In addition, due to the amount of space *Panicum virgatum* takes to grow, it would be ideal to grow larger quantities for biofuel production. One interesting concept is to grow switchgrass so that it can solely be utilized in biofuel production while not taking space allocated for food crops is to grow it in environments which are less suitable for crops such as places with high salinity in their soil or places with drought. A preliminary step to determine if growing switchgrass in less suitable environments is feasible is to examine which genetic factors affect the growth and yield of *Panicum virgatum* and if those factors can be manipulated in order to maximize the potential yield.

Objective

This study focuses on utilizing genetic data from researcher-published studies and examining genetic factors that could potentially influence the growth and yield of *Panicum virgatum*. This data is sourced from NCBI's GEO datasets database (NCBI). Using statistical programming languages, the genetic data is able to be visualized. With a proper understanding of *Panicum virgatum* growth factors, researchers will be one step closer to being able to achieve the long-term goal of being able to mass produce biofuel in areas where plants are not typically grown.

Materials & Methods

GSE55462: This dataset is sourced from a study performed by Rahman H et al. and it examines the differential gene expression of finger millet or *Eleusine coracana* the crop is in its natural environment and when salt stress is present, which was measured with RNA sequencing. This helps identify which genes are over-expressed in the presence of salt. Although *Eleusine coracana* is not the same thing as *Panicum virgatum*, it is similar to *Panicum virgatum* genetically. This study tested two genotypes of *Eleusine coracana*, the salinity stress susceptible CO 12 and the salt tolerant Trichy 1. The data from this study produced four files containing the gene expression data for *Eleusine coracana*, two for each genotype tested with one file displaying gene expression for the control and the other file displaying gene expression for the specimen under salt stress (Rahman et al.).

This already normalized data was processed using the Python3 programming language, with the use of the pandas library and the plotly library run in a Jupyter notebook environment. In total, 55,986 genes were observed in *Eleusine coracana* and each of the four files were loaded into pandas data frames. For each of the two genotypes tested, the control gene expression data, the genotypes of *Eleusine coracana* not exposed to salt, was merged with the salt stress gene expression data of its corresponding genotype. The merge allows for the gene expression difference between the salt stress data and control data to be easily calculated as both gene expression numbers are in the same data frame as according to their matching entries from each genotype of *Eleusine coracana*. The difference was computed between the gene expression of the genes in the salt stressed *Eleusine coracana* and the control *Eleusine coracana*. This data was displayed in a pandas data frame and was filtered out by positive gene expression changes for this study (Figure 2). The final data frame with the feature ID, which contains the gene location in *Eleusine coracana*, and the difference in gene expression were plotted in an interactive scatterplot using plotly.

