

Data mining for pathogenic and symbiotic plant microbe interaction signatures using co-expression networks

Wissenschaftliche Arbeit zur Erlangung des Grades M.Sc. an TUM-School of Life Sciences der
Technischen Universität München.

Betreuer/-in

Dr. Amit Fenn
Lehrstuhl für Computational Plant Biology

Aufgabensteller/-in

Prof. Dr. rer. nat. Nadia Kamal
Lehrstuhl für Computational Plant Biology

Eingereicht von

Vaishnavi Mohanadas Nair
Fritz Walter straße 5F
85375 Neufahrn bei freising
+49 15510479427

Eingereicht am

Freising, den 15.07.2025

Statutory Declaration

I confirm that this Master's thesis is my own work and I have documented all sources and materials used.

This thesis was not previously presented to another examination board and has not been published.

Neufahrn, 15.07.2025

Ort, Datum

A handwritten signature in black ink, slanted upwards from left to right, reading "Vaishnavi. M".

Unterschrift

Table of Contents

Statutory Declaration	2
List of Abbreviation	6
List of figures	7
List of tables	9
List of equations	9
1. Abstract.....	10
2. Introduction.....	11
a. Background on <i>Solanum lycopersicum</i>	11
b. Plant-microbe interaction.....	11
i. Defense mechanisms activated by microbes.....	11
ii. Types of microbes and its interaction with Tomato within this study	12
c. Abiotic stress responses included within this study	13
d. Weighted correlation network analysis.....	13
e. Gene network inference with Ensemble of trees	14
3. Aim of the Thesis	16
4. Methods	17
a. Dataset collection	18
b. Quality control and preprocessing	18
i. Fastp and FastQC	18
c. Transcript quantification	19
i. Kallisto	19
d. Normalization	21
i. DESeq2 & VST	21
e. Co-expression Network analysis	21
i. WGCNA	21
ii. Signed hybrid & Soft threshold	22
iii. Module detection using hierarchical clustering.....	22
f. Regulatory network analysis	22
i. GENEI3	22
ii. Louvain clustering & Resolution	23
iii. Cosine similarity & Cosine threshold.....	23
g. Functional enrichment	23
i. Mercator and clusterProfiler	23
5. Results	25
a. Principal Component Analysis (PCA).....	25
i. PCA plotted using TPM values	25
ii. PCA with tissue labels	27
iii. PCA plotted using VST data.....	29
b. Co-expression modules	30
i. Overview of all the plots generated through WGCNA.....	30

ii.	Sample clustering	31
iii.	Soft-thresholding power selection	32
iv.	Merged modules	34
v.	Module-trait associations and eigengene expression across interaction types ...	35
vi.	Gene significance plot	39
vii.	Module-trait relationships and module eigengene expression across microbes .	48
viii.	Functional enrichment of WGCNA	53
c.	Regulatory modules	64
i.	Overview of all the plots generated through GENIE3.....	64
ii.	Cosine similarity	65
iii.	Functional enrichment of GENIE3	66
6.	Discussion.....	73
a.	Limitations and Future work	73
i.	Data variability and experimental constraints	73
ii.	Limitations of WGCNA.....	73
iii.	Limitations of GENIE3.....	74
iv.	Louvain clustering.....	74
v.	Limitations of using cosine similarity.....	74
vi.	Advantage of implementing WGCNA and GENIE3	75
b.	Interpretation of key findings.....	76
i.	Biological interpretation of enriched pathways	76
ii.	Role of TF found in regulatory modules	76
7.	Conclusion	77
8.	Acknowledgements.....	77
9.	References	78
Copyright license	85	
Supplementary	85	
i.	Metadata containing the accession numbers.	85

List of Abbreviation

Pathogen-Associated Molecular Pattern	PAMP
PAMP- Triggered Immunity	PTI
Transcription Factors	TF
Resistance	R
Effector Triggered Immunity	ETI
Hypersensitive Response	HR
Pattern Recognition Receptor	PRR
Arbuscular Mycorrhizal Fungi	AMF
Cucumber Mosaic Virus	CMV
Serendipita Indica	SI
Funneliformis Mosseae	FM
Cladosporium Fulvum	CF
Meloidogyne incognita	MI
Oidium Neolycopersici	ON
Botrytis cinerea	BC
Drought	D
Gene network inference with ensemble of trees	GENIE3
Genetic regulatory network	GRN
Sequence read archive	SRA
Transcripts per million	TPM
Variance stabilizing transformation	VST
Topological overlap measure	TOM
Module eigengene	ME
Module-Trait Relation	MTR
Principal component analysis	PCA
Gene significance	GS
Module membership	MM
Reactive oxygen species	ROS
Gene ontology	GO

