

Exploring the Effects of Plasma Treatment on *Sorghum bicolor*: An RNA-Seq Analysis

Fangyi Li

Student Number: 1005937774

Course: BCB430Y 2023-2024
Advanced Special Project in Bioinformatics and Computational Biology

Advisory Team:

Professor Nicholas J. Provert
Professor Takamasa Okumura
Professor Eiji Nambara
Vincent Lau, MSc.

April 4, 2024

Abstract

Plasma treatment, encompassing various technologies such as cold plasma and atmospheric plasma, holds promise in agricultural research for enhancing plant growth, improving crop productivity, and alleviating environmental stresses. Despite its potential, the molecular mechanisms governing the response of plants to plasma treatment, especially in crops like *Sorghum bicolor*, remain poorly understood. In this study, we utilized RNA sequencing (RNA-Seq) analysis coupled with bioinformatics tools to explore the transcriptional response of *Sorghum bicolor* to plasma treatment. Our differential gene expression analysis unveiled significant alterations in gene expression profiles across different conditions, with a notable number of genes showing up and down-regulation in response to plasma treatment compared to untreated (wild-type) plants. Functional enrichment analysis, employing established databases or tools such as agriGO, g:Profiler, and PlantRegMap, elucidated the biological processes, molecular functions, and cellular components associated with the differentially expressed genes. Noteworthy enrichments included processes related to photosynthesis, response to environmental stimuli, and metabolic pathways. Moreover, visualization of gene expression patterns using the Sorghum eFP Browser provided insights into the spatial and temporal dynamics of gene expression changes induced by plasma treatment. Overall, our findings deepen the understanding of the molecular mechanisms underlying *Sorghum bicolor*'s response to plasma treatment, with potential implications for enhancing crop productivity and sustainability in agricultural practices.

Keywords: *Sorghum bicolor*, Plasma Treatment, RNA-Seq, eFP Browser

1 Introduction

Sorghum bicolor, commonly known as sorghum, stands out as a vital cereal crop globally, prized for its versatility and resilience across diverse agroecological landscapes. Its capacity to thrive under various climatic conditions and its multifaceted utility in food, feed, fodder, and bioenergy production underscore its pivotal role in bolstering global food security and advancing agricultural sustainability (McCormick et al. 2018).

Plasma treatment, an emerging technology in agricultural research, has garnered increasing attention for its potential to bolster plant growth, enhance crop productivity, and mitigate biotic and abiotic stresses. Functioning as the fourth state of matter, plasma comprises a highly reactive ionized gas housing an array of reactive oxygen and nitrogen species, ultraviolet radiation, and electromagnetic fields. When applied to plants, plasma treatment induces physiological and biochemical transformations at the molecular level, triggering shifts in gene expression patterns and activating defense mechanisms that fortify plant resilience against environmental stressors (Attri et al. 2020).

However, despite the burgeoning interest in plasma treatment as a sustainable agricultural modality, the molecular intricacies underlying the plant's response to plasma treatment still need to be more adequately understood, particularly in crops like sorghum. Understanding how plasma treatment influences gene expression patterns in sorghum holds promise for shedding light on the molecular underpinnings of plasma-induced stress tolerance and growth promotion, thus offering potential avenues for enhancing crop productivity and sustainability.

In this study, we employed RNA sequencing (RNA-Seq) analysis in conjunction with bioinformatics tools to probe the transcriptional response of *Sorghum bicolor* to plasma treatment. By characterizing differential gene expression patterns, functional enrichment analyses, and gene expression visualization, we aimed to elucidate the molecular mechanisms governing the plant's response to plasma treatment and pinpoint potential candidate genes associated with plasma-induced stress tolerance and growth promotion in sorghum.

2 Methods

2.1 Experimental Design

Condition	Label Name	Number of Replicates
Dry Seed	Seed Dry	3
Imbibed Seed (no plasma treatment)	Seed WT	3
Imbibed Seed with Plasma Treatment	Seed Plasma	3
Shoot from Imbibed Seed	Shoot WT	3
Shoot from Imbibed Seed with Plasma	Shoot Plasma	3

Table 1: Overview of Experimental Conditions and Replicates.

2.1.1 Dry Seed Condition

RNA-Seq data for the dry seed condition were obtained from a previous study titled "Gene expression analysis of *Sorghum bicolor* BTx623 (Federhen 2012)." These reads represented the initial gene expression state before undergoing any experimental treatment, providing a comparable condition for comparative analyses with seeds subjected to imbibition and other treatments.

2.1.2 Imbibed Seed Conditions

Seeds were subjected to controlled imbibition conditions to simulate natural germination processes. Each condition involved the incubation of ten seeds per bed, with the bed comprising paper pulp (1.2 g) mixed with water (2.7 mL) in polymer dishes. The incubation environment was maintained under a 16-hour light cycle with a photosynthetic photon flux density (PPFD) of $100 \mu\text{mol m}^{-2} \text{s}^{-1}$ and a temperature of 28°C. Seeds in specific conditions were treated with plasma during imbibition, with the "1-day imbibed seed with plasma treatment" condition aimed at investigating any potential effects of plasma treatment on gene expression compared to seeds without plasma treatment. Sampling was conducted one day after seed imbibition, with RNA-Seq data collected from three seeds per condition, replicated three times for two conditions. This baseline data represented the gene expression profile at the onset of seed imbibition.

2.1.3 Shoot Conditions

Following seed imbibition, seeds were replanted into pods to facilitate shoot growth. Each pod accommodated one seed and was filled with a soil mixture comprising culture soil (Sakata Super Mix, Sakata Seed Co., Tokyo, Japan) and vermiculite at a ratio of 2:1. Nutrient supplementation was provided by adding 1000-fold diluted HYPONEX 3L to the soil of 6L. Pod growth conditions included a 16-hour light cycle with a PPFD of $80 \mu\text{mol m}^{-2} \text{s}^{-1}$ and a temperature of 28°C. It is important to note that shoots were not subjected to plasma treatment directly. Instead, shoots were grown from seeds that had undergone either plasma or no treatment. Sampling of shoot tissues was performed four weeks after seed imbibition. RNA-Seq data were collected from one individual per condition and replicated three times for two conditions. This baseline data represented the gene expression profile of shoots grown from seeds subjected to the imbibition process, with and without plasma treatment.

2.2 Bioinformatics Analysis

2.2.1 Read Alignment with HISAT2

The raw sequencing reads obtained from the experimental design section underwent alignment using HISAT2 (Hierarchical Indexing for Spliced Alignment of Transcripts 2). HISAT2 was selected for its fast and sensitive alignment capabilities. It is specifically designed for aligning RNA-Seq reads to reference genomes while accounting for known splice junctions (Kim et al. 2019).

```
#!/bin/bash

# Loading modules needed
module load hisat2
module load samtools

# 1. Convert GTF to HISAT2's known-splice-site format
hisat2_extract_splice_sites.py Sbicolor_annotations.gtf > splice_sites.txt
hisat2_extract_exons.py Sbicolor_annotations.gtf > exons.txt

# 2. Build the HISAT2 index
hisat2-build -p 8 --ss splice_sites.txt --exon exons.txt Sbicolor_reference.fa
Sbicolor_index

# 3. Align RNA-Seq reads using HISAT2 (paired-end)
hisat2 -p 8 --phred33 -x Sbicolor_index -1 read_1.fq -2 read_2.fq -S sample.sam

# 4. Convert SAM to BAM and sort
samtools view -@ 8 -b -o sample_unsorted.bam sample.sam
samtools sort -@ 8 -o sample.bam sample_unsorted.bam
```

Listing 1: Example Bash Code for Read Alignment with HISAT2 and Samtools.

The alignment process began with indexing the reference genome using HISAT2. For this study, the genome annotation file in GTF (Gene Transfer Format) and the genome reference file in FASTA format were obtained and converted from Phytozome (Goodstein et al. 2012). This step involved generating an index of the reference genome to facilitate rapid alignment.

Once the reference genome was indexed, the sequencing reads in FASTQ format were aligned to the indexed reference genome using HISAT2. HISAT2 employs a two-step approach to identify potential splice junctions and align reads accordingly during alignment.

The alignment process output was a file in SAM (Sequence Alignment/Map) format containing information about the alignment of each read to the reference genome. Subsequently, the SAM files were converted into sorted BAM (Binary Alignment/Map) files using Samtools (Danecek et al. 2021). Sorted BAM files are advantageous for downstream analyses due to their compressed format and indexed alignment positions.