Results

Both the CO 12 and Trichy 1 genotypes of *Eleusine coracana*, a tetraploid organism, had different numbers of genes that showed a positive change in the presence of salt stress. For the CO 12 genotype, there were 9,642 genes with a positive change in expression after being exposed to salt stress while for the Trichy 1 genotype, there were 10,358 genes with a positive change in expression after being exposed to salt stress. In the interactive scatter plots, there was a small number of genes that shows an intense positive change of gene expression (Figures 3 to 8).
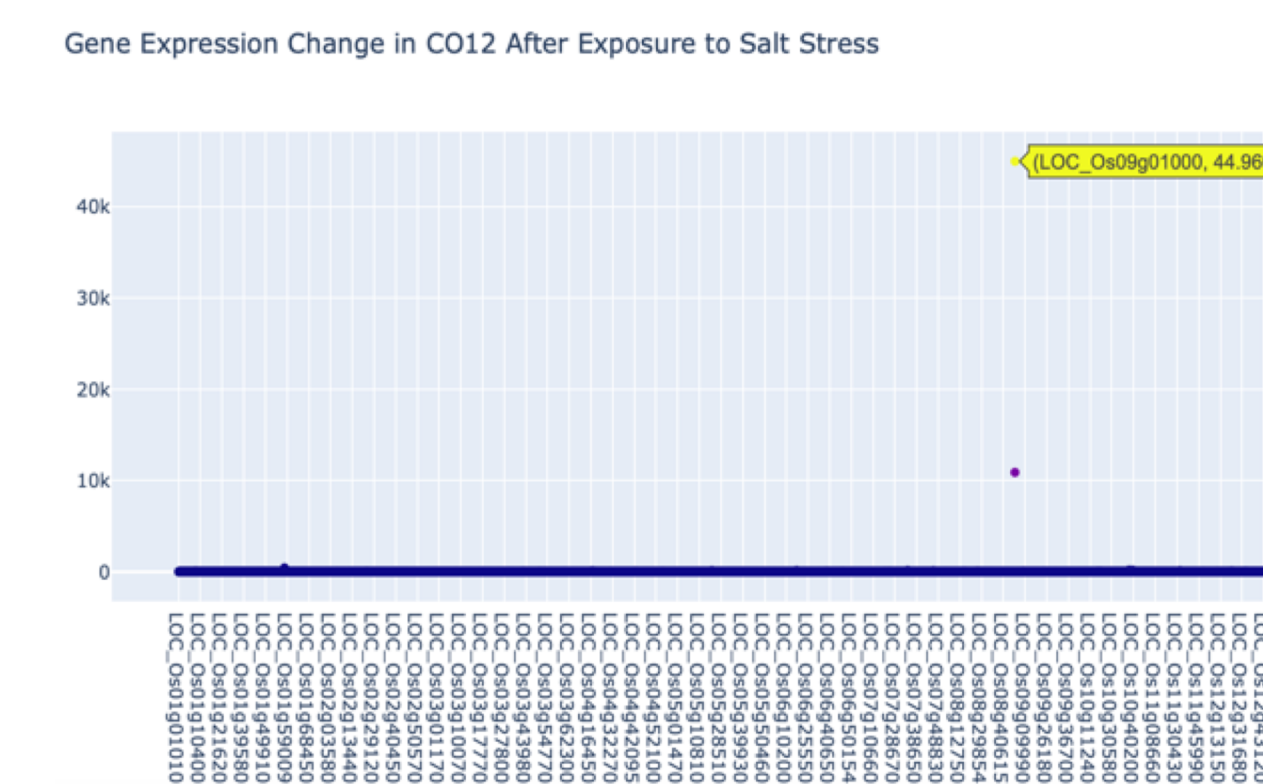
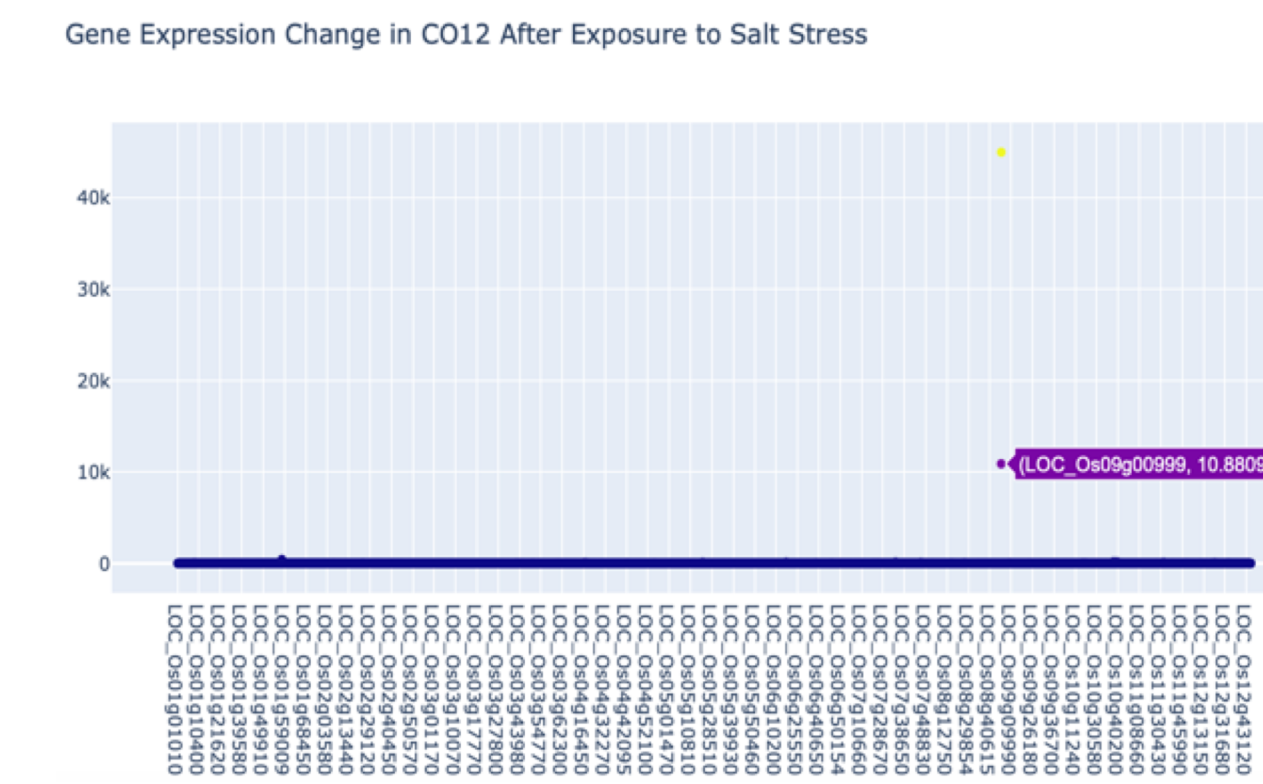**Figure 3.** The LOC_Os09g01000 gene shows an expression change of ~45,000 in CO 12