List of Figures

FIGURE 1 PLANT IMMUNE SYSTEM - PTI AND ETI	12
FIGURE 2 BIOINFORMATICS PIPELINE USED IN THIS STUDY.....	17
FIGURE 3 KALLISTO ALIGNMENT SUMMARY OF 86 FASTQ FILES.....	19
FIGURE 4 KALLISTO ALIGNMENT SUMMARY OF 73 FASTQ FILES.....	20
FIGURE 5 PCA OF TPM UNFILTERED.....	26
FIGURE 6 PCA OF TPM FILTERED.	27
FIGURE 7 PCA OF TISSUE SAMPLE (PC1 AND 2).	28
FIGURE 8 PCA OF TISSUE SAMPLE (PC 3 AND 4).....	29
FIGURE 9 PCA OF VST DATA.	30
FIGURE 10 SAMPLE CLUSTERING DENDROGRAM.....	32
FIGURE 11 SCALE INDEPENDENCE PLOT.	33
FIGURE 12 MEAN CONNECTIVITY.	33
FIGURE 13 GENE DENDROGRAM AND MODULE COLORS.	34
FIGURE 14 MODULE-TRAIT RELATIONSHIP WITH DIFFERENT TYPES OF INTERACTIONS.	36
FIGURE 15 DENDROGRAM AND HEATMAP OF MODULE EIGENGENE AND INTERACTION TYPE.....	38
FIGURE 16 SCATTER PLOT OF MM vs GS IN M _E RED MODULE.	40
FIGURE 17 SCATTER PLOT OF MM vs GS IN M _E GREEN MODULE.....	41
FIGURE 18 SCATTER PLOT OF MM vs GS IN M _E YELLOWGREEN MODULE.	42
FIGURE 19 SCATTER PLOT OF MM vs GS IN M _E BROWN4 MODULE.....	43
FIGURE 20 SCATTER PLOT OF MM vs GS IN M _E TURQUOISE MODULE.....	43
FIGURE 21 SCATTER PLOT OF MM vs GS IN M _E BLACK MODULE.	44
FIGURE 22 SCATTER PLOT OF MM vs GS IN M _E LIGHTSTEELBLUE MODULE.	45
FIGURE 23 SCATTER PLOT OF MM vs GS IN M _E PALEVIOLETRED3 MODULE.....	46
FIGURE 24 SCATTER PLOT OF MM vs GS IN M _E ORANGERED3 MODULE.	47
FIGURE 25 SCATTER PLOT OF MM vs GS IN M _E BISQUE4 MODULE.	48
FIGURE 26 A MODULE TRAIT RELATIONSHIP PLOT FOR DIFFERENT MICROBES.....	51
FIGURE 27 DENDROGRAM AND HEATMAP OF MODULE EIGENGENE AND MICROBES.	52
FIGURE 28 FUNCTIONAL ENRICHMENT OF RED MODULE (WGCNA).	54
FIGURE 29 FUNCTIONAL ENRICHMENT OF GREEN MODULE (WGCNA).	55
FIGURE 30 FUNCTIONAL ENRICHMENT OF YELLOWGREEN MODULE (WGCNA).	56
FIGURE 31 FUNCTIONAL ENRICHMENT OF BROWN4 MODULE (WGCNA).	57
FIGURE 32 FUNCTIONAL ENRICHMENT OF BLACK MODULE (WGCNA).	58
FIGURE 33 FUNCTIONAL ENRICHMENT OF LIGHTSTEELBLUE MODULE (WGCNA).	59
FIGURE 34 FUNCTIONAL ENRICHMENT OF PALEVIOLETRED3 MODULE (WGCNA).	61
FIGURE 35 FUNCTIONAL ENRICHMENT OF ORANGERED3 MODULE (WGCNA).	62
FIGURE 36 FUNCTIONAL ENRICHMENT OF BISQUE4 MODULE (WGCNA).	63
FIGURE 37 FUNCTIONAL ENRICHMENT OF TURQUOISE MODULE (WGCNA).	64
FIGURE 38 COSINE SIMILARITY COMPUTED BETWEEN WGCNA MODULES AND GENIE3-LOUVAIN CLUSTERS.....	66
FIGURE 39 FUNCTIONAL ENRICHMENT OF BROWN4 MODULE (GENIE3).	67

