



A GWAS study of color and texture on natto soybeans

Gabrielle Jones, Jade Walker-Wilson, Joshua Winter, Leandro Mozzoni



Abstract

Soybeans are an essential component to the contemporary Japanese diet. Natto is a Japanese dish made by fermenting soybeans, and the physical properties of the unprocessed soybeans will affect the natto produced. Pale yellow soybeans that are slightly firm in texture are preferred. The genetic improvement of these qualities is achieved through breeding. A genome-wide association study (GWAS) is an important tool for soybean breeding and can be used to associate genomes with a certain genotype. Color and texture are the selected phenotypes in this study and a GWAS program was ran to associate the genotype markers with the phenotypes. None of the markers in the study reached the threshold of significance and therefore are not considered to be associated with one of the phenotypes. Further research is necessary.

Introduction

Arkansas's key row crop is soybeans, making up more in acreage than all other crops combined. In 2015, Arkansas farmers harvested 3.1 million acres of soybean crop, valued at \$1.5 billion (Encyclopedia of Arkansas, 2020).

Soybeans have been a staple in the Japanese diet for centuries. Fermented foods such as natto are popular dishes that require soybeans for production. In 2020, the United States exported 53 percent of the world market share in soybeans to Japan (Sasatani, 2021).

The morphological and physiological traits of the soybeans affect the quality of the natto that is ultimately produced (Farnworth, 2003). Pale yellow color and slightly firm texture are preferred (Shih et al., 2009). The genetic improvement of these quantitative traits is an important goal for soybean breeders.

The purpose of this study was to identify the relationship between the alleles and phenotype in the soybean varieties that affect color and texture. A genome-wide association study (GWAS) is an important tool in the field of genetic research. A GWAS software performs a statistical analysis between SNPs to find the specific genomic loci that are linked to the trait of interest (Contreras-Soto et al., 2017; Varshney et al., 2014). An SNP, or single nucleotide polymorphism, is a single nucleotide base change in a DNA sequence ("Polymorphism," n.d.). GWAS is also used for the improvement of desired soybean traits by finding the relationship between phenotypes and genotypes (Contreras-Soto et al., 2017). Essentially, the study searches the soybean genome to find the SNPs related to the quantitative traits: color and texture.

Materials & Methods

This study was conducted by the collection of color data of 101 soybean genotypes and texture data of 105 genotypes. The dried soybeans were prepared by being soaked in water overnight and autoclaved at 121 degrees Celsius for 20 minutes the following day. The color was analyzed using a reader that measured the L* coordinates of the soybean samples. Texture was analyzed using a TA.XT2i Texture Analyzer from Texture Technology Corporation. The blade on the machine tested the shearing force when passed through the soybean sample.