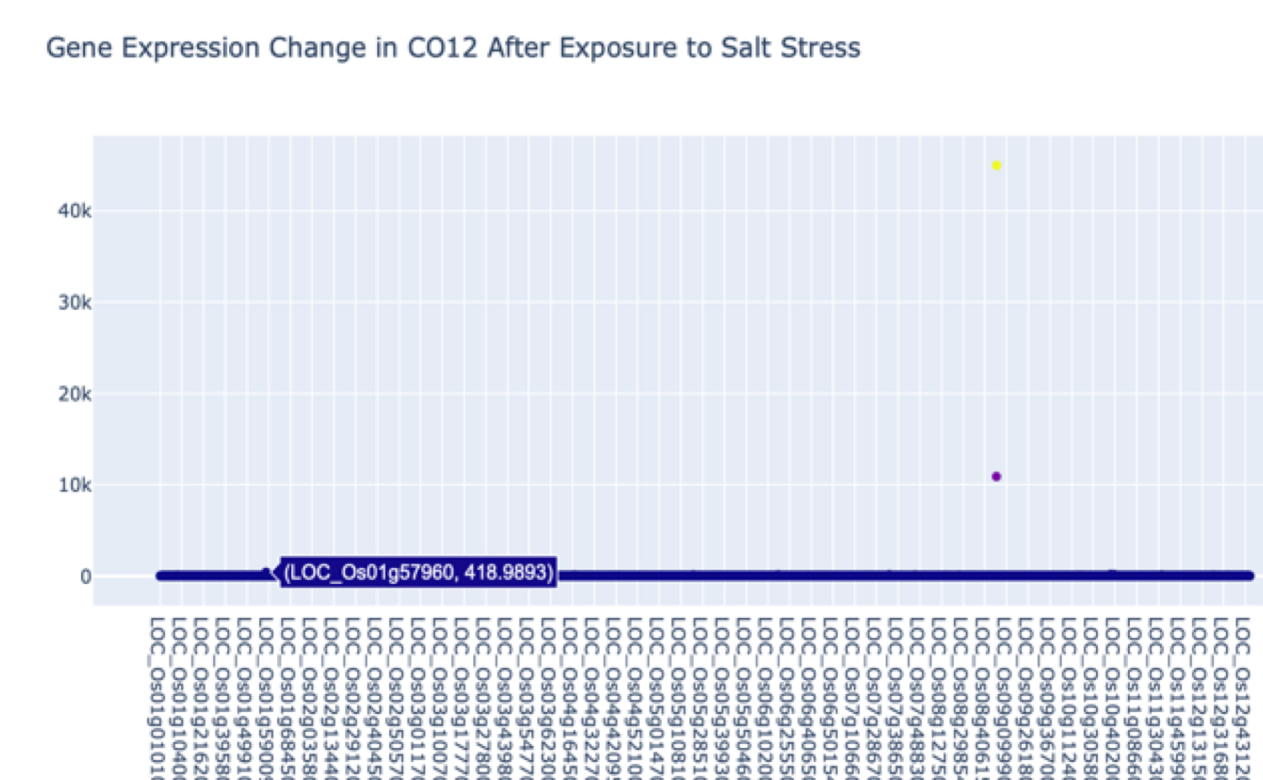**Figure 4.** The LOC_Os09g00999 gene shows an expression change of ~11,000 in CO 12