FIGURE 40 FUNCTIONAL ENRICHMENT OF ORANGERED3 MODULE (GENIE3)	68
FIGURE 41 FUNCTIONAL ENRICHMENT OF LIGHTSTEELBLUE MODULE (GENIE3).....	69
FIGURE 42 FUNCTIONAL ENRICHMENT OF TURQUOISE MODULE (GENIE3).....	70
FIGURE 43 FUNCTIONAL ENRICHMENT OF RED MODULE (GENIE3)	71
FIGURE 44 FUNCTIONAL ENRICHMENT OF GREEN MODULE (GENIE3)	72

List of Tables

TABLE 1 SOLANUM LYCOPERSICUM INFECTED WITH DIFFERENT MICROBES.....	18
TABLE 2 DIFFERENT MODULES THAT ARE CORRELATED WITH THEIR CORRESPONDING MICROBES.	49
TABLE 3 FIGURE LICENSES	85
TABLE 4 METADATA CONTAINING THE ACCESSION NUMBERS.....	85

List of Equations

EQUATION 1 ADJACENCY FUNCTION	14
-------------------------------------	----

1. Abstract

Tomato (*Solanum lycopersicum*) serves as an important crop model for understanding plant-microbe interactions due to its economic significance and rich genomic resources. This study applies systems level transcriptomic approach to uncover the molecular networks underlying tomato responses to a range of microbial conditions, including biotrophic (fungal, viral, nematode), mutualistic (arbuscular and root endophytic fungi), necrotrophic (fungal) and abiotic stress (drought). A total of 73 RNA-seq samples were collected from public repositories, preprocessed and variance-stabilized for downstream analysis.

To understand how to build gene regulatory networks, two computational frameworks were employed independently: Weighted Gene Co-expression Network Analysis (WGCNA) to identify co-expressed gene modules, and Gene network inference with ensemble of trees (GENIE3), a Random forest-based method to infer directed gene regulatory networks. WGCNA identified 39 distinct gene modules, several of which showed specific or shared correlations with microbial interactions. For example, in this study a module was uniquely associated with *Oidium neoly copersici* infection and enriched for redox-related glutaredoxin activity and there were other modules linked to arbuscular mycorrhizal symbiosis and enriched in photosynthetic and nutrient exchange pathways. In contrast yet another gene modules appeared to mediate generalized stress responses across both beneficial and pathogenic conditions. GENIE3 based regulator inference highlighted transcription factors such as bZIPs, NACs, WRKYs, and AP2/ERFs as key regulators of modules associated with photosynthesis, stress signaling and defense.

To compare the gene groups from WGCNA and GENIE3, a method called cosine similarity was used. This showed that while there was some overlap, each method also finds unique gene clusters. This thesis explores how both methods are useful and work better in synchrony.

Overall, this study shows that combining WGCNA and GENIE3 can help us better understand how tomato plants respond to different microbes. It also found some genes that are not yet well understood, which could be interesting for future lab experiments. This approach can also be used to study other crops and stress conditions.

2. Introduction

a. Background on *Solanum lycopersicum*

Tomato (*Solanum lycopersicum*) is a widely cultivated horticultural crop with significant agricultural and nutritional value. As an herbaceous perennial, it offers many advantages for physiological, genetic, and cytological research due to its relatively simple growth requirements and amenability to laboratory culture conditions. These characteristics make tomato a preferred model system for plant biology studies, particularly in dicotyledonous species (Rick, 1980).

Moreover, its fully sequenced genome and the availability of genetic tools have established tomato as one of the best-studied dicot plants, enabling researchers to explore complex biological processes such as stress responses, fruit development, and host-microbe interaction (Bai & Lindhout, 2007). These properties make it an excellent system for studying how plants respond to different environmental cues, including biotic and abiotic stress.

b. Plant-microbe interaction

i. Defense mechanisms activated by microbes.

Regardless of the varieties of tomatoes, they have a sweet taste and as a result of that richness in carbohydrates in a matrix of free water, they become susceptible to pathogen attacks(Colmán et al., 2018). Due to this fact, many plants, including tomatoes, have developed complex mechanisms to protect themselves against pathogens. During the pathogen attack, the conserved molecular patterns are recognized by the plants (pathogen-associated molecular patterns (PAMPs)) at the pathogen cell surface level and that triggers the basal immune responses (PAMP-triggered immunity (PTI)). This effective two-way communication makes up the plants' immune system(Guttman et al., 2014) (refer Figure 1) which modulates regulatory proteins (such as transcription factors (TFs) and protein kinases) and pathogenesis related proteins(X. Yao et al., 2023). When the infection takes places the plants with the corresponding resistance (R) proteins either directly or indirectly recognize the effectors and start the immune response known as effector triggered immunity (ETI), including hypersensitive response (HR), in response to some pathogens secreting effectors to increase their pathogenicity into the host cells and to suppress PTI(Campos et al., 2022).