2.2.2 Read Counting with featureCounts

Following the alignment of sequencing reads to the reference genome, the subsequent step in the RNA-Seq analysis pipeline involved quantifying the number of reads aligned to each genomic feature, such as genes or transcripts. This quantification was conducted using featureCounts, a widely utilized software tool for read counting in RNA-Seq analysis known for its accuracy and efficiency (Liao, Smyth, and W. Shi 2014).

```

# Run featureCounts
fc_result <- Rsubread::featureCounts(files = list.files(path = bam_path, pattern
= "\\\\.bam$", full.names = TRUE),
                                      annot.ext = gtf,
                                      isGTFAnnotationFile = TRUE,
                                      GTF.featureType = "exon",
                                      GTF.attrType = "gene_id",
                                      useMetaFeatures = TRUE,
                                      isPairedEnd = TRUE)

# Extract raw counts
counts <- fc_result$counts

```

Listing 2: Example R code for Read Counting with featureCounts.

FeatureCounts takes the aligned reads in BAM format as input and a genome annotation file in GTF format. In this study, the genome annotation file in GTF format was obtained from Phytozome. It furnishes vital information about the genomic coordinates of genes and other genomic features essential for accurate read counting.

The featureCounts process began with input preparation, where aligned reads in BAM format, and the genome annotation file in GTF format were provided as input to featureCounts. Subsequently, featureCounts assigned each aligned read to genomic features based on their alignment positions, handling reads that overlapped with multiple features or fell within ambiguous regions according to specified parameters.

The output of featureCounts was a count matrix, where each row represented a genomic feature (e.g., gene), and each column represented a sample. This count matrix was the foundation for downstream analyses, such as differential gene expression analysis using tools like DESeq2.

2.2.3 Differential Gene Expression Analysis with DESeq2

Following read counting with featureCounts, the count matrix containing the read counts for each genomic feature across samples underwent differential gene expression analysis using DESeq2. DESeq2, a widely used R/Bioconductor package, is designed explicitly for analyzing RNA-Seq data, particularly for identifying genes that are differentially expressed between conditions (Love, Huber, and Anders 2014).

```

# Perform DESeq2 analysis
dds <- DESeq2::DESeqDataSetFromMatrix(counts, colData, design = ~ condition)
dds <- DESeq2::DESeq(dds)

# Get Results
res <- DESeq2::results(dds)

```

Listing 3: Example R Code for Differential Gene Expression Analysis with DESeq2.

The DESeq2 analysis began with the import of the count matrix generated by featureCounts into R using the DESeq2 package. Upon import, DESeq2 normalized the count data to correct for differences in library size and composition across samples, ensuring that gene expression levels were comparable.

Subsequently, DESeq2 utilized a negative binomial distribution to model read counts and employed a generalized linear model framework to identify differentially expressed genes. The model accounted for variability in sequencing depth and biological variability between replicates. Statistical testing was then conducted to determine whether the expression of each gene was significantly different between conditions.

Genes with adjusted p-values below a specified threshold (e.g., FDR <0.05) were considered statistically significant. DESeq2 output included lists of differentially expressed genes, corresponding fold changes and statistical significance values. These results provided insights into the up-regulated or down-regulated genes under different experimental conditions.

2.2.4 Functional Analysis Tools

Functional analysis tools are essential for interpreting the biological significance of differentially expressed genes identified through RNA-Seq analysis. These tools allow researchers to explore the functional roles and biological processes associated with the identified genes.

This study utilized several functional analysis tools to gain insights into the biological functions, pathways, and regulatory networks associated with the differentially expressed genes. The functional analysis tools employed include:

1. **agriGO**: agriGO is a web-based tool specifically designed for Gene Ontology (GO) enrichment analysis in agricultural species. It enables the identification of overrepresented GO terms among a list of genes of interest, providing insights into the biological processes, molecular functions, and cellular components associated with the genes (T. Tian et al. 2017).
2. **g:Profiler**: A versatile tool for functional enrichment analysis of gene lists. It integrates multiple databases, including GO, KEGG, and Reactome pathways, to identify enriched functional categories and pathways, offering comprehensive annotations and visualizations for interpreting gene lists (Reimand et al. 2007).
3. **PlantRegMap**: A comprehensive database of plant transcriptional regulatory networks and regulatory elements. It provides insights into transcription factors, target genes, and regulatory interactions in various plant species, aiding in exploring regulatory networks associated with differentially expressed genes (F. Tian et al. 2020).

2.2.5 eFP Browser

The eFP (electronic Fluorescent Pictograph) browser is a valuable tool for visualizing gene expression data (Winter et al. 2007). In our study, we utilized TPM (Transcripts Per Million) values calculated with the eFP browser to visualize gene expression patterns of interest (Wagner, Kin, and Lynch 2012). This visualization allowed researchers to observe spatial and temporal gene expression patterns across various tissues and developmental stages.

TPM values were calculated using the formula:

$$TPM = \frac{n_g}{N \cdot l_g} (10^3) (10^6)$$

where n_g is the number of reads/fragments for the gene, N is the sum of reads/fragments, and l_g is the length of the gene (Wagner, Kin, and Lynch 2012).

Moreover, correlation analysis was conducted between TPM values of dry seed samples and those from previous databases. This comparative analysis provided valuable insights by aligning current gene expression profiles with established data from prior studies. By leveraging the eFP browser for visualization and conducting correlation analysis, researchers better understood gene expression dynamics in plant growth, development, and environmental responses.

3 Results

3.1 Differential Gene Expression Analysis

3.1.1 DESeq2 Results

```
> print(dds_seed_wt_dry)
log2 fold change (MLE): condition seed_dry vs seed_wt
wald test p-value: condition seed dry vs seed wt
DataFrame with 28717 rows and 6 columns
  baseMean log2FoldChange lfcSE stat pvalue padj
  <numeric> <numeric> <numeric> <numeric> <numeric>
Sobic.001G000100.v3.2 3.97598 -3.643486 1.425678 -2.55562 1.06000e-02 2.04099e-02
Sobic.001G000200.v3.2 632.37911 -1.215341 0.195499 -6.21662 5.07986e-10 3.04781e-09
Sobic.001G000400.v3.2 282.89857 -0.782365 0.222048 -3.52341 4.26033e-04 1.08414e-03
Sobic.001G000300.v3.2 4.60058 4.407810 1.104089 3.99226 6.54465e-05 1.92489e-04
Sobic.001G000700.v3.2 182.04448 -1.587959 0.267573 -5.93468 2.94411e-09 1.60347e-08
...
Sobic.K043600.v3.2 0.769405 1.288839 2.119533 0.608077 0.5431365 0.649893
Sobic.K043700.v3.2 27.843352 0.803814 0.483482 1.662551 0.0964023 0.148734
Sobic.K043800.v3.2 2.430546 -1.370751 2.205958 -0.621386 0.5343460 0.641335
Sobic.K044200.v3.2 39.730762 -0.898586 0.496942 -1.808230 0.0705707 0.112760
Sobic.K044300.v3.2 207.139737 0.170289 0.222827 0.764220 0.4447361 0.550420
```

(a) Seed WT vs. Dry

```
> print(dds_seed_wt_plasma)
log2 fold change (MLE): condition seed_plasma vs seed_wt
wald test p-value: condition seed plasma vs seed wt
DataFrame with 28717 rows and 6 columns
  baseMean log2FoldChange lfcSE stat pvalue padj
  <numeric> <numeric> <numeric> <numeric> <numeric>
Sobic.001G000100.v3.2 3.97598 -3.9679900 1.421504 -2.791404 0.00524800 0.0169289
Sobic.001G000200.v3.2 632.37911 0.0105073 0.181010 0.058048 0.95371042 1.00000000
Sobic.001G000400.v3.2 282.89857 -0.6107034 0.208899 -2.923440 0.00346187 0.0118489
Sobic.001G000300.v3.2 4.60058 3.0883923 1.105207 2.794402 0.00519958 0.0167967
Sobic.001G000700.v3.2 182.04448 -0.6802701 0.244964 -2.777015 0.00548606 0.0175889
...
Sobic.K043600.v3.2 0.769405 -1.698837 2.293805 -0.740620 0.45892395 NA
Sobic.K043700.v3.2 27.843352 -1.574370 0.505798 -3.112645 0.00185419 0.00690263
Sobic.K043800.v3.2 2.430546 -2.654903 2.205958 -1.203515 0.22877717 0.36992413
Sobic.K044200.v3.2 39.730762 -0.113607 0.426500 -0.266371 0.78995345 0.89252750
Sobic.K044300.v3.2 207.139737 -0.383588 0.215036 -1.783833 0.07445078 0.15288196
```

