Integrating Transcriptomics, Phenomics, and Ionomics to Study Nitrogen Deficiency Response in Maize

Abstract

Our project utilizes Weighted Gene Co-expression Network Analysis (WGCNA) to examine a combined dataset of RNA sequencing, morphological trait, and ionomics data from maize under nitrogen stress conditions. The primary goal is to use a multi-omics approach to identify gene modules associated with specific traits and to understand the interaction between gene expression patterns and ionomic profiles. Using the WCGNA package in R, we constructed a gene co-expression network, identified several significant modules, correlated these modules with external phenotypic traits, and performed Gene Ontology (GO) enrichment analysis on the identified modules. Our analysis found 24 separate modules, 8 of which differed significantly from the controls. 7 of these significant modules had associated GO terms of interest.

Introduction

Maize is one of the most widely cultivated cereal crops worldwide, with significant importance for both food and bioenergy production (Erenstein et al., 2022). Understanding the inner workings of its growth, yield, and adaptation to different environments is crucial. Here, we explore how to integrate RNAseq, phenomics and ionomics data to provide a comprehensive view of the gene expression, mineral nutrient status and phenomic effects of maize under nitrogen deficiency conditions.

Weighted Gene Co-expression Network Analysis (WGCNA) is a systems biology method for describing the correlation patterns among genes across microarray and RNAseq samples. WGCNA can be used to find clusters (modules) of highly correlated genes, which may be linked to specific traits. This analysis aims to uncover the underlying genetic networks associated with adaptation strategies to nitrogen deficiency, as well as how they relate to measurable changes in the maize seedling phenotype.

Methods

Data Collection

The data supplied for use in this project were collected previously as part of the USDA funded project titled "An Integrated Multi-Omics Approach To Understand Drought, Temperature, And Nutrient Stress Responses". We will be using a subset of the total datasets produced in this project, which includes:

1. RNAseq reads from four maize genotypes with three biological replicates each, subjected to 13 days of low nitrogen conditions or a control treatment.
2. Ionomics data from the same genotypes and conditions, with four biological replicates per genotype.
3. Phenomics data detailing morphological traits calculated from images taken at the experiment's conclusion, with 6-9 replicates per genotype. This includes total plant area (cm^2), plant height (cm) and mean plant hue (°). In this context, hue is a measurement made in degrees that represents a color in the visible electromagnetic spectrum. Hue

degrees of ~90 represent green, healthy leaf tissue, which lower degrees (~80-60) reflect leaf yellowing or loss of green tissue, also known as chlorosis.

Data Preparation

1. RNAseq reads were mapped to the latest version of the maize genome (B73_v5), normalized and a variance stabilizing transformation was applied using the vst() function in the DESeq2 package in R. Genes with variance less than 1 across the dataset were filtered out.
2. The ionomics dataset was log transformed and ions with variance less than 1 were filtered out.
3. The morphological traits dataset was previously filtered for outliers based on a Cook's distance cutoff of 0.5.

The RNAseq and ionomics datasets were merged, and a center and scale transformation across samples was applied to normalize the different types of data. The final merged dataset consisted of a csv file with rows as samples and columns identifying information for each sample including: genotype, treatment, replicate, and each gene count or ion level as a dependent feature.

WGCNA

We used the WGCNA package in R to construct a co-expression network, selecting an appropriate soft-thresholding power of 16 using the pickSoftThreshold() function (Fig. 1) to ensure network scale-free topology. We then ran the blockwiseModules() function to generate the co-expression network.

Module significance and correlation with morphological traits

Module eigengene values were extracted from the co-expression network and used to run linear models to determine if they were significantly different between the low nitrogen and the control treatment. Pearson correlations between all modules and the morphological traits area, height, and hue, were estimated using the module eigengene values.

Module gene ontology enrichment

To understand the genetic basis of the WGCNA clustering, we conducted a gene ontology enrichment analysis using the gprofiler2 package in R, using the false discovery rate correction to adjust term *p* value. Python and Excel were used to isolate the top 4 most enriched GO terms for each of the modules found to be significantly different between treatments. Modules 0 and 12 were omitted given that 0 contained unassigned features, and 12 did not have enriched GO terms in our initial analysis. The term ranking was made using the calculated GeneRatio statistic, as per the methodology found in Kolberg et. al, 2020.