Effector-Triggered Immunity (ETI)

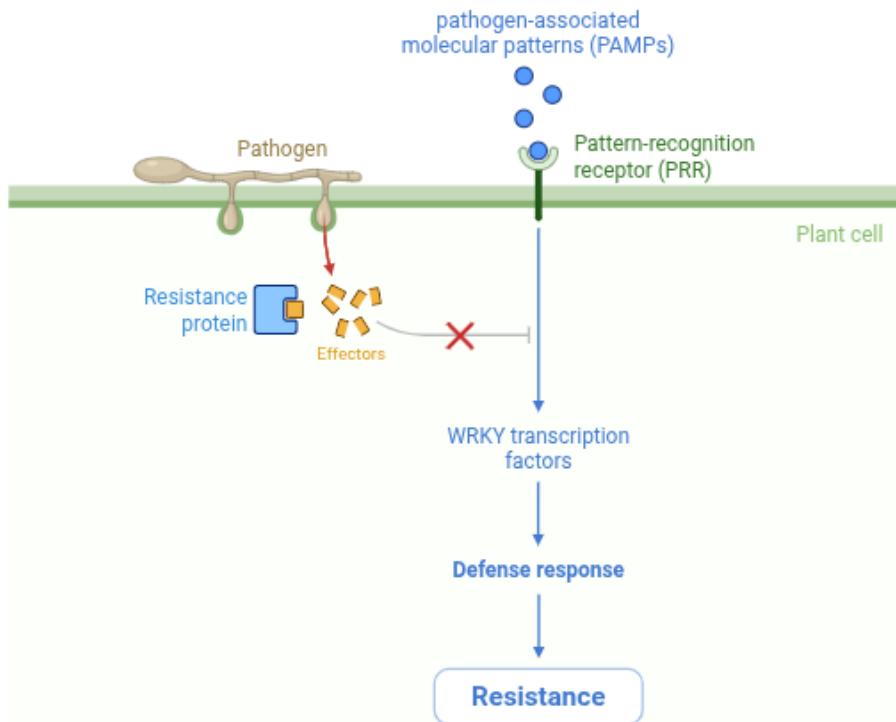


Figure 1 Plant Immune system - PTI and ETI

Figure 1 description: Pathogens are detected through surface localized pattern recognition receptors (PRRs) that recognize pathogen associated molecular patterns (PAMPs), initiating a PTI. In response to this, pathogens secrete effector proteins (Avr) to suppress PTI and manipulate host hormone signaling. R genes detect these effectors, activating effector-triggered immunity (ETI). This leads to rapid transcriptional activation of defense genes, often accompanied by chromatin modifications that enhance TF accessibility. Image created using Biorender template(Coelho et al., 2011).

ii. Types of microbes and its interaction with Tomato within this study

Before diving deeper, it's important to understand the main types of microbes that interact with plants. Some microbes, called **biotrophs**, feed on living plant cells without killing them(Mapuranga et al., 2022). They often form long-term relationships with their host and are typical of many fungal pathogens and viruses. In contrast, **necrotrophs** take a more aggressive approach — they kill plant cells and then feed on the dead tissue(De Cal et al., 2022). This usually leads to visible damage and decay, like what we see with grey mold caused by *Botrytis cinerea*. Then there are **symbiotic** microbes, such as a kind of fungi called arbuscular mycorrhizal fungi (AMF), which benefit the plant(Khalil et al., 2022). AMF are soil microorganisms that form a mutualistic association with the plant's root system(Singh et al., 2024). The symbiotic relationship is formed by penetrating the cortical cells of the roots and forming arbuscules, these are tree-like structures that further facilitate nutrient exchange between the fungus and the plant(Plouznikoff et al., 2019).