(b) Seed WT vs. Plasma

```
> print(dds_shoot_wt_plasma)
log2 fold change (MLE): condition shoot_plasma vs shoot_wt
wald test p-value: condition shoot plasma vs shoot wt
DataFrame with 28717 rows and 6 columns
  baseMean log2Foldchange lfcSE stat pvalue padj
  <numeric> <numeric> <numeric> <numeric> <numeric>
Sobic.001G000100.v3.2 3.97598 -0.9843752 0.850624 -1.1572387 0.247175 1
Sobic.001G000200.v3.2 632.37911 -0.0959605 0.178707 -0.5369710 0.591288 1
Sobic.001G000400.v3.2 282.89857 -0.0751542 0.204381 -0.3677165 0.713085 1
Sobic.001G000300.v3.2 4.60058 -0.6001341 1.470825 -0.4080254 0.683255 1
Sobic.001G000700.v3.2 182.04448 -0.0159357 0.250892 -0.0635161 0.949356 1
...
Sobic.K043600.v3.2 0.769405 0.958375 2.473182 0.387507 0.698381 1
Sobic.K043700.v3.2 27.843352 -0.425415 0.906269 -0.469414 0.638774 1
Sobic.K043800.v3.2 2.430546 -0.403215 1.790657 -0.225177 0.821841 1
Sobic.K044200.v3.2 39.730762 -0.148845 0.407127 -0.365599 0.714664 1
Sobic.K044300.v3.2 207.139737 -0.150428 0.204674 -0.734965 0.462361 1
```

(c) Shoot WT vs. Plasma

Figure 1: DESeq2 Results of (a) Seed WT vs. Dry, (b) Seed WT vs. Plasma, and (c) Shoot WT vs. Plasma Comparisons.

Figure 1 presents result tables for each comparison, containing key parameters such as the mean of normalized counts for all samples (baseMean), log2 fold change (log2FoldChange), standard error (lfcSE), Wald statistic (stat), Wald test p-value (pvalue), and Benjamini-Hochberg adjusted p-value (padj). The DESeq2 analysis utilized an adjusted p-value threshold of <0.05 . Differentially expressed genes were identified based on a log2 fold change threshold of >1 for up-regulated and <-1 for down-regulated genes. These results unveiled significant transcriptomic changes across different experimental conditions, indicating the dynamic nature of gene regulation in the studied organism (Love, Huber, and Anders 2014).

Comparison	Up-regulated Genes	Down-regulated Genes
Seed WT vs. Dry	4737	6852
Seed WT vs. Plasma	1976	3603
Shoot WT vs. Plasma	0	0

Table 2: Number of Up and Down-regulated Genes for Each Comparison.

Table 2 summarizes the number of up and down-regulated genes identified for each comparison, highlighting the magnitude of gene expression alterations in response to various experimental treatments.

3.1.2 PCA Plot

Principal Component Analysis (PCA) was utilized to visualize the overall patterns of variation in gene expression across samples. The PCA plot reveals distinct clustering patterns among the samples based on their transcriptional profiles (Pearson 1901).

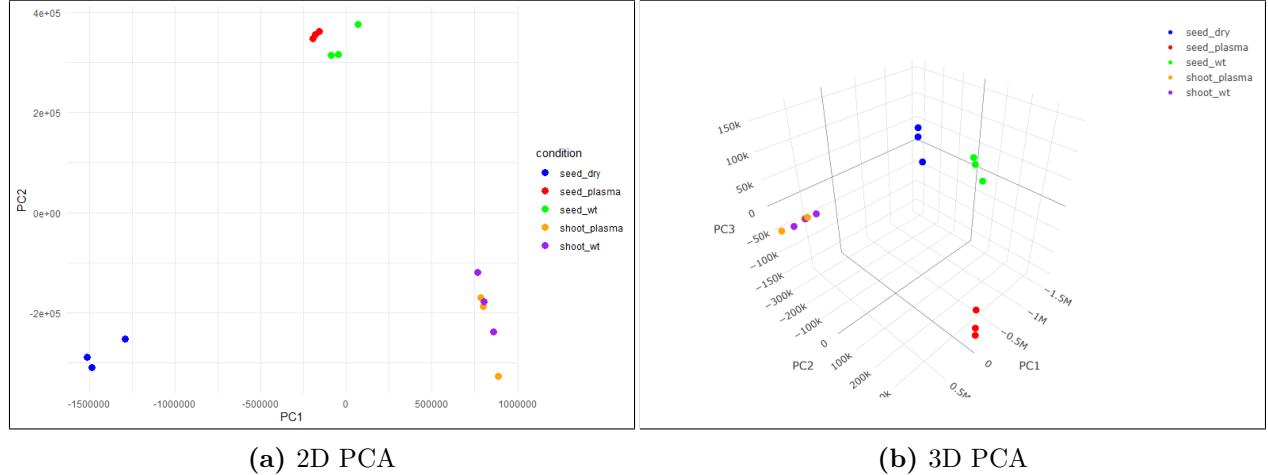


Figure 2: PCA Plots in (a) 2D and (b) 3D.

Figure 2 illustrates that replicates of Seed Dry, Seed WT, and Seed Plasma samples form three distinct clusters, indicating similar gene expression profiles within each experimental condition. Conversely, replicates of Shoot WT and Shoot Plasma samples cluster together as a single group, suggesting similar gene expression profiles between these conditions.

The observed clustering patterns in the PCA plot provide valuable insights into the relationships between samples and highlight the transcriptional differences between the experimental conditions.

3.1.3 MA Plot

MA plots were utilized to visualize gene expression differences between conditions. Each point on the plot represents a gene, with the x-axis indicating the average expression level (A) and the y-axis showing the log-fold change (M) in gene expression between conditions.

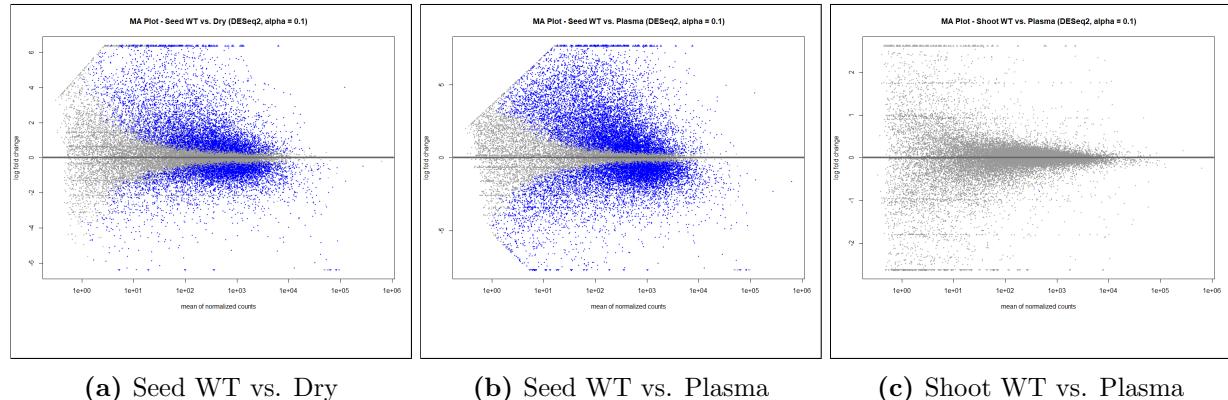


Figure 3: MA Plots of (a) Seed WT vs. Dry, (b) Seed WT vs. Plasma, and (c) Shoot WT vs. Plasma Comparisons.

In Figure 3, comparing (a) Seed WT vs. Dry and (b) Seed WT vs. Plasma, two distinct sets of points are observed: grey and blue. Grey points represent genes that do not meet the criteria for differential expression, while blue points highlight genes that are significantly differentially expressed (adjusted p-value <0.05 and log₂ fold change >1 or <-1). Notably, only the comparisons of (a) Seed WT vs. Dry, as well as (b) Seed WT vs. Plasma, exhibit blue points, indicating significant changes in gene expression between these conditions.