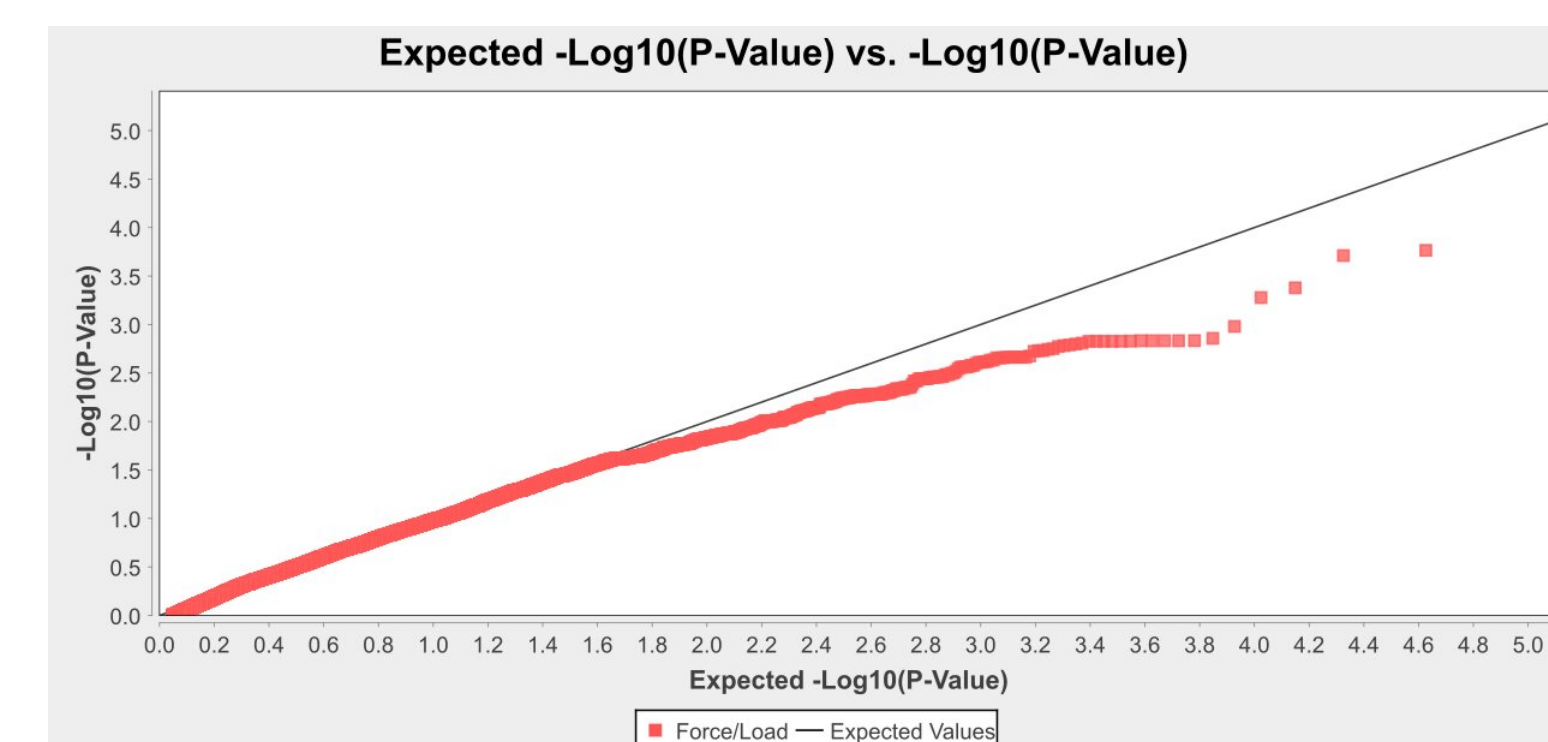


Fig A: Texture analyzer and color reader used in the study.

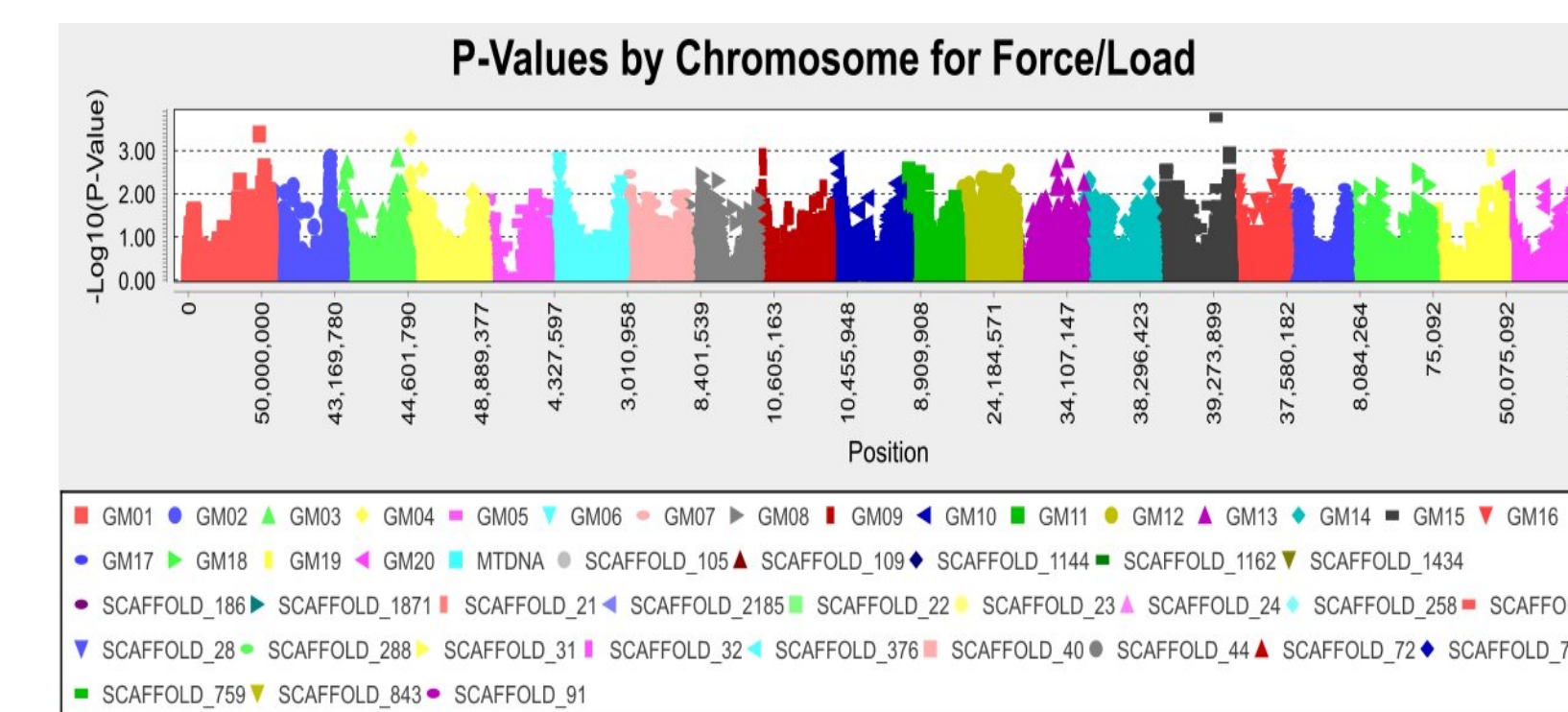
A genome wide association study (GWAS) software, Tassel, was used to analyze the color and texture data. GWAS takes an average of the quantitative trait and the SNP marker. The GWAS was graphed as a general linear model (GLM) and multi-linear model (MLM). The GLM compares phenotype and genotype values. The MLM compares phenotype and genotype values and accounts for kinship.