Further microbes explored for this study include, the root endophyte *Serendipita Indica* (*SI*) (formerly known as *Piriformospora indica*, order Sebacinales, phylum Basidiomycota) fungus is a root-colonizing endophytic symbiotic fungus. The fungus colonizes the root by penetrating the root epidermis and the cortex intracellularly without causing cell death(Varma et al., 1999). *Funneliformis mosseae* (FM) is an AM fungus forming mutualistic symbiosis with plant roots. It colonizes roots by forming hyphopodia on the root surface and grows intercellularly to form arbuscules inside cortical cells (Rodrigues & Rodrigues, 2015) (Miozzi et al., 2020). *Cucumber mosaic virus* (CMV) is a positive-sense single-stranded biotrophic RNA virus in the family *Bromoviridae*, genus *Cucumovirus*. CMV infects plants via aphid transmission and spreads systemically through plasmodesmata and phloem (Miozzi et al., 2020). *Cladosporium fulvum* (CF) is a biotrophic fungal pathogen that causes tomato leaf mold. It colonizes the apoplastic space of tomato leaves, avoiding host cell penetration and secreting effectors to suppress host defense(Jiang et al., 2022). *Meloidogyne incognita* (MI) is a biotrophic sedentary endoparasitic root-knot nematode. It penetrates the root tip, migrates to the vascular cylinder, and induces the formation of giant cells for feeding(Meidani et al., 2025). *Oidium neolycopersici* (ON) is an obligatory biotrophic fungus causing powdery mildew on tomatoes. It forms appressoria to penetrate host epidermal cells and grow epiphytically, extracting nutrients via haustoria(Jones et al., 2001). *Botrytis cinerea* (BC) is a necrotrophic fungal pathogen responsible for gray mold disease in many plant species. It infects by secreting enzymes and toxins that kill host cells, then colonizes the dead tissue(Borges et al., 2014).

c. Abiotic stress responses included within this study

Plants face a variety of environmental challenges, including both biotic and abiotic stressors. Abiotic stresses such as drought, salinity, and temperature extremes can significantly impact plant growth and productivity. At the same time, plants also encounter biotic threats from pathogens with diverse lifestyles, including necrotrophy that actively kill host tissue. Interestingly, the molecular response to these different stress types often overlaps. Many genes involved in abiotic stress responses are also activated during pathogen attack, which makes it difficult to isolate pathogen-specific pathways. To address this, drought (D) and necrotrophic pathogen samples were included in this study. By comparing gene expression across these conditions, the study aims to identify core stress-responsive pathways by using network-based tools like WGCNA and GENIE3 to distinguish genes and regulatory modules that are uniquely associated with biotrophic pathogen interactions.

d. Weighted correlation network analysis

Weighted Gene Co-expression Network Analysis (WGCNA) is a systems biology method that describes correlation patterns among genes across microarray samples and RNA-seq data. Functions in WGCNA can be divided into the following categories: 1. network construction; 2. module detection; 3. association of genes with sample traits. WGCNA is designed to model a network that is biologically representative, and this is approximated by scale free topology. A network is set to be scale free if the fraction of nodes with degree k follows a power-law distribution $k^{-\alpha}$, where $\alpha > 1$ (Broido & Clauset, 2019). In a WGCNA network, each gene is

represented as a node, and the connection between two genes based on their expression similarity is called an edge (Ko & Brandizzi, 2023).

WGCNA network is specified by its adjacency matrix a_{ij} , a symmetric $n \times n$ matrix, where a_{ij} encodes the strength of network connection between the nodes i and j . To calculate the adjacency matrix, an intermediate quantity called the co-expression similarity s_{ij} is defined, typically as the absolute Pearson correlation between gene profiles. An unweighted network arises from setting a “hard” threshold parameter where two genes are linked ($a_{ij} = 1$) if the absolute correlation between their expression profiles exceeds the hard threshold. While this is used widely, it doesn’t reflect the continuous nature of the underlying co-expression of information and may lead to information loss. In contrast, weighted network allows the adjacency to take continuous values between 0 and 1 and is defined by raising the co-expression similarity to a power termed as “soft” threshold,

$$a_{ij} = s_{ij}^{\beta}$$

Equation 1 Adjacency function

with $\beta \geq 1$ (Langfelder & Horvath, 2008)(Zhang & Horvath, 2005).

The next step in WGCNA is module detection. There are many possible ways for defining a network module proposed in various literature (Bar-Joseph et al., 2003) (Segal et al., 2003)(Xu et al., 2004). Modules are defined as clusters that result from using topological overlap-based dissimilarity as an input of average linkage hierarchical clustering. Branches in the resulting cluster tree (dendrogram) are referred to as modules (Langfelder & Horvath, 2007).

In this study, the co-expression network of genes connected to plant-microbe interaction was built using the WGCNA algorithm. The research will aid in the identification of novel or unique molecular pathways specific to that microbe interaction.

e. Gene network inference with Ensemble of trees

Gene network inference with ensemble of trees (GENIE3), is a computational approach that uses decision trees to analyze gene expression data and predict which genes regulate others. The prediction of a regulatory network between p genes is broken down into p independent regression issues by GENIE3. Using the tree-based ensemble approaches such as Random Forest or Extra-trees, the expression pattern of one gene (the target gene) is predicted from the expression patterns of the other genes (the input genes) in each regression issue(Huynh-Thu et al., 2010). Random forest is an ensemble algorithm based on learning a collection of decision trees. Each gene is ranked based on the measure of importance that results from the forest(Petralia et al., 2015)(Huynh-Thu & Geurts, 2018). The relative importance of each gene is quantified producing an edge score that reflects its potential regulatory influence on the target. GENIE3 can generate a ranked list of regulatory links but is unable to compute functional modules unlike WGCNA, hence Louvain clustering must be introduced.