Blue points in these MA plots suggest that certain genes undergo significant changes in expression levels during seed imbibition and in response to plasma treatment. These findings provide valuable insights into the transcriptional dynamics associated with these experimental conditions.

3.1.4 Volcano Plot

Volcano plots were employed to visualize the relationship between statistical significance and fold change of gene expression between two conditions. In these plots, each point represents a gene, with the x-axis indicating the log₂ fold change and the y-axis representing the statistical significance, typically measured as the adjusted p-value.

Three distinct sets of points are observed in the volcano plots: black, blue, and red. Black points represent genes that do not meet the criteria for differential expression. Blue points indicate genes with an adjusted p-value <0.05 , while red points highlight genes that not only have an adjusted p-value <0.05 but also have a log₂ fold change >1 or <-1 .

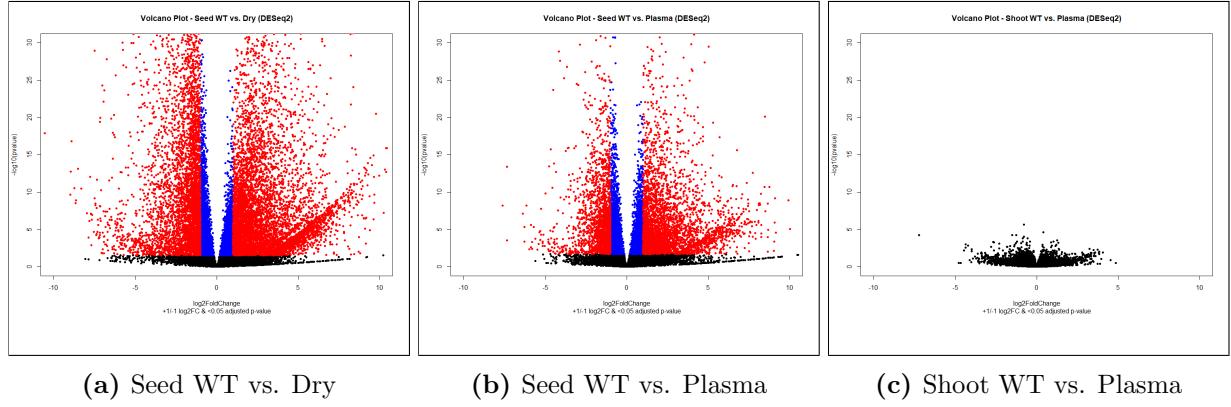


Figure 4: Volcano Plots of (a) Seed WT vs. Dry, (b) Seed WT vs. Plasma, and (c) Shoot WT vs. Plasma Comparisons.

Figure 4 compares (a) Seed WT vs. Dry and (b) Seed WT vs. Plasma, exhibiting both blue and red points, indicating significant changes in gene expression between these conditions. In contrast, Figure 4 comparing (c) Shoot WT vs. Plasma only contains black points, indicating no significant differential gene expression between these conditions.

The presence of blue and red points in the volcano plots highlights genes with significant changes in expression levels, providing valuable insights into the transcriptional responses associated with seed imbibition and plasma treatment.

3.2 Functional Analysis

Functional analysis tools were utilized to uncover the biological processes and pathways influenced by gene expression changes in response to plasma treatment. The analysis concentrated on data comparing Seed WT and Seed Plasma conditions, as no differentially expressed genes were identified in the Shoot WT vs. Plasma comparison.

3.2.1 agriGO

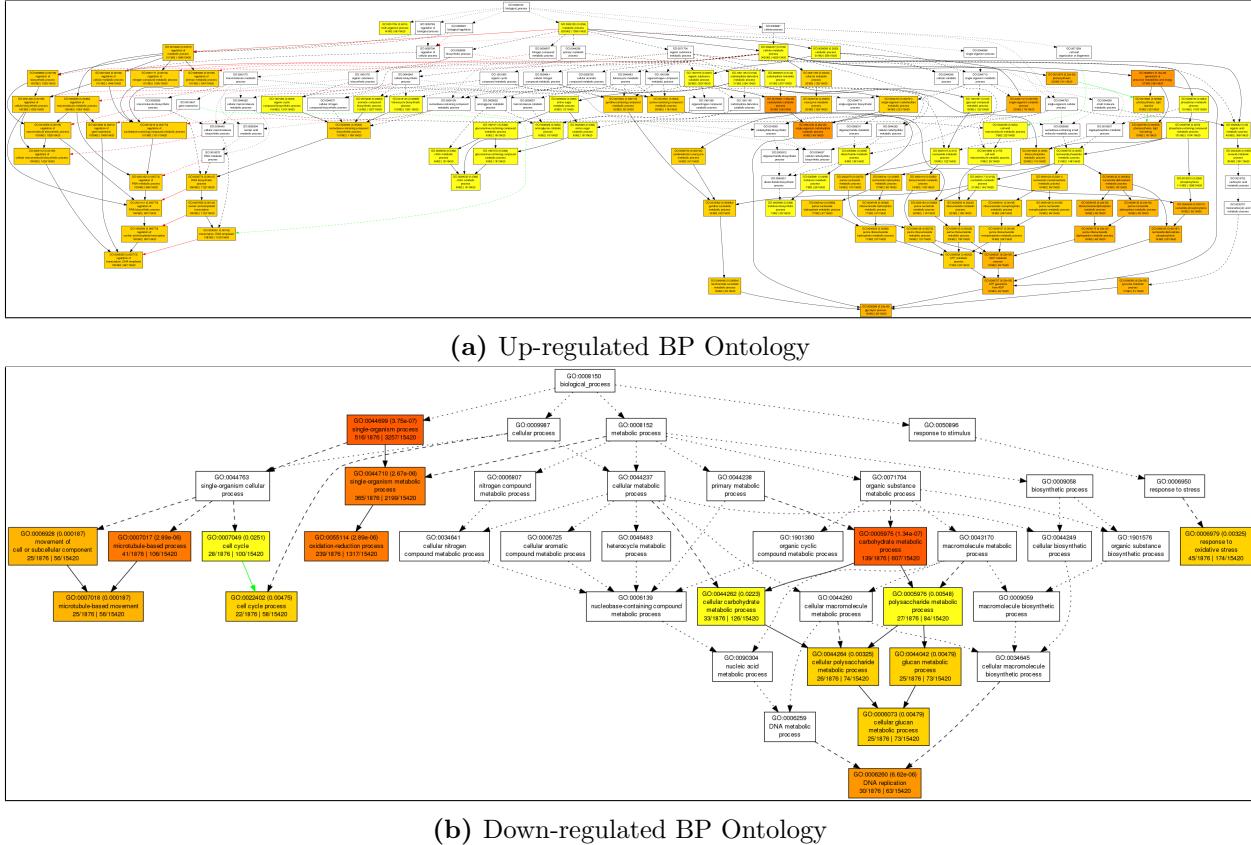
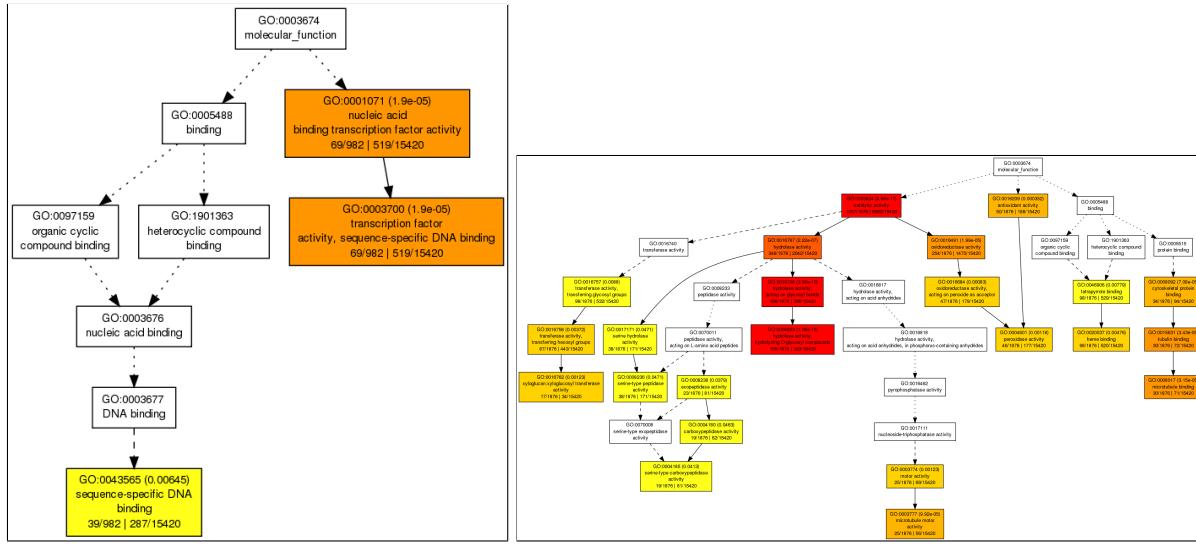


Figure 5: agriGO Biological Process Ontology of (a) Up-regulated and (b) Down-regulated Genes.