Individual gene differential expression analysis

To further the analysis of the transcriptomic data, we also conducted a differential expression analysis on the individual gene level using the DESeq() function in the DESEq2 R

package. We then examined the relationship between genes found to be differentially expressed and the modules in which the WGCNA analysis placed them.

Results

Co-expression network feature clustering

WGCNA produced a scale-free network of 24 modules, each with varying amounts of features: Module 1 contained 6119 features, while Module 24 contained 31. Module 0 contained features that were not assigned to a specific module in the network. The ionic levels from the ionomics dataset were all clustered between Modules 0 or 1, while the rest of the modules contained gene levels only.

Correlation Between Gene Expression Modules and Morphological Traits

We investigated the relationships between the co-expression modules and the three morphological traits extracted at the end of the experiment. The analysis was visualized in a heatmap of correlation coefficients (Fig. 1) to assess the strength and direction of these relationships. This revealed varied patterns of correlation between the gene modules and the assessed morphological traits. Plant area showed negative correlations with several modules, notably ME6 and ME10 ($r = -0.41$ and -0.45 respectively), suggesting an inverse relationship with the expression profiles within these modules. Plant area was positively correlated with several modules as well, the strongest of these being ME11, ME13 and ME19 ($r = 0.44, 0.55, 0.5$), which shows a moderately strong positive correlation with the genes in these modules.

Plant height exhibited a particularly strong negative correlation with module ME3 ($r = -0.64$), indicating that as gene expression in this module increases, plant height tends to decrease. Another noteworthy negative correlation was observed with module ME24 ($r = -0.47$). Conversely, positive correlations were not as evident, with no module showing a positive correlation with plant height of $r \geq 0.4$. Overall, this trait had the weakest correlations aside from the ME3 negative correlation.

Mean plant hue displayed a pronounced set of correlations. Module ME6 exhibited a significant negative correlation ($r = -0.6$), indicating a link between gene expression within this module and hue variation. This suggests potential involvement of ME6 in the pigmentation or coloration processes of the plants, considering chlorosis is one of the most notable effects of nitrogen deficiency. ME12 also showed a negative correlation ($r = -0.46$). ME15 and ME17 were both positively correlated to a significant degree with mean plant hue ($r = 0.52, 0.53$).

Overall, the heatmap suggests a complex interaction between gene expression modules and plant morphological traits. The significant correlations point towards a potential regulatory role for specific modules in the development and phenotypic expression of these traits. The nature of the relationships, whether direct or indirect, requires further investigation to elucidate the underlying biological mechanisms.