Louvain clustering is a modularity-based community detection method which is suitable for large, sparse network (Blondel et al., 2008). It starts by creating communities of size 1 where each node in the network forms a community. The next step is then divided into two sub steps. In the first sub step, a node v is assigned to a community of a neighbor u , such that the modularity of the partition increased. This process is then repeated until the modularity can be improved. This sub step generates the initial partition. In the second sub step, each community within the partition is considered as supernode. If there is at least one edge connecting the nodes of each community that the supernodes represent, then the supernodes are linked. Once this step is concluded, the algorithm iterates and stops when the modularity cannot increase(Perrin & Zuccon, 2018)(Traag et al., 2019).

3. Aim of the Thesis

The aim of this thesis is to investigate the biological pathways and gene regulation mechanisms activated during plant-microbe interactions, with a particular focus on pathogenic and mutualistic relationships. By utilizing both co-expression networks and gene regulatory networks, the study seeks to identify common patterns and unique plant responses under various microbial interactions.

This work utilizes publicly available RNA-seq datasets across multiple studies involving different pathogens and environmental conditions. A central goal is to evaluate how scalable and integrative different network analysis tools are when applied to heterogeneous data sources. We aim to understand whether combining diverse datasets can lead to the identification of robust, potentially universal molecular responses to biotic stress such as those triggered by biotrophic and symbiotic interactions as well as abiotic stress factors.

The research is guided by the following questions:

- Can we identify specific co-expression modules with the correlation network that are consistently associated with biotrophic interactions across multiple datasets?
- Among the genes that are part of the pathogen associated co-regulated network, how many are also involved in the gene regulatory networks, indicating a deeper role in transcriptional control?

Ultimately, this thesis aims to contribute to the understanding how plants orchestrate complex gene expression programs in response to microbial interactions and how this knowledge can be generalized or refined across species and environmental contexts.

4. Methods

The complete bioinformatics pipeline used in this study is summarized in Figure 2. The workflow includes raw data retrieval, quality control, transcript quantification, normalization, co-expression network, gene regulatory network and functional enrichment. This visual representation provides a step-by-step overview of the analysis process, from raw FASTQ file to functional enrichment results.