Figure 5 illustrates the functional analysis results conducted using agriGO, focusing on the biological processes associated with differentially expressed genes between Seed WT and Seed Plasma conditions (T. Tian et al. 2017).

Subfigure (a) displays the biological process graph of up-regulated genes, highlighting key GO terms related to metabolic processes and energy generation. Notable terms include "generation of precursor metabolites and energy," "carbohydrate catabolic process," and "photosynthesis," indicating enrichment of pathways involved in energy metabolism and carbohydrate utilization in response to plasma treatment.

Conversely, Subfigure (b) presents the biological process graph of down-regulated genes, revealing GO terms associated with diverse cellular processes such as carbohydrate metabolism, oxidation-reduction processes, and DNA replication. These findings suggest suppressing metabolic and cellular activities in response to plasma treatment, potentially reflecting adaptive responses to altered environmental conditions.



(a) Up-regulated MF Ontology

(b) Down-regulated MF Ontology

Figure 6: agriGO Molecular Function of Ontology (a) Up-regulated and (b) Down-regulated Genes.

Figure 6 provides insights into the molecular functions associated with differentially expressed genes between Seed WT and Seed Plasma conditions, as revealed by agriGO analysis (T. Tian et al. 2017).

Subfigure (a) illustrates the molecular function graph of up-regulated genes, highlighting key GO terms related to transcriptional regulation and enzymatic activities. Notable terms include "transcription factor activity," "sequence-specific DNA binding," and "carbohydrate kinase activity," suggesting an enrichment of regulatory and metabolic functions in response to plasma treatment.

Conversely, Subfigure (b) depicts the molecular function graph of down-regulated genes, revealing GO terms associated with catalytic activities and microtubule binding. These findings indicate suppressing enzymatic and structural functions in response to plasma treatment, potentially reflecting cellular processes and organization alterations.

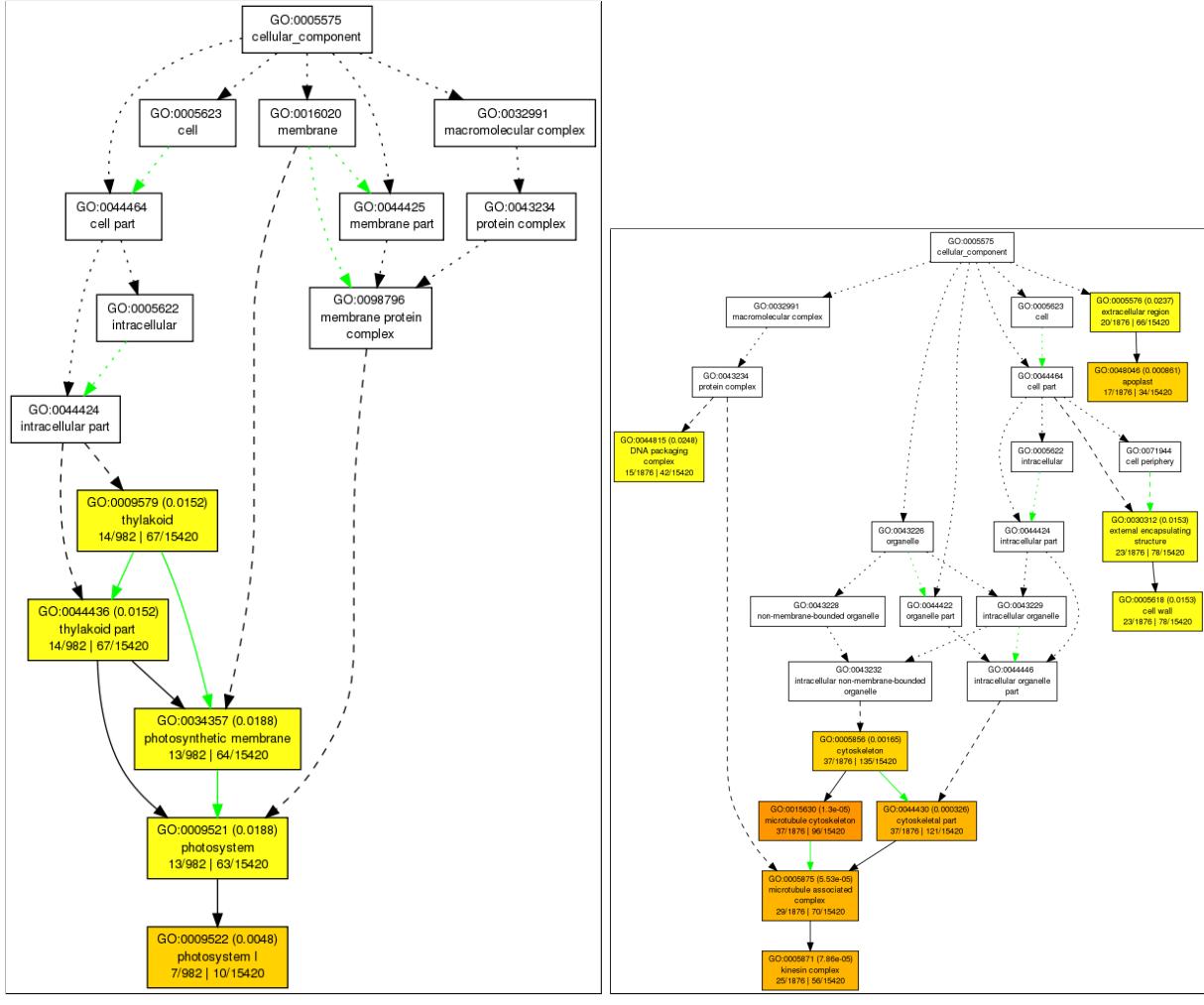


Figure 7: agriGO Cellular Component Ontology of (a) Up-regulated and (b) Down-regulated Genes.

Figure 7 illustrates the cellular components associated with differentially expressed genes between Seed WT and Seed Plasma conditions, as identified through agriGO analysis (T. Tian et al. 2017).

Subfigure (a) displays the cellular component graph of up-regulated genes, highlighting key GO terms related to photosynthetic structures and membranes. Notable terms include "photosystem I," "thylakoid," and "photosynthetic membrane," indicating the enrichment of components involved in photosynthesis and energy production in response to plasma treatment.

Conversely, Subfigure (b) presents the cellular component graph of down-regulated genes, revealing GO terms associated with cytoskeletal elements and extracellular structures. Terms such as "microtubule cytoskeleton" and "cell wall" suggest alterations in cytoskeletal organization and cell wall composition in response to plasma treatment.

3.2.2 g:Profiler



Figure 8: g:Profiler g:GOST Result of Up-regulated Genes.

Figure 8 presents the functional enrichment analysis results performed using the g:Profiler g:GOST tool on the up-regulated genes between Seed WT and Seed Plasma conditions. The analysis identified several significantly enriched GO terms associated with diverse biological processes, molecular functions, and cellular components (Reimand et al. 2007).

Among the top enriched terms, "DNA-binding transcription factor activity," "ubiquitin protein ligase activity," and "chitin binding" highlight the regulatory and protein modification functions potentially involved in response to plasma treatment. Additionally, "response to temperature stimulus" and "response to hypoxia" suggest adaptive responses to environmental cues induced by plasma treatment.

Notably, the enrichment of terms such as "proteasome-mediated ubiquitin-dependent protein" and "oxoacid metabolic process" underscores the involvement of protein degradation and metabolic processes in the cellular response to plasma treatment.



Figure 9: g:Profiler g:GOSt Result of Down-regulated Genes.