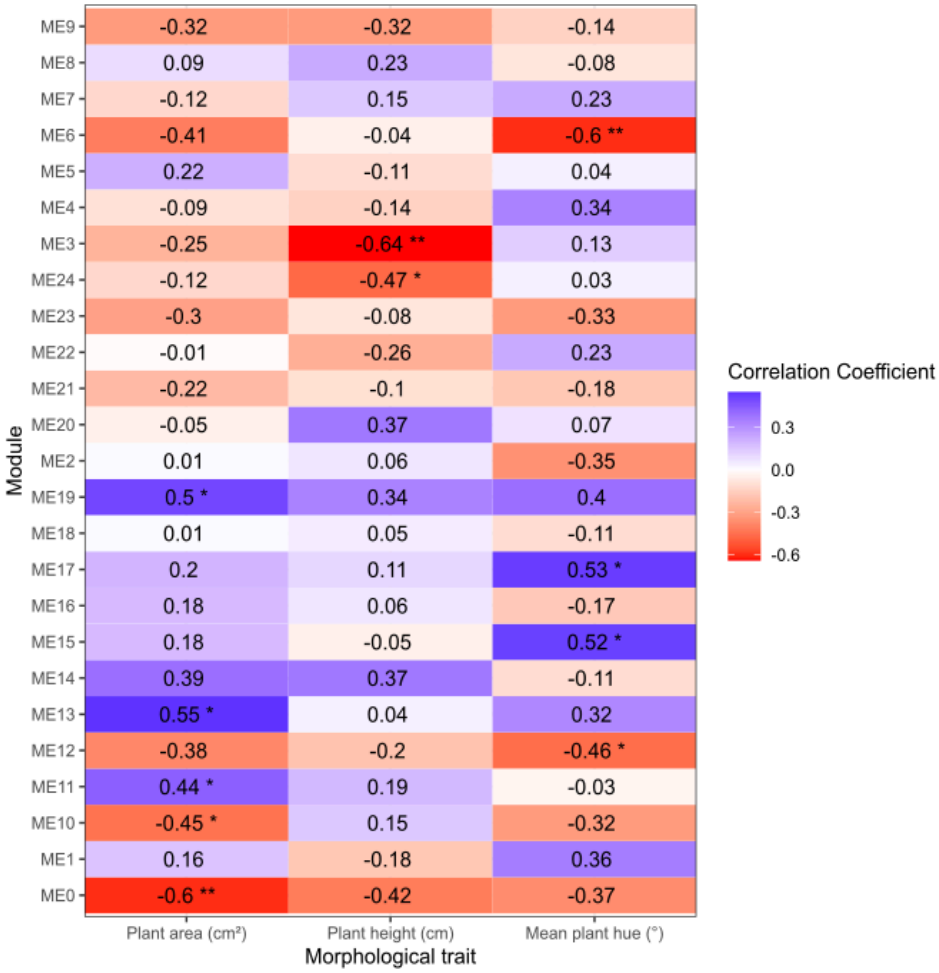

Figure 2: Correlation Heatmap of Gene Modules with Plant Morphological Traits. This heatmap depicts the correlation coefficients between gene expression modules (ME1 to ME24) and morphological traits of plant area (cm²), plant height (cm), and mean plant hue (°). Red to blue gradient represents the range of correlation coefficients from positive (1.0) to negative (-1.0), with asterisks indicating levels of statistical significance ($p < 0.05$, *$p < 0.01$, **$p < 0.001$). Modules and traits with no significant correlation are represented in white.

Differential Module Expression Under Control and Low Nitrogen Conditions

Distinct modules exhibited significant differences in eigengene values when the plants were subjected to the low nitrogen treatment as compared to those grown under control conditions. Eigengenes serve as a composite representation of the gene expression profiles within a module, effectively summarizing the module's overall activity. The boxplots depicted in Figure 3 illustrate the modules with significantly different eigengene values between the two different treatments, pinpointing the impact of nitrogen availability on gene expression. These modules provided the basis for the subsequent gene ontology (GO) enrichment analysis.

Notably, Module ME6 and ME12 showed significant increases in eigengene values under low nitrogen conditions, which could implicate these modules in nitrogen response mechanisms. Module ME10 showed a less pronounced yet significant increase in eigengene value from the control treatment. In contrast, Modules ME13, ME15, ME17 and ME19 all demonstrated a notable decrease in eigengene values under low nitrogen conditions. Module 1 had a non-significant reduction in eigengene value.

These differential expressions highlight the complexity of the plant's transcriptional response to nitrogen levels, indicating that specific modules are either upregulated or

downregulated to adapt to nutrient stress. The observed variations underscore the potential regulatory functions these modules may have in maintaining nitrogen homeostasis and plant development under varying nutrient conditions.