Figure 5. LOC_Os01g57960 gene shows an expression change of ~419 in CO 12
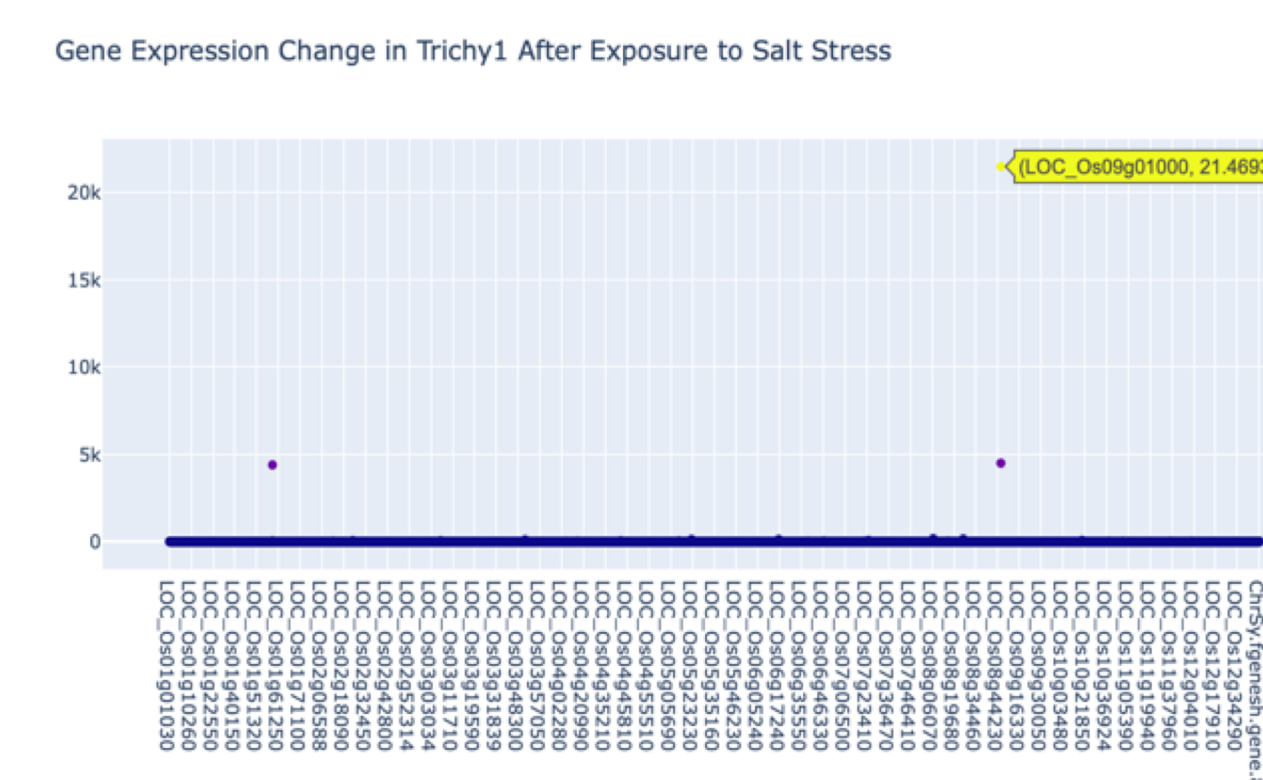
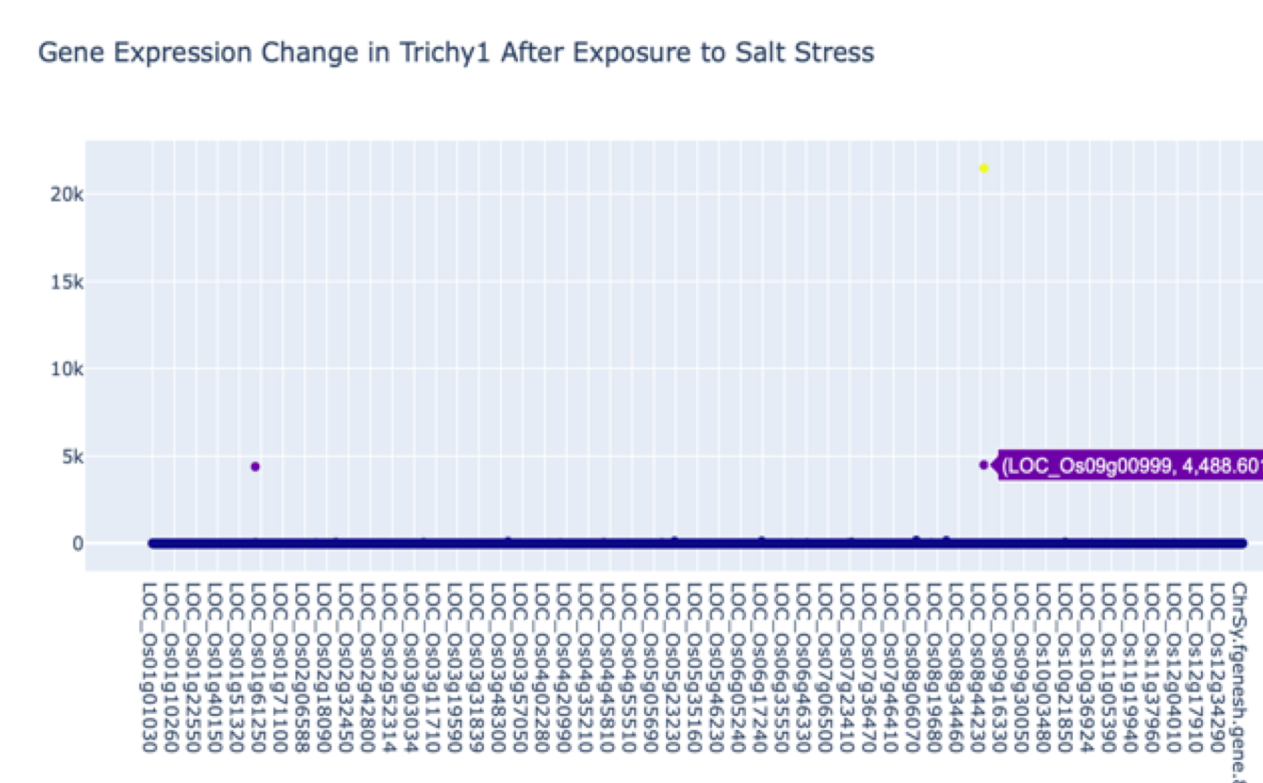**Figure 6.** LOC_Os09g01000 gene shows an expression change of ~21,000 in Trichy 1