Figure 9 displays the functional enrichment analysis results conducted using the g:Profiler g:GOSt tool on the down-regulated genes between Seed WT and Seed Plasma conditions. The analysis identified several significantly enriched GO terms associated with various biological processes, molecular functions, and cellular components (Reimand et al. 2007).

Among the top enriched terms, "catalytic activity," "structural constituent of chromatin," and "heme binding" highlight functional roles related to enzymatic activities and chromatin organization. Additionally, terms such as "cell wall organization or biogenesis," "mitotic cell cycle," and "secondary metabolite biosynthetic process" indicate perturbations in cellular processes and metabolic pathways associated with plasma treatment-induced gene downregulation.

Enriching terms such as "extracellular region" and "microtubule cytoskeleton" suggest cellular organization and extracellular matrix composition alterations in response to plasma treatment.

3.2.3 PlantRegMap

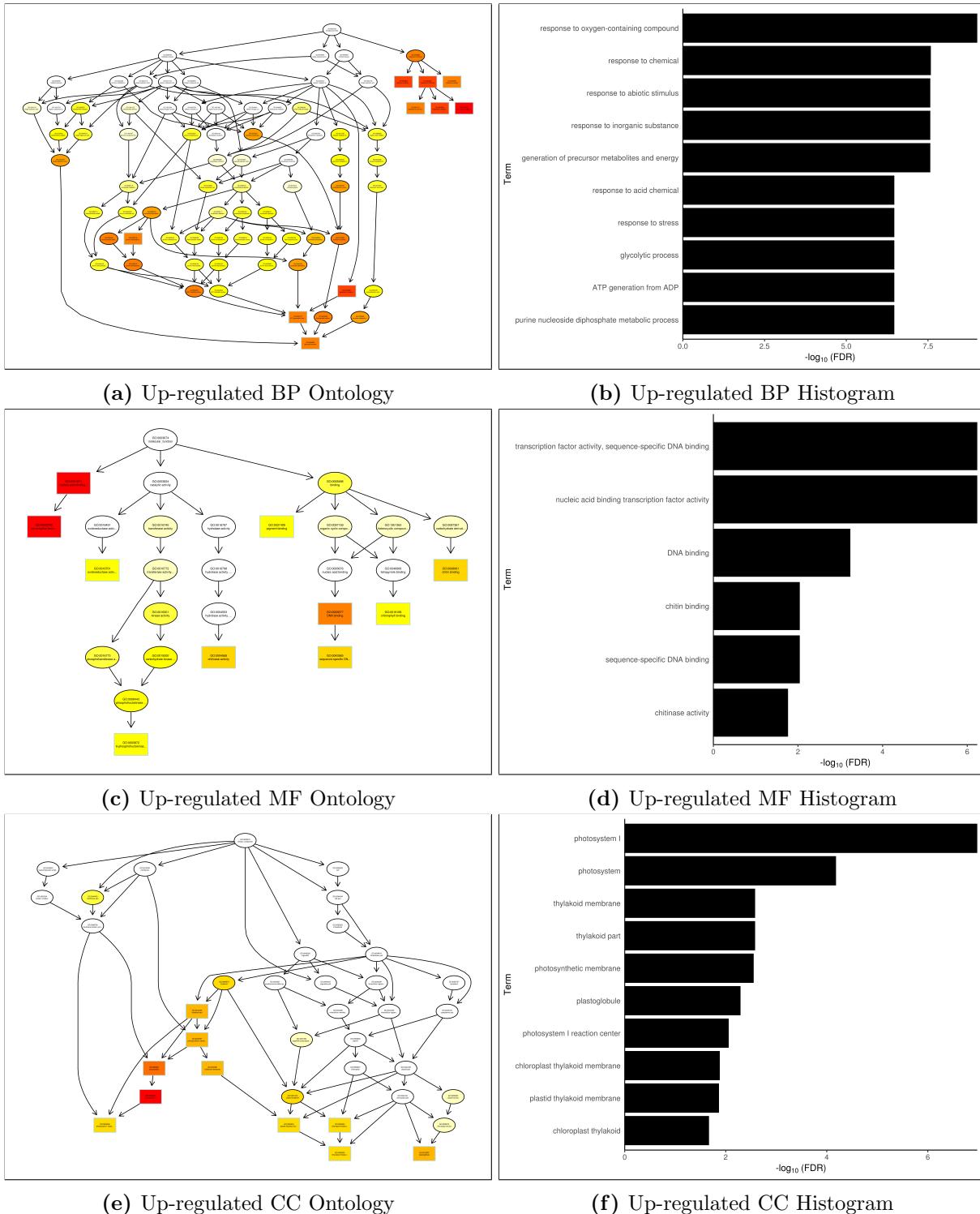


Figure 10: PlantRegMap Up-regulated Genes' (a) Biological Process Ontology, (b) Biological Process Histogram, (c) Molecular Function Ontology, (d) Molecular Function Histogram, (e) Cellular Component Ontology, and (f) Cellular Component Histogram.

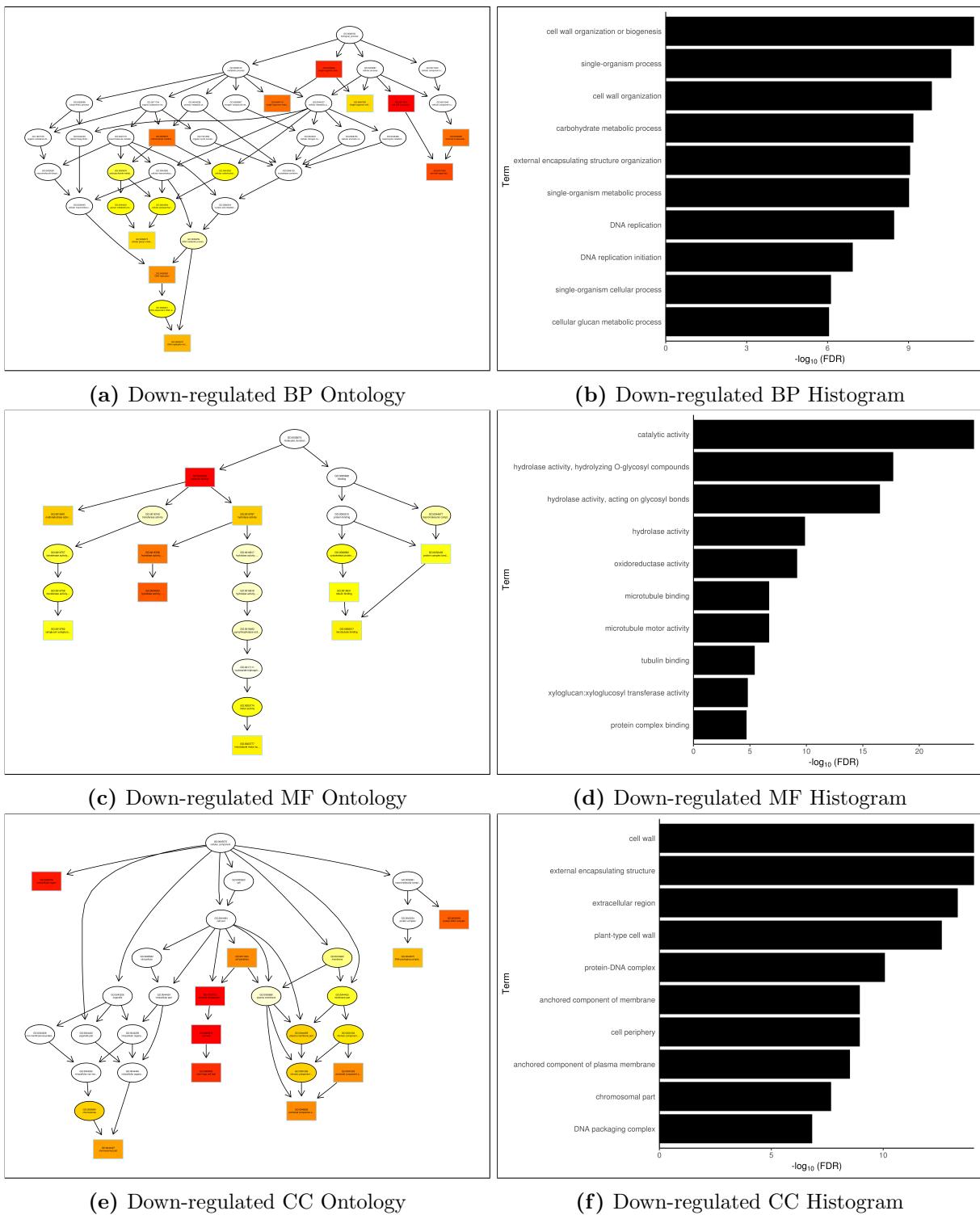


Figure 11: PlantRegMap Down-regulated Genes' (a) Biological Process Ontology, (b) Biological Process Histogram, (c) Molecular Function Ontology, (d) Molecular Function Histogram, (e) Cellular Component Ontology, and (f) Cellular Component Histogram.