Results

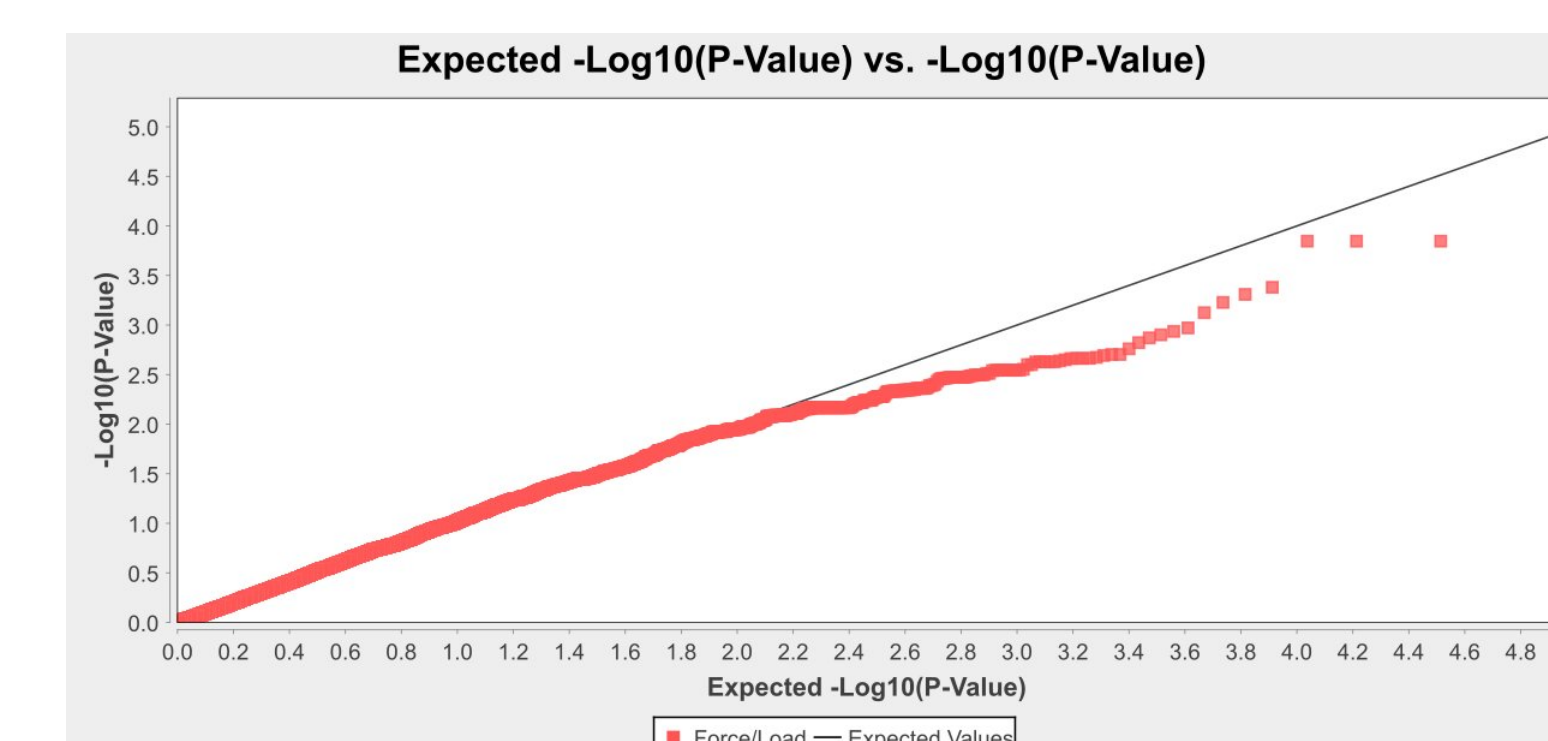
Texture:



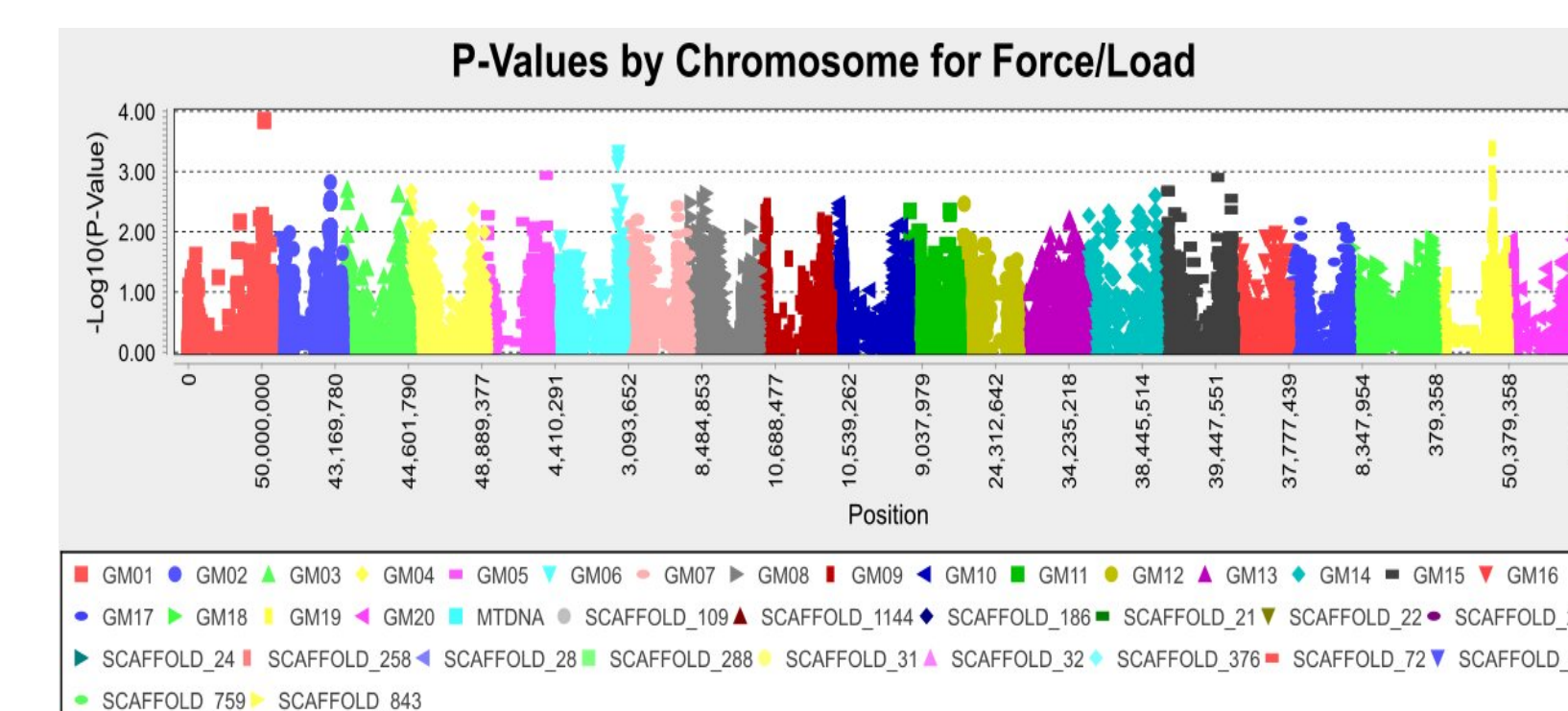
QQ Plot (GLM)



Manhattan Plot (GLM)