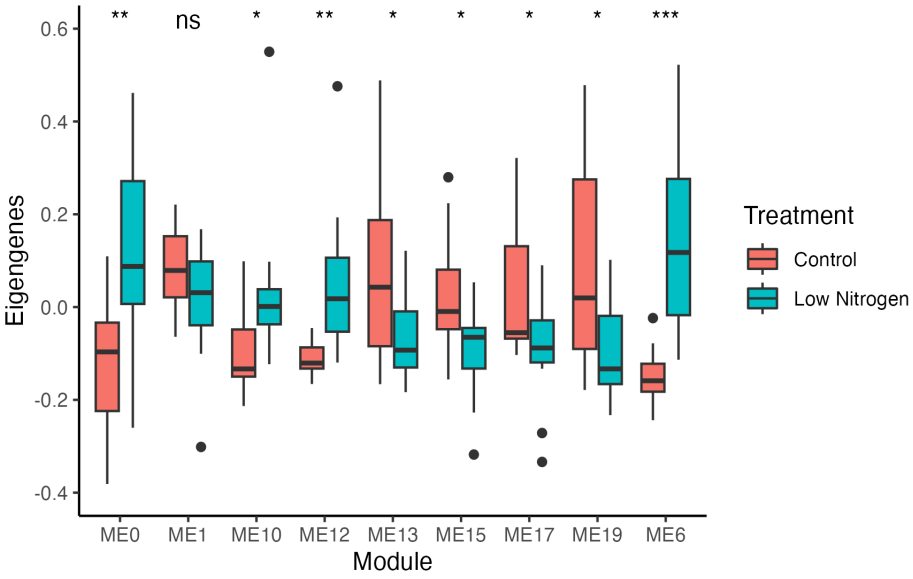

Figure 3: Differential Eigengene Values Across Modules by Treatment. The boxplot compares eigengene values for modules ME0, ME1, ME6, ME10, ME12, ME13, ME15, ME17, ME19 between control (red) and low nitrogen (blue) treatments. Modules were selected based on their significant differential expression from a linear regression against controls. The y-axis represents eigengene values, which are summary expressions of the genes within each module.

Gene Ontology Enrichment Analysis Reveals Modular Functional Specificity

Our enrichment analysis of GO terms across differentially expressed gene modules identified specific biological processes and cellular components that are uniquely overrepresented in each module. The following outlines the common themes within the top four GO terms for each significantly different module compared to control:

Module 6 exhibits upregulated genes with a convergence of terms related to cellular anatomy and metabolic pathways. The presence of "intracellular anatomical structure" and "cytoplasm" implies an intracellular localization of module-associated functions, while "protein-containing complex" and "catabolic process" align with protein metabolism and degradation, highlighting a potential role for this module in cellular metabolism and protein turnover.

Module 10's GO terms suggest the upregulation of genes related to cellular constituents and bioenergetic processes, with a clear focus on photosynthesis. The co-enrichment of "cytoplasm," "plastid," and "chloroplast" point towards the module's involvement in photosynthetic energy conversion, whereas "oxidoreductase activity" underscores the module's role in electron transport and redox reactions fundamental to metabolic energy transduction.

Module 13 presents a singular enrichment in "Carotenoid biosynthesis," indicative of the down-regulation of a specialized metabolic pathway for the production of carotenoids. This suggests a distinct module function in synthesizing these compounds, which are known to have roles in plant photoprotection and antioxidative responses (Pérez-Gálvez et al, 2020).

Modules 15 and 19 are distinguished by their specific association with lipid metabolism. Enriched terms such as "glycerophospholipid metabolism," "inositol phosphate metabolism,"

"lipid biosynthetic process," and "phospholipid metabolic process" collectively point to genes involved in the biosynthesis and transformation of lipids being down-regulated under low nitrogen conditions. This strongly suggests the relevance of lipid dynamics and cellular membrane homeostasis in the response to low nitrogen.

Module 17 is representative of downregulated genes tied to signal transduction and regulatory mechanisms within plant cells. Notably, the "MAPK signaling pathway - plant" and "plant hormone signal transduction" reflect the module's putative role in cellular signaling, while terms like "protein dephosphorylation" and "dephosphorylation" are indicative of the module's function in protein modification and the modulation of signal transduction pathways.

The identified enriched GO terms within each module denote not only the potential biological roles but also suggest a coordination of gene expression tailored to the physiological demands of the plant's response to nitrogen availability. These findings provide a framework for further investigation into the molecular underpinnings of nitrogen utilization and assimilation in plants.

Differential expression of genes under low nitrogen

To identify differential expression at the individual gene level, we separately analyzed the differential expression of each of our mapped genes between the low nitrogen and the control treatment samples. We found a total of 1453 up-regulated, and 1488 down-regulated loci. Of these, 1241 were assigned to one of the modules, 451 remained unassigned, and 434 were filtered out of the initial dataset, due to having a variance lower than 1. Figure 4 shows the genes with log2 fold changes over 5, found through the DESeq method, and their corresponding modules in the co-expression network. Remarkably, most of the top differentially expressed genes are being placed in network modules that were not significantly different between the low nitrogen and the control treatments. Nevertheless, the most down-regulated gene was found in Module 15. The *Zm00001eb132640* locus was found to have a log2 fold change of -8.34 in the low nitrogen treatment. It encodes a putative protein with GO terms for organic cyclic compound biosynthesis and the sterol metabolic process, congruent with the enriched terms for its module. Although its function in maize and closely related grasses is unknown, its *Arabidopsis* ortholog encodes a sterol 4 alpha-methyl oxidase involved in sterol composition and the balance of auxin and cytokinin activities during embryogenesis. On the other hand, the second highest up-regulated gene was placed in Module 6. Consistently with the enriched terms for its module, the *Zm00001eb423830* locus has a 6 log2 fold-change. It has associated GO terms for “pteridine-containing compound metabolic process” and is a putative membrane-integrated protein. Despite lacking further functional information, genetic variants in this locus have been reportedly associated to the agronomically-important traits kernel row number and growing degree days to anthesis in cultivated corn (Wallace, et al., 2014).