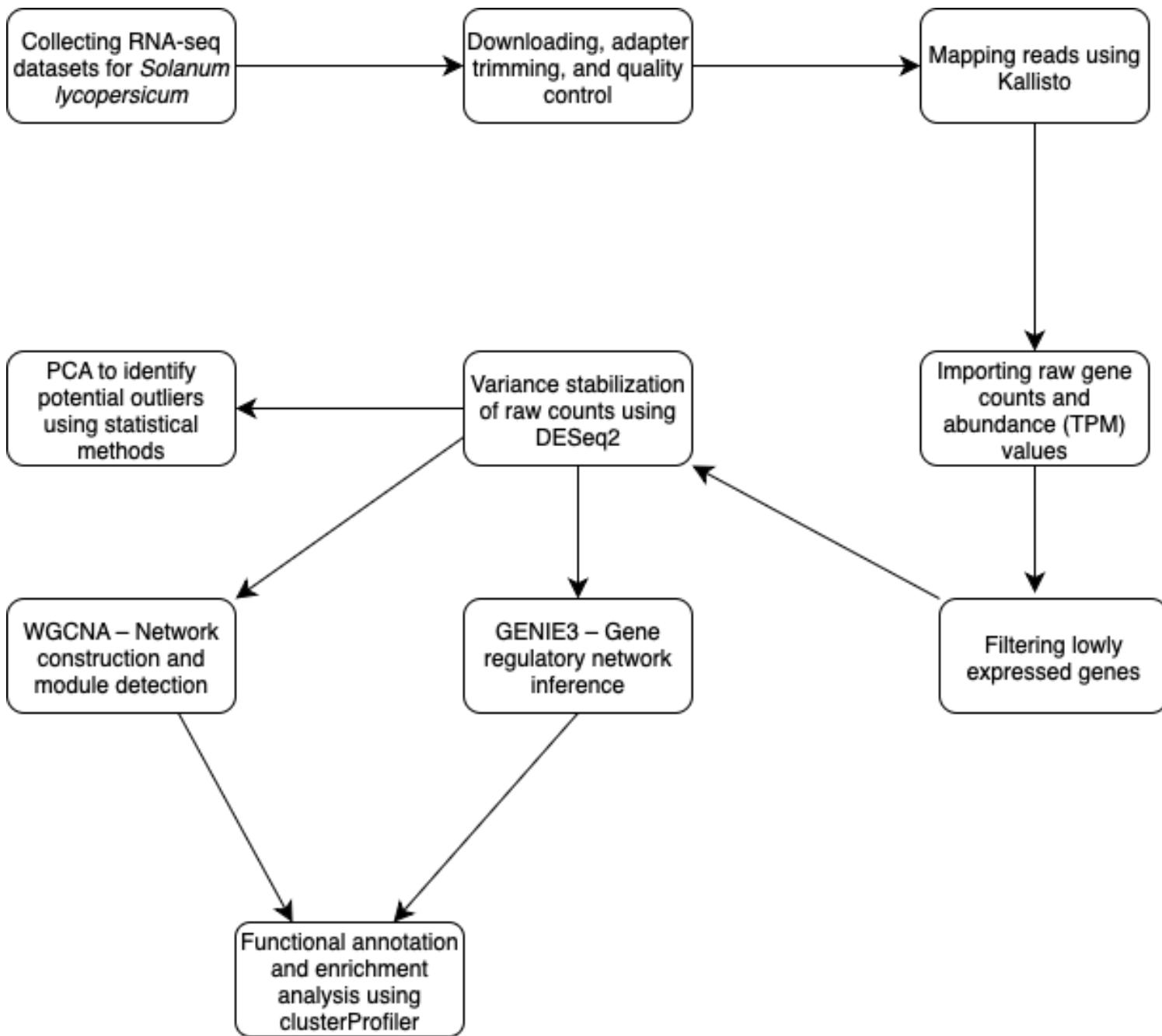


Figure 2 bioinformatics pipeline used in this study.

To promote transparency and reproducibility, all scripts, commands, and input files used for the analysis are available in LRZ Gitlab repository: [Gitlab link](#)

a. Dataset collection

A total of 86 SRA files were obtained for *solanum lycopersicum* infected with various microbes (refer Table 1) from sequence read archive (SRA). The dataset includes both paired-end and single-end read types. The SRA files were converted to fastq files with the help of SRA toolkit's (v3.2.1) Fasterq-dump command.

Table 1 Solanum lycopersicum infected with different microbes.

Sample	No. of FASTQ Interaction	Tissue	Pathogen
Arbuscular Mycorrhizal (AM)	6	Symbiotic	Root
Serendipita Indica (SI)	4	Symbiotic	Root and Leaf
Funneliformis Mosseae (FM)	3	Symbiotic	Leaf
Cucumber Mosaic Virus (CMV)	3	Biotrophic	Leaf
Cladosporium Fulvum (CF)	12	Biotrophic	Leaf
Meloidogyne Incognita (MI)	6	Biotrophic	Root
Oidium Neolycopersici-powdery mildew (ON)	6	Biotrophic	Leaf
Botrytis Cinerea (control) (BC)	9	Necrotrophic	Leaf
Drought stress (control) (D)	12	Drought	Leaf
			Abiotic stress

Table 1 description: List of microbes, number of fastq files for each sample, interaction type and plant tissue type are discussed here. All data obtained from NCBI SRA.

Among the different interactions studied, control samples are also included to unify pathway level interactions related to stress and plant death hormone. This allows for identification of both novel and or shared pathways among different plant-microbe interaction.

b. Quality control and preprocessing

i. Fastp and FastQC

Adapter trimming using Fastp (v0.26.0) command was carried out for the 86 FASTQ files. For the trimmed FASTQ files, quality control was carried out using FastQC (v0.11.8) and MultiQC (v1.6) to visualize the read quality before and after trimming.

c. Transcript quantification

i. Kallisto

The reference FASTA file for *solanum lycopersicum* was downloaded from Sol Genomics database(*Sol Genomics Network*, n.d.) which follows ITAG version 4. This FASTA file was then used to build the Kallisto index, which serves as the initial step of Kallisto alignment. Using the generated index, Kallisto (v0.51.1) was then run in the FASTQ files to perform transcript quantification. The results of the alignment were subsequently visualized using MultiQC reports (refer Figure 3).