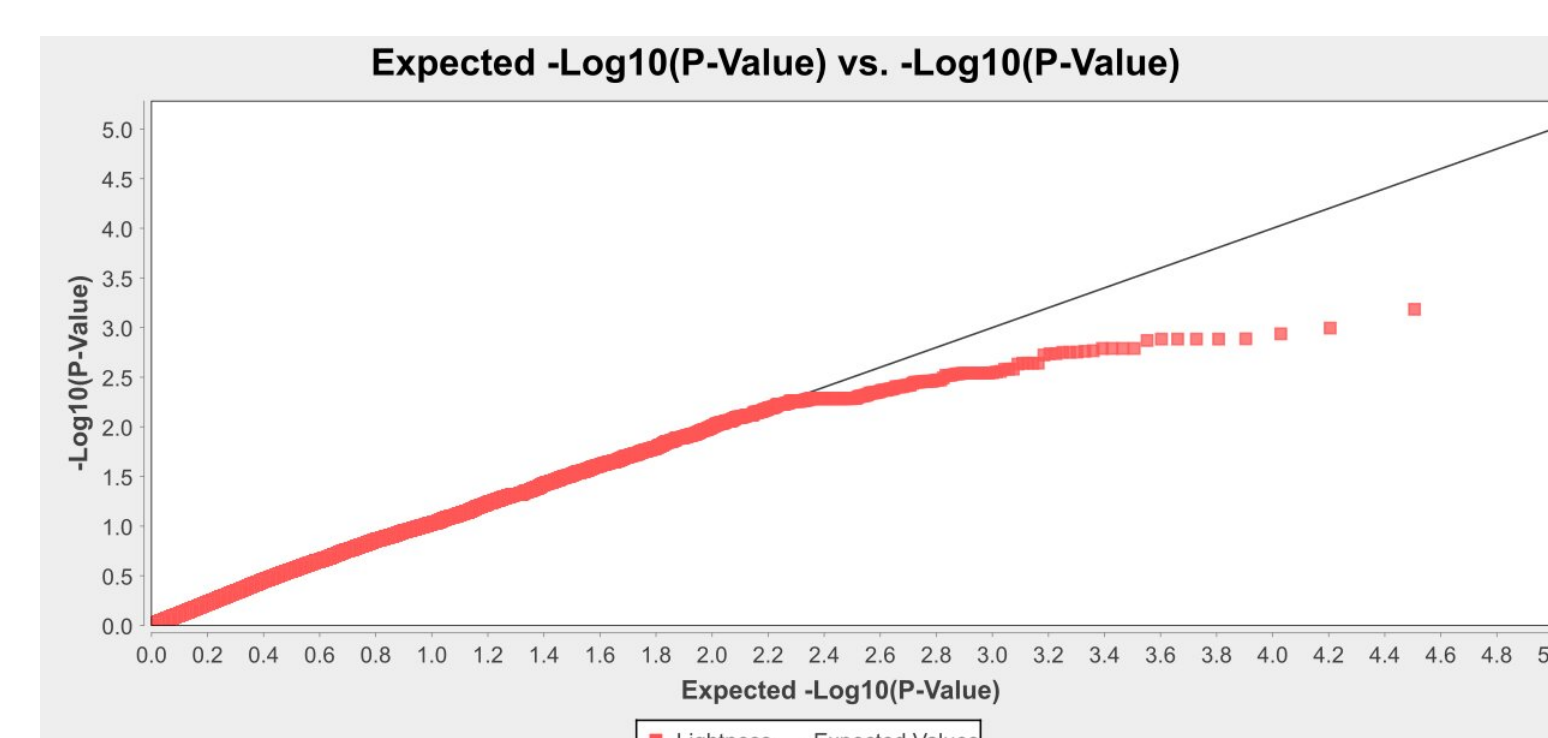


QQ Plot (MLM)