Figure 10 presents the functional analysis results of PlantRegMap on the up-regulated genes between Seed WT and Seed Plasma conditions. Subfigures (a), (c), and (e) depict the Biological Process Ontology, Molecular Function Ontology, and Cellular Component Ontology, respectively, highlighting the enriched GO terms associated with different aspects of gene function and localization (F. Tian et al. 2020).

Notable GO terms identified in the histograms (Subfigures (b), (d), and (f)) include "response to oxygen-containing compound," "transcription factor activity," and "photosystem I," suggesting enrichment of processes related to environmental responses, transcriptional regulation, and photosynthesis in response to plasma treatment. These findings provide insights into the functional roles and cellular localization of up-regulated genes in the studied organism under plasma treatment conditions.

Similarly, Figure 11 illustrates the functional analysis results conducted using PlantRegMap on the down-regulated genes between Seed WT and Seed Plasma conditions. Subfigures (a), (c), and (e) display the Biological Process Ontology, Molecular Function Ontology, and Cellular Component Ontology, respectively, highlighting the enriched GO terms associated with various biological processes, molecular functions, and cellular components (F. Tian et al. 2020).

Notable GO terms identified in the histograms (Subfigures (b), (d), and (f)) include "cell wall organization or biogenesis," "catalytic activity," and "cell wall," suggesting perturbations in processes related to cell wall organization, enzymatic activities, and cellular localization in response to plasma treatment. These findings provide insights into the functional roles and cellular localization of down-regulated genes in the studied organism under plasma treatment conditions.

3.3 eFP Browser

3.3.1 TPM Correlation Analysis

To assess the reliability of the transcript abundance estimates obtained from RNA-Seq data, a correlation analysis was performed comparing the TPM (Transcripts Per Million) values derived from the RNA-Seq experiment with the databased TPM values available in the eFP Browser's database (Wagner, Kin, and Lynch 2012).

	seed_dry_1	seed_dry_2	seed_dry_3	mean_tpm
databased_seed_dry_1	0.9053431	0.9104441	0.8561029	0.9101076
databased_seed_dry_2	0.8524849	0.8229024	0.7690464	0.8346996
databased_seed_dry_3	0.8700211	0.8568043	0.8022809	0.8624846
databased_mean_tpm	0.8797852	0.8668061	0.8122952	0.8726687

Table 3: Correlation Coefficient Matrix of TPM and Databased TPM.

The correlation analysis revealed strong positive correlations between the TPM values of samples from the RNA-Seq experiment and the databased TPM values. For instance, the correlation coefficients between samples within the RNA-Seq experiment ranged from approximately 0.77 to 0.91, indicating high consistency in transcript abundance estimates

across replicates. Furthermore, the correlation coefficients between the RNA-Seq-derived TPM values and the databased TPM values ranged from approximately 0.80 to 0.91, suggesting a high degree of concordance between the two datasets.

Overall, the TPM correlation analysis underscores the reliability of the transcript abundance estimates obtained from the RNA-Seq experiment and validates the consistency of gene expression patterns observed across samples.

3.3.2 eFP Browser Visualization

Utilizing the Sorghum eFP Browser, gene expression patterns across different conditions were visually represented (Winter et al. 2007). By searching for *Sorghum bicolor* gene IDs, users can observe the gene expression levels depicted by a colour gradient ranging from red to yellow. In this representation, red signifies high gene expression levels, while yellow indicates low or negligible gene expression.

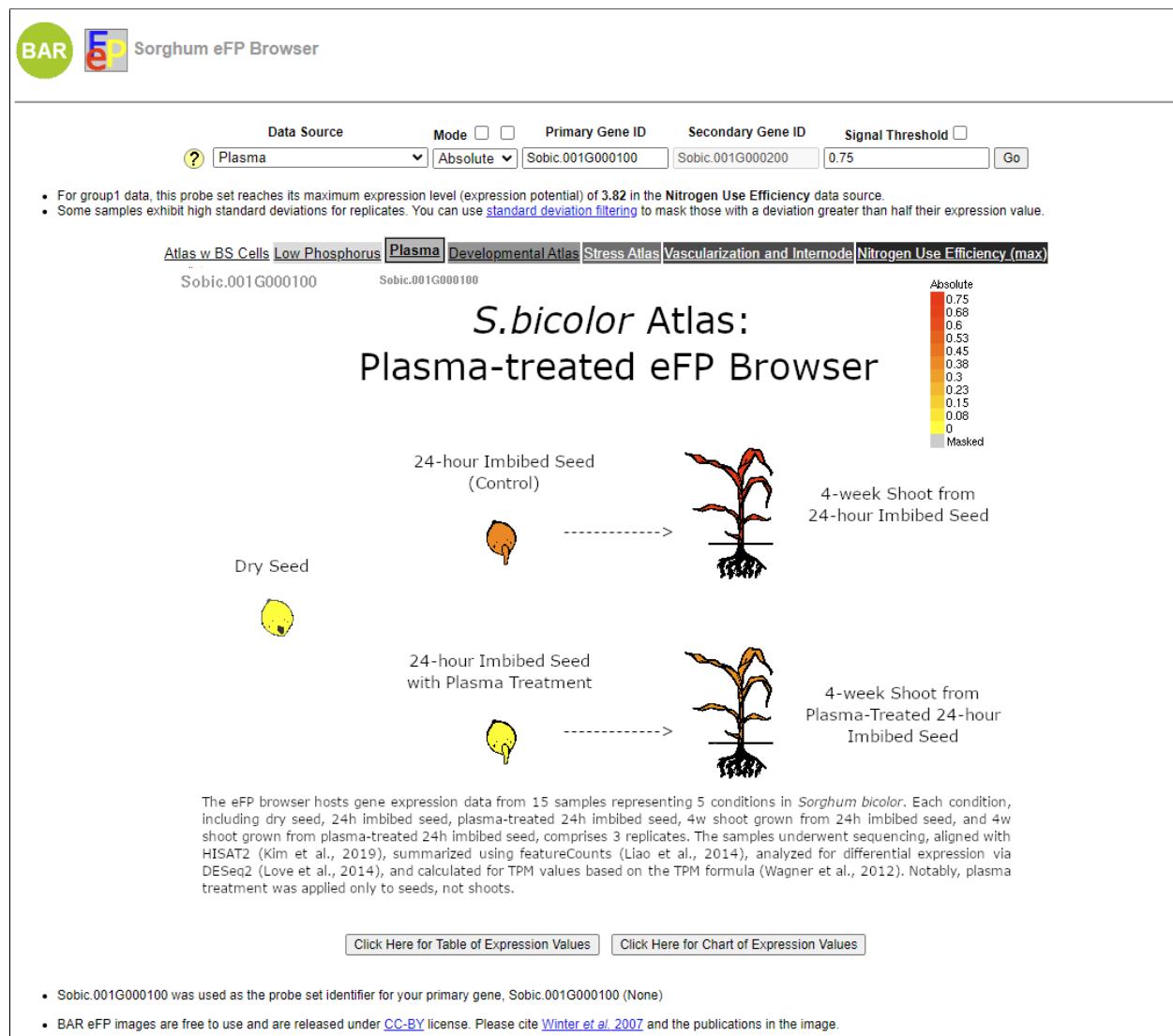


Figure 12: Sorghum eFP Browser Plasma Atlas.

The Sorghum eFP Browser Plasma Atlas comprehensively visualizes gene expression patterns in response to plasma treatment. Through interactive exploration, users can gain insights into the spatial and temporal dynamics of gene expression changes induced by plasma treatment across the *Sorghum bicolor* genome.

4 Discussion

The current study employed RNA-Seq analysis coupled with bioinformatics tools to explore the transcriptional response of *Sorghum bicolor* to plasma treatment, revealing dynamic changes in gene expression patterns across various conditions and developmental stages.