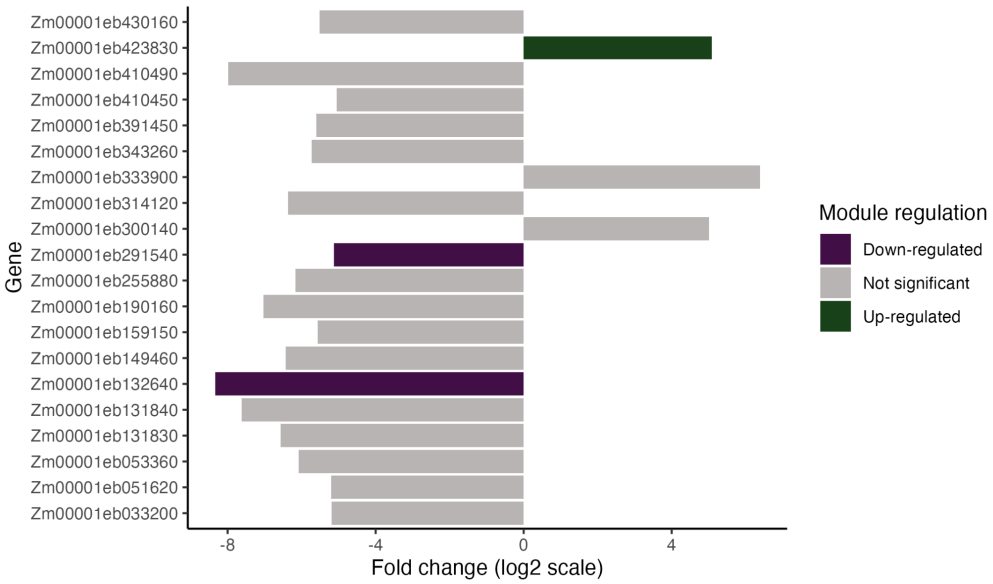

Figure 4: Highest differentially expressed genes. Bar-plots of the genes with higher than 5 log2 fold change (negative and positive). Bar color indicates whether the gene's corresponding module was down- or up-regulated, or if it was found to be not significantly different between treatments.

Discussion

Plants respond to abiotic stresses in a complex manner that would benefit from more holistic approaches aiming to study the interaction between morphological, physiological, and molecular changes instead of studying these in isolation. Here, we navigate the integration of phenomics, ionomics, and transcriptomics datasets to study the response to nitrogen deficiency in maize seedlings using a weighted correlation network analysis. We were able to identify clusters of differentially expressed genes that were correlated with each other as well as with morphological changes in the treated plants. The gene ontology enrichment analysis of these modules allowed us to determine what molecular processes these genes were involved in. The significantly up-regulated modules were found to be involved in protein metabolism and the chloroplast electron transfer chain machinery. Their negative correlation to plant area and mean hue suggests that under low nitrogen conditions, the plants prioritize energy production and respiration processes at the expense of a reduction in photosynthetic light-harvesting reactions, which compromises shoot growth and results in the loss of chlorophyll molecules that causes leaf yellowing (Mu & Chen 2021). On the other hand, gene modules related to lipid and cell membrane homeostasis as well as signal transduction were found to be significantly down-regulated and positively correlated to plant area and mean hue. Studies in *Arabidopsis*, wheat, and soybean have previously found that lipid content and regulation of chloroplast membrane lipid molecules is affected by nitrogen deficiency (Gaude et al. 2007, Narasimhan et al. 2013, Li et al. 2019). More recently, a positive correlation between shoot dry weight and linoleic acid content in wheat seedlings grown under low nitrogen conditions (Liu et al. 2020). These findings indicate that membrane remodeling plays a pivotal role in the plant's ability to support further shoot growth under the stress.