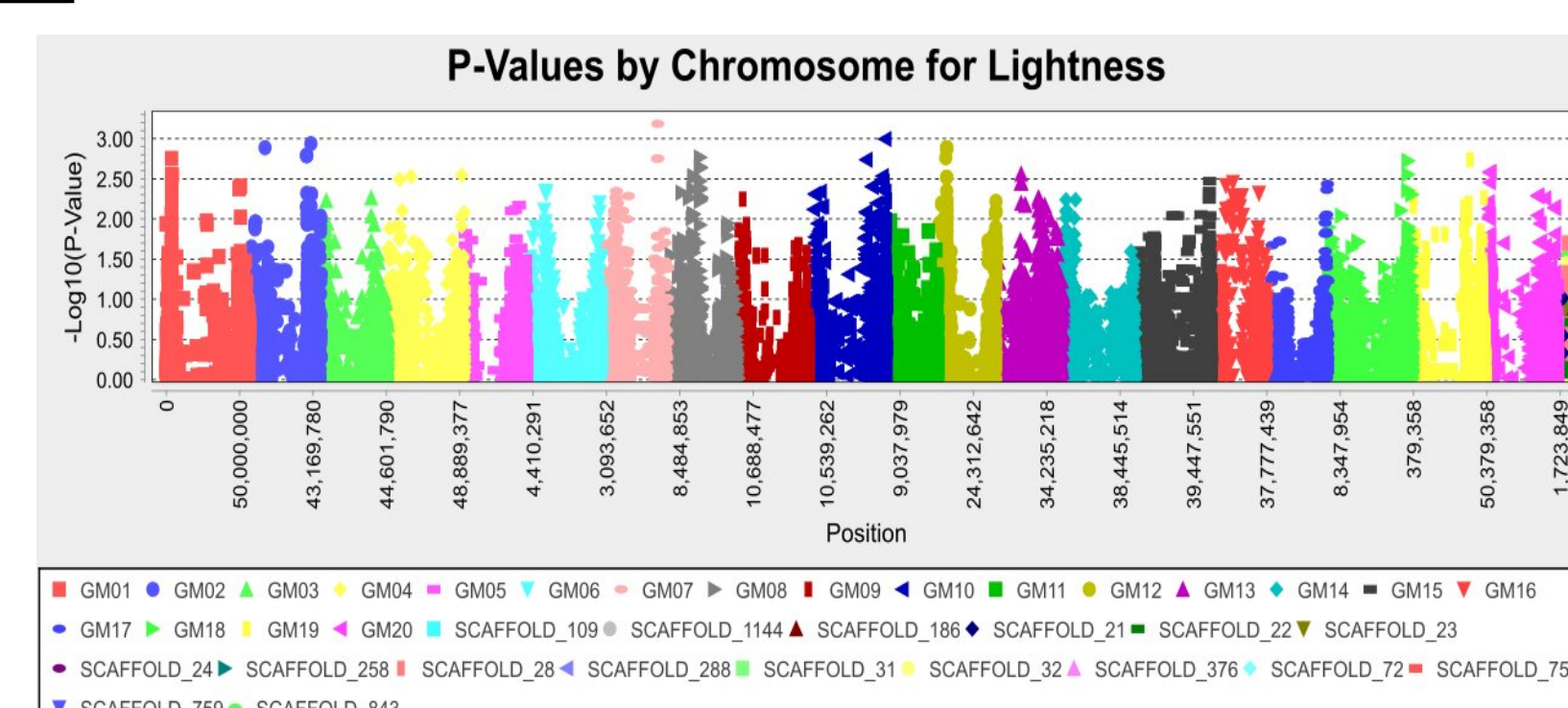


Manhattan Plot (MLM)