4.1 Differential Gene Expression Analysis

Our analysis unveiled substantial alterations in gene expression profiles, notably identifying a significant number of up and down-regulated genes in response to plasma treatment compared to the wild-type condition. These findings offer insights into the molecular mechanisms underlying the biological effects of plasma treatment on *Sorghum bicolor*.

4.2 Functional Analysis

Functional enrichment analysis using agriGO, g:Profiler, and PlantRegMap elucidated the biological processes, molecular functions, and cellular components associated with the differentially expressed genes. Enrichment of processes related to photosynthesis, response to environmental stimuli, and metabolic pathways underscored the multifaceted nature of the plant's response to plasma treatment. Additionally, the down-regulation of genes involved in cell wall organization and carbohydrate metabolism suggests the potential remodelling of cell structure and metabolism in response to plasma treatment.

4.3 eFP Browser

Using the Sorghum eFP Browser, visualizing gene expression patterns provided valuable insights into the spatial and temporal dynamics of gene expression changes induced by plasma treatment. Observing gene expression levels in different tissues and developmental stages enhances our understanding of the systemic response of *Sorghum bicolor* to plasma treatment.

4.4 Implications and Future Directions

These findings contribute to understanding the molecular mechanisms underlying *Sorghum bicolor*'s response to plasma treatment. Insights gained could improve crop resilience and productivity under various environmental conditions. Future research may focus on elucidating specific regulatory networks and signalling pathways involved in response to plasma treatment, as well as exploring applications in agricultural practices to enhance crop growth and stress tolerance.

4.5 Limitations

Acknowledging the limitations of this study is crucial. While RNA-Seq analysis provides valuable insights into gene expression patterns, further experimental validation is necessary to confirm the functional significance of identified differentially expressed genes. Additionally, our analysis focused on specific conditions and developmental stages, warranting further studies encompassing a broader range of conditions and time points for a comprehensive understanding of the transcriptional response to plasma treatment in *Sorghum bicolor*.

4.6 Conclusion

In conclusion, our study sheds light on the transcriptional response of *Sorghum bicolor* to plasma treatment, revealing dynamic changes in gene expression patterns. The comprehensive analysis provided valuable insights into the molecular mechanisms underlying the response to plasma treatment, offering potential implications for agricultural practices and crop improvement strategies.

References

- Allen, Bryce et al. (2012). “Software as a service for data scientists.” In: *Communications of the ACM* 55.2, pp. 81–88.
- Attri, P et al. (2020). “Plasma Agriculture from Laboratory to Farm: A Review. Processes 2020, 8, 1002.” In: *Processes* 8.1002.
- Attri, Pankaj et al. (2021). “Impact of atmospheric pressure plasma treated seeds on germination, morphology, gene expression and biochemical responses.” In: *Japanese journal of applied physics* 60.4, p. 040502.
- Binder, Joseph B and Ronald T Raines (2010). “Fermentable sugars by chemical hydrolysis of biomass.” In: *Proceedings of the National Academy of Sciences* 107.10, pp. 4516–4521.
- Chen, Yuxin et al. (2018). “SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data.” In: *Gigascience* 7.1, gix120.
- Conesa, Ana et al. (2016). “A survey of best practices for RNA-seq data analysis.” In: *Genome biology* 17.1, pp. 1–19.
- Cutler, Adrian J et al. (1997). “Induction of (+)-abscisic acid 8’ hydroxylase by (+)-abscisic acid in cultured maize cells.” In: *Journal of Experimental Botany* 48.10, pp. 1787–1795.
- Danecek, Petr et al. (Feb. 2021). “Twelve years of SAMtools and BCFtools.” In: *GigaScience* 10.2. giab008. ISSN: 2047-217X. DOI: 10.1093/gigascience/giab008. eprint: <https://academic.oup.com/gigascience/article-pdf/10/2/giab008/36332246/giab008.pdf>. URL: <https://doi.org/10.1093/gigascience/giab008>.
- Federhen, Scott (2012). “The NCBI taxonomy database.” In: *Nucleic acids research* 40.D1, pp. D136–D143.
- Foster, Ian (2011). “Globus Online: Accelerating and democratizing science through cloud-based services.” In: *IEEE Internet Computing* 15.3, pp. 70–73.
- Goodstein, David M et al. (2012). “Phytozome: a comparative platform for green plant genomics.” In: *Nucleic acids research* 40.D1, pp. D1178–D1186.
- Kanehisa, Minoru and Susumu Goto (2000). “KEGG: kyoto encyclopedia of genes and genomes.” In: *Nucleic acids research* 28.1, pp. 27–30.
- Kim, Daehwan et al. (2019). “Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype.” In: *Nature biotechnology* 37.8, pp. 907–915.
- Krochko, Joan E et al. (1998). “(+)-Abscisic acid 8’-hydroxylase is a cytochrome P450 monooxygenase.” In: *Plant physiology* 118.3, pp. 849–860.
- Liao, Yang, Gordon K Smyth, and Wei Shi (2014). “featureCounts: an efficient general purpose program for assigning sequence reads to genomic features.” In: *Bioinformatics* 30.7, pp. 923–930.
- (2019). “The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads.” In: *Nucleic acids research* 47.8, e47–e47.
- Loken, Chris et al. (2010). “SciNet: lessons learned from building a power-efficient top-20 system and data centre.” In: *Journal of Physics: Conference Series*. Vol. 256. 1. IOP Publishing, p. 012026.
- Love, Michael I, Wolfgang Huber, and Simon Anders (2014). “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.” In: *Genome biology* 15.12, pp. 1–21.

- Mathur, Supriya et al. (2017). "Sweet sorghum as biofuel feedstock: recent advances and available resources." In: *Biotechnology for Biofuels* 10, pp. 1–19.
- McCormick, Ryan F et al. (2018). "The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization." In: *The Plant Journal* 93.2, pp. 338–354.
- Oikawa, Tetsuo et al. (2004). "A role of OsGA20ox1, encoding an isoform of gibberellin 20-oxidase, for regulation of plant stature in rice." In: *Plant molecular biology* 55, pp. 687–700.
- Pearson, Karl (1901). "LIII. On lines and planes of closest fit to systems of points in space." In: *The London, Edinburgh, and Dublin philosophical magazine and journal of science* 2.11, pp. 559–572.
- Pertea, Mihaela et al. (2016). "Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown." In: *Nature protocols* 11.9, pp. 1650–1667.
- Qin, Xue et al. (2013). "Gibberellin 20-oxidase gene OsGA20ox3 regulates plant stature and disease development in rice." In: *Molecular plant-microbe interactions* 26.2, pp. 227–239.
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.
- Reimand, Jüri et al. (2007). "g: Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments." In: *Nucleic acids research* 35.suppl_2, W193–W200.
- Robinson, James T et al. (2011). "Integrative genomics viewer." In: *Nature biotechnology* 29.1, pp. 24–26.
- Robinson, Mark D, Davis J McCarthy, and Gordon K Smyth (2010). "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data." In: *bioinformatics* 26.1, pp. 139–140.
- Souza, Paula Monteiro de et al. (2010). "Application of microbial α -amylase in industry-A review." In: *Brazilian journal of microbiology* 41, pp. 850–861.
- Suriyasak, Chetphilin et al. (2021). "Alterations of DNA methylation caused by cold plasma treatment restore delayed germination of heat-stressed rice (*Oryza sativa L.*) seeds." In: *ACS Agricultural Science & Technology* 1.1, pp. 5–10.
- Thomas, Paul D et al. (2022). "PANTHER: Making genome-scale phylogenetics accessible to all." In: *Protein Science* 31.1, pp. 8–22.
- Tian, Feng et al. (2020). "PlantRegMap: charting functional regulatory maps in plants." In: *Nucleic acids research* 48.D1, pp. D1104–D1113.
- Tian, Tian et al. (2017). "agriGO v2. 0: a GO analysis toolkit for the agricultural community, 2017 update." In: *Nucleic acids research* 45.W1, W122–W129.
- Trapnell, Cole et al. (2010). "Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation." In: *Nature biotechnology* 28.5, pp. 511–515.
- Wagner, Günter P, Koryu Kin, and Vincent J Lynch (2012). "Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples." In: *Theory in biosciences* 131, pp. 281–285.
- Winter, Debbie et al. (2007). "An ‘Electronic Fluorescent Pictograph’ browser for exploring and analyzing large-scale biological data sets." In: *PloS one* 2.8, e718.