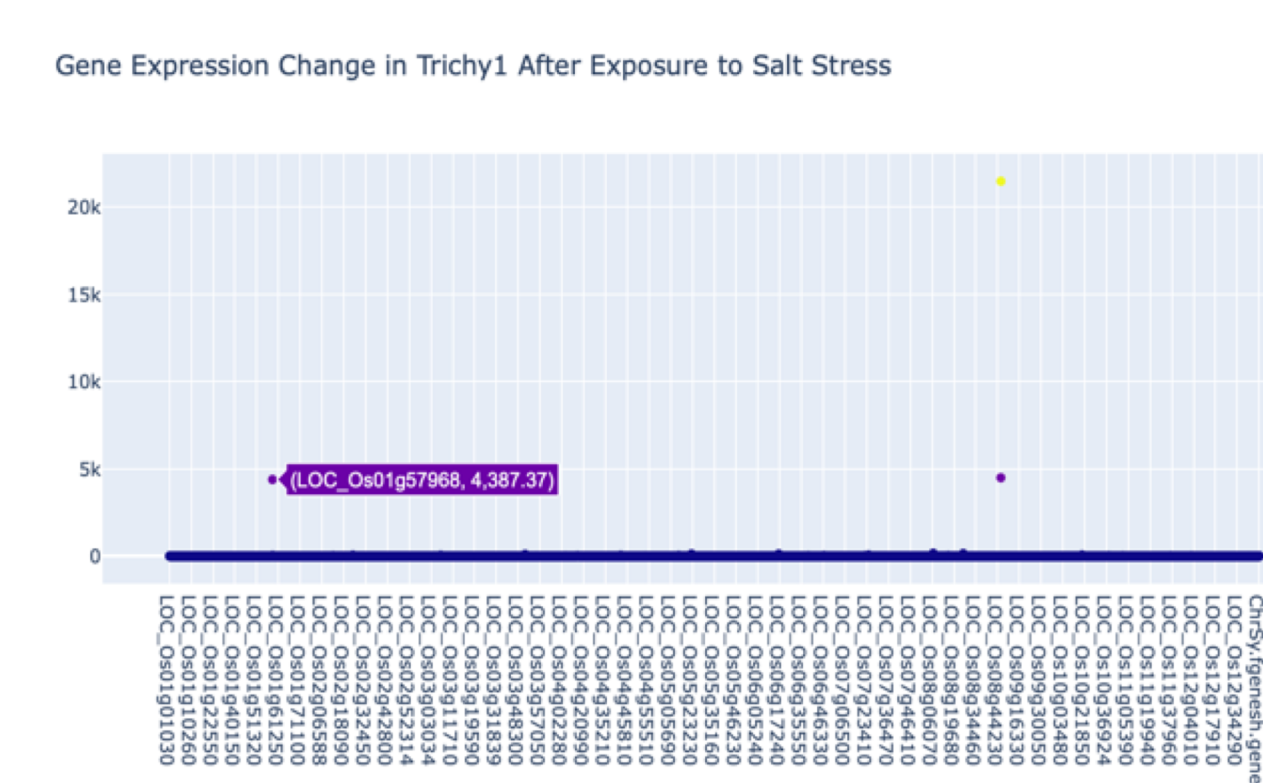**Figure 7.** LOC_Os09g00999 gene shows an expression change of ~4,000 in Trichy 1