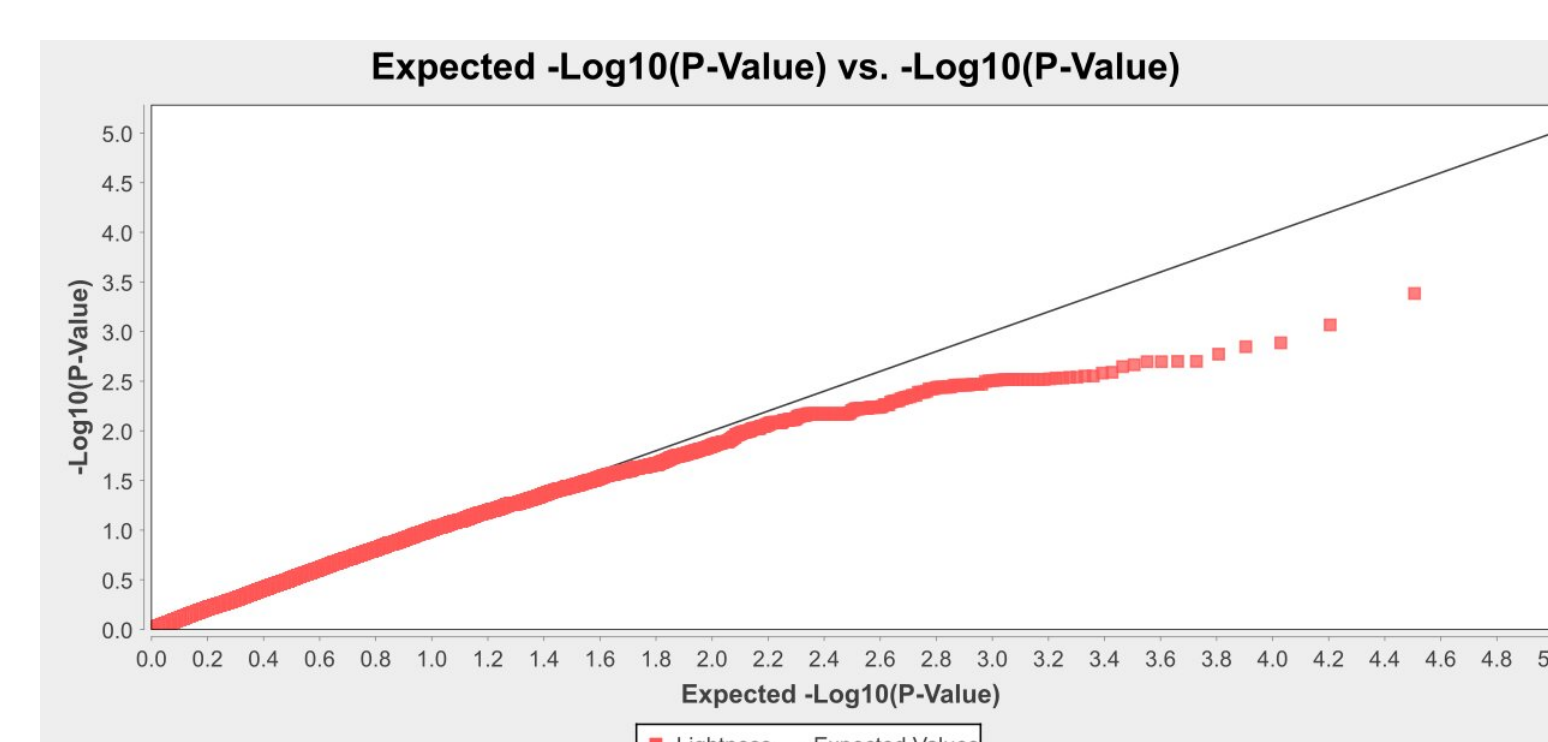
Color:



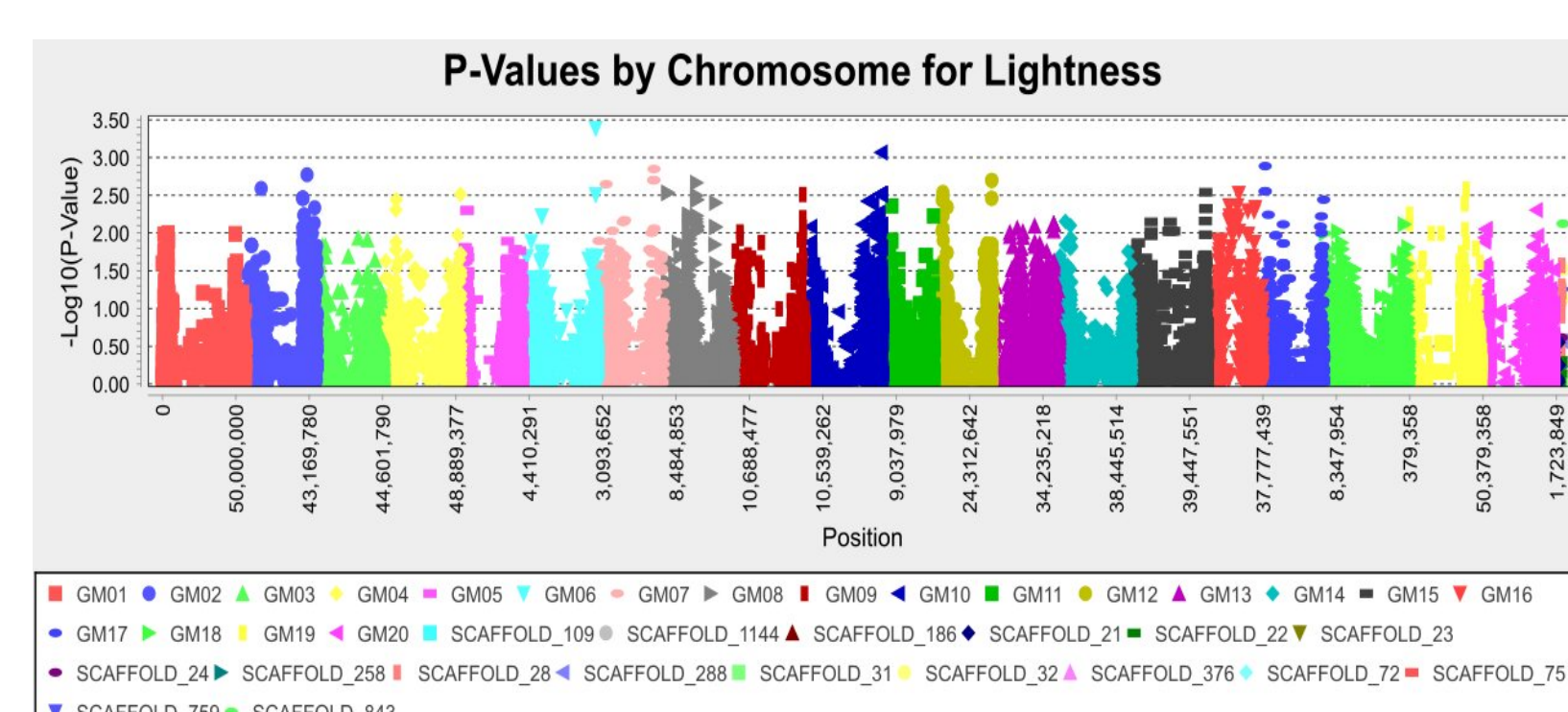
QQ Plot (GLM)