Complementing our correlation network with a classic differential gene expression analysis (DESeq method) allowed us to identify individual genes that were highly up- or down-regulated within our network modules. The most up- or down-regulated genes common in both analyses may be great targets for further functional characterization through genetic transformation and for the search of naturally occurring variants in non-cultivated maize accessions. Notably, several of the significantly differentially expressed genes found through the DESeq method were placed by the WGCNA in Module 1, which was found to be not significantly different between the treatments. Being the module with the largest number of associated features, the absence of a $p<0.05$ is understandable yet it suggests our network analysis could benefit from parameter tuning.

On the same vein, the ionic levels extracted from the ionomics dataset were either placed in Module 1 or remained unassigned. During the data exploration phase of this project we conducted a principal component analysis in our datasets and found that even after filtering out low variance ions our samples did not seem to be clustering by treatment or genotype. This all hints at the ions measured not being strongly correlated to the mapped genes, and/or our ionomics dataset not capturing differences between treatments. For future integrative approaches, we believe a better design would be to make different correlation networks for the transcriptomics and ionomics datasets separately and find correlations between the modules themselves.

Overall, this project was successful in identifying molecular processes that affect morphological features of maize seedlings under nitrogen deficient conditions, allowing us to integrate visual observations with molecular changes in the treated plants. Supplementing our network analysis with the DESeq method to find individually differentially expressed genes provided us with specific loci involved in relevant biological processes that could potentially be crop improvement targets. Future work should explore other data normalization and partition techniques to better leverage ionomics datasets to further our understanding of the interconnectedness of the plant response to abiotic stresses.

References

Erenstein, Olaf, Moti Jaleta, Kai Sonder, Khondoker Mottaleb, & B.M. Prasanna. “Global Maize Production, Consumption and Trade: Trends and R&D Implications.” *Food Security* 14, no. 5 (October 1, 2022): 1295–1319. https://doi.org/10.1007/s12571-022-01288-7.

Gaude, N., Brehelin, C., Tischendorf, G., Kessler, F., & Dormann, P. (2007). Nitrogen deficiency in Arabidopsis affects galactolipid composition and gene expression and results in accumulation of fatty acid phytyl esters. The Plant Journal, 49, 729–739.

Kolberg, Liis, Uku Raudvere, Ivan Kuzmin, Jaak Vilo, & Hedi Peterson. “Gprofiler2 -- an R Package for Gene List Functional Enrichment Analysis and Namespace Conversion Toolset

g:Profiler." F1000Research, November 17, 2020. https://doi.org/10.12688/f1000research.24956.2.

Langfelder, Peter, & Steve Horvath. "WGCNA: an R package for weighted correlation network analysis." BMC Bioinformatics 9 (2008): 559. https://doi.org/10.1186/1471-2105-9-559.

Li, Q., Qi, L. Y., Yin, L. N., Wang, S. W., & Deng, X. P. (2019). Relationship between lipid and flag leaf senescence induced by low nitrogen stress during grain filling of wheat. Acta Agronomica Sinica, 44, 1221–1228.

Liu, X., Wang, S., Deng, X., Zhang Z., & Yin L. (2020). Comprehensive evaluation of physiological traits under nitrogen stress and participation of linolenic acid in nitrogen-deficiency response in wheat seedlings. BMC Plant Biology, 20, 501. https://doi.org/10.1186/s12870-020-02717-5

Mu, X., & Chen, Y. (2021). The physiological response of photosynthesis to nitrogen deficiency. Plant Physiology and Biochemistry, 158, 76-82. https://doi.org/10.1016/j.plaphy.2020.11.019.

Narasimhan, R., Wang, G. L., Li, M. Y., Roth, M., Welti, R., & Wang, X. M. (2013). Differential changes in galactolipid and phospholipid species in soybean leaves and roots under nitrogen deficiency and after nodulation. Phytochemistry, 96, 81–91.

Pérez-Gálvez, Antonio, Isabel Viera, & María Roca. "Carotenoids and Chlorophylls as Antioxidants." *Antioxidants* 9, no. 6 (June 9, 2020): 505. https://doi.org/10.3390/antiox9060505.

Wallace, J. G., Bradbury, P. J., Zhang, N., Gibon, Y., Stitt, M., & Buckler, E. S. (2014). Association Mapping across Numerous Traits Reveals Patterns of Functional Variation in Maize. PLoS Genetics, 10(12), e1004845. https://doi.org/10.1371/journal.pgen.1004845.