Figure 8. LOC_Os01g57968 gene shows an expression change of ~4,000 in Trichy 1

	Feature ID	Change of gene expression with salt stress
0	LOC_Os01g01010	0.166683
3	LOC_Os01g01040	0.135428
4	LOC_Os01g01050	0.132182
5	LOC_Os01g01060	0.914965
7	LOC_Os01g01080	0.195016
...	...	...
55933	ChrSy.fgenesh.gene.37	0.430306
55938	ChrSy.fgenesh.gene.42	0.164287
55945	ChrSy.fgenesh.gene.49	0.016529
55948	ChrSy.fgenesh.gene.52	0.042125
55978	ChrSy.fgenesh.gene.82	0.135876

Figure 2

Discussion

- In CO 12, the LOC_Os09g01000, LOC_Os09g00999, and LOC_Os01g57960 genes showed the highest change in expression and in Trichy 1, the LOC_Os09g01000, LOC_Os09g00999, and LOC_Os01g57968 genes showed the greatest positive change in expression.
- In the two genotypes, both LOC_Os09g01000 and LOC_Os09g00999 were present as genes with high levels of expression changes. This is a good indicator that these genes play an important role in salt stress because they consistently stood out with a high change of gene expression in two different types of *Eleusine coracana* in the presence of salt stress.
- The LOC_Os01g57960 and LOC_Os01g57968 genes are not the same, but are very close in relation to each other.
- It is notable however that the expression levels of LOC_Os01g57968 in CO 12 decreases by about 10,000 after salt stress is present while the expression levels of LOC_Os01g57960 in Trichy 1 decreases by about 100 after salt stress is present.
- The change of the LOC_Os09g01000 gene expression in CO 12 is around 45,000 while the expression level change of the same gene in Trichy 1 is around 21,000. The same trend can be noticed for the LOC_Os09g00999 gene, which has a gene expression difference of approximately 11,000 in CO 12 and 4,000 in Trichy 1.
  - This difference in gene expression can be explained by the fact that the Trichy 1 genotype is more tolerant to salt, unlike the CO 12 variant (Rahman et al.).
  - Nevertheless, this trend demonstrates that the LOC_Os09g01000 and LOC_Os09g00999 genes are viable markers of salt stress.

Conclusion

Through examining genomic data on the gene expression of *Eleusine coracana* in the presence of salt stress, there can be a better understanding on which genetic factors can be manipulated in the genetically similar biofuel crop, *Panicum virgatum* in order to be able to grow it in areas which contain more salinity. This allows for more potential space to be able to grow the biofuel source without intruding on the occupied space that is being utilized by other crops that may be important for things such as food production. As a result, there can be a higher yield of biofuel and the rapidly growing need for biofuel production could be met. The only barrier at the moment is the lack of knowledge and information on the genes being examined. Once those genes are understood, they can be manipulated and applied to *Panicum virgatum* provided that the equivalent genes could be identified and modified in that crop. This solution has the potential to allay the issue of biofuel production for years to come.

References

Fike, John H., et al. "Long-Term Yield Potential of Switchgrass-for-Biofuel Systems." *Biomass and Bioenergy*, vol. 30, no. 3, 2006, pp. 198–206., doi:10.1016/j.biombioe.2005.10.006.

"Home - GEO DataSets - NCBI." *National Center for Biotechnology Information*, U.S. National Library of Medicine, www.ncbi.nlm.nih.gov/gds.

Rahman, H. "GEO Accession Viewer." *National Center for Biotechnology Information*, U.S. National Library of Medicine, 15 Aug. 2014, www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE55462.

Susan.mahr. "Switch Grass, Panicum Virgatum." *Master Gardener Program*, 30 Dec. 2015, wimastergardener.org/article/switch-grass-panicum-virgatum/.

"U.S. Energy Information Administration - EIA - Independent Statistics and Analysis." *Biofuels Explained - U.S. Energy Information Administration (EIA)*, www.eia.gov/energyexplained/biofuels/.