Manhattan Plot (GLM)



QQ Plot (MLM)



Manhattan Plot (MLM)

Texture Results:

The GWAS was graphed as a general linear model (GLM) and multi-linear model (MLM). The GLM compares phenotype and genotype values. The MLM compares phenotype and genotype values and accounts for kinship.

The GLM QQ shows that the data points are normally distributed. The GLM Manhattan plot shows the significance of certain markers in a position on a chromosome. A marker on chromosome 14 is the closest to being significant.

The MLM QQ plot shows that the data is normally distributed. The MLM Manhattan plot shows that none of the markers reached the threshold of significance. A marker on chromosome 1 is the closest to the threshold.

Color Results:

The GWAS was graphed as a GLM and MLM. The GLM QQ plot shows that the data is normally distributed. The GLM Manhattan plot shows that none of the markers reach the threshold of significance. A marker on chromosome 6 is closest to the threshold.

The MLM QQ plot shows that the data is normally distributed. The MLM Manhattan plot shows that none of the markers reach the threshold of significance. Markers on chromosome 6 and chromosome 10 are closest to the threshold.

Discussion & Conclusion

- The aim of this study was to identify markers in the soybean varieties that affect color and texture.
- Only a few markers came close to the threshold of significance, but none of the markers reached the actual threshold.
- Therefore, there were no markers identified in this study that can be attributed to a phenotype.
- Limitations to the study include a small sample size, which weakens the data quality, the GLM plots not accounting for kinship, and GWAS not accounting for linkage disequilibrium.
- Further research is necessary.

References

- Contreras-Soto, R.I., Mora, F., de Oliverira, M.A.R., Higashi, W., Scapim, C.A., & Schuster, I. 2017. A genome-wide association study for agronomic traits in soybean using SNP markers and SNP-based haplotype analysis. *PLoS ONE*, 12 (2): e0171105.
- Encyclopedia of Arkansas. 2020. Arkansas Soybean Promotion Board. <https://encyclopediaofarkansas.net/entries/arkansas-soybean-promotion-board-12286/>
- Farnworth, E.R. 2003. Handbook of Fermented Functional Foods (pp. 228-243). Taylor and Francis Group, LLC.
- National Human Genome Research Institute. n.d.. Polymorphism. <https://www.genome.gov/genetics-glossary/Polymorphism>
- Sasatani, D. 2021. Utilization of Food-Grade Soybeans in Japan. USDA.
- Shih, M.C., Yang, K.T., & Kuo, S.T. 2009. Optimization process of black soybean natto using response surface methodology. *Journal of Food Science*, 74 (6).
- Varshney, R.K., Terauchi, R., & McCouch, S.R. 2014. Harvesting the promising fruits of genomics: applying genome sequencing technologies to crop breeding. *PLoS Biology*, 12(6): e1001883.

Acknowledgements

I would like to express my gratitude to Dr. Mozzoni and the Soybean Breeding Program at the University of Arkansas for allowing me to utilize their materials, equipment, and laboratory spaces to complete this research. This material is based upon work supported by the National Science Foundation under the Research Experiences for Undergraduates (REU) program. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation