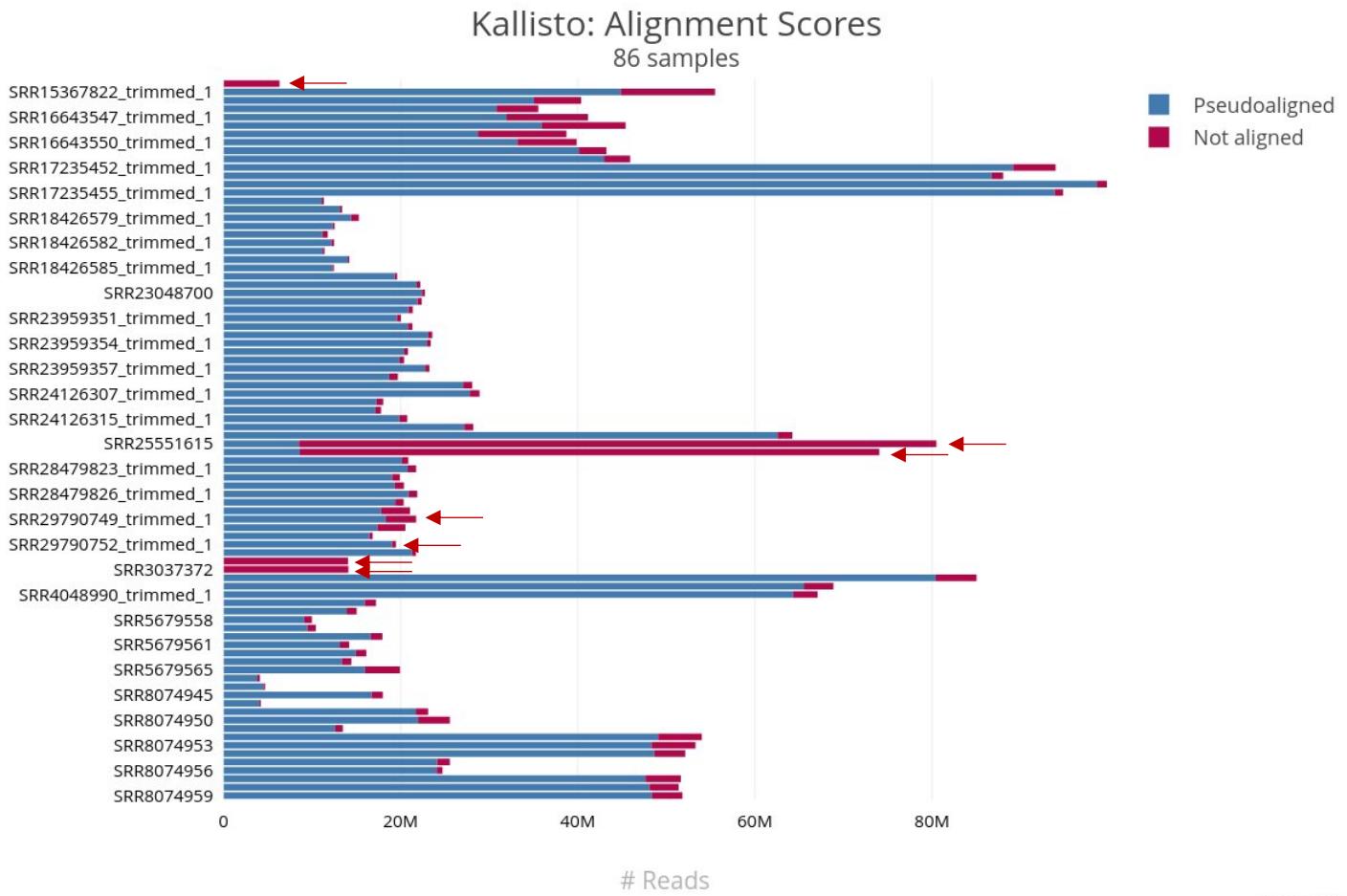


Figure 3 Kallisto alignment summary of 86 FASTQ files.

Figure 3 description: The y-axis lists the sample identifiers (SRR IDs), x-axis shows the number of reads (in millions). Blue bars indicate pseudo aligned reads, and red bars represent unaligned reads. Red arrow represents the some of the samples removed from the study.

Note: Not all 86 sample IDs are labelled here due to MultiQC download issue.

The MultiQC report revealed that several samples exhibited poor alignment quality relative to their read lengths. As a result, the following samples were excluded from further analysis: SRR12775028, SRR3037371, SRR3037372. In addition to the low alignment quality, some samples were also removed - SRR25551615, SRR25551616, SRR25521634, SRR25521635, SRR29790750, SRR29790749, SRR29790748, SRR29790753, SRR29790752, SRR29790751 due to the absence of a supporting published article. After these exclusions, Kallisto alignment

was re-evaluated on the remaining samples, and a new MultiQC report was generated to visualize the updated alignment metrics (refer Figure 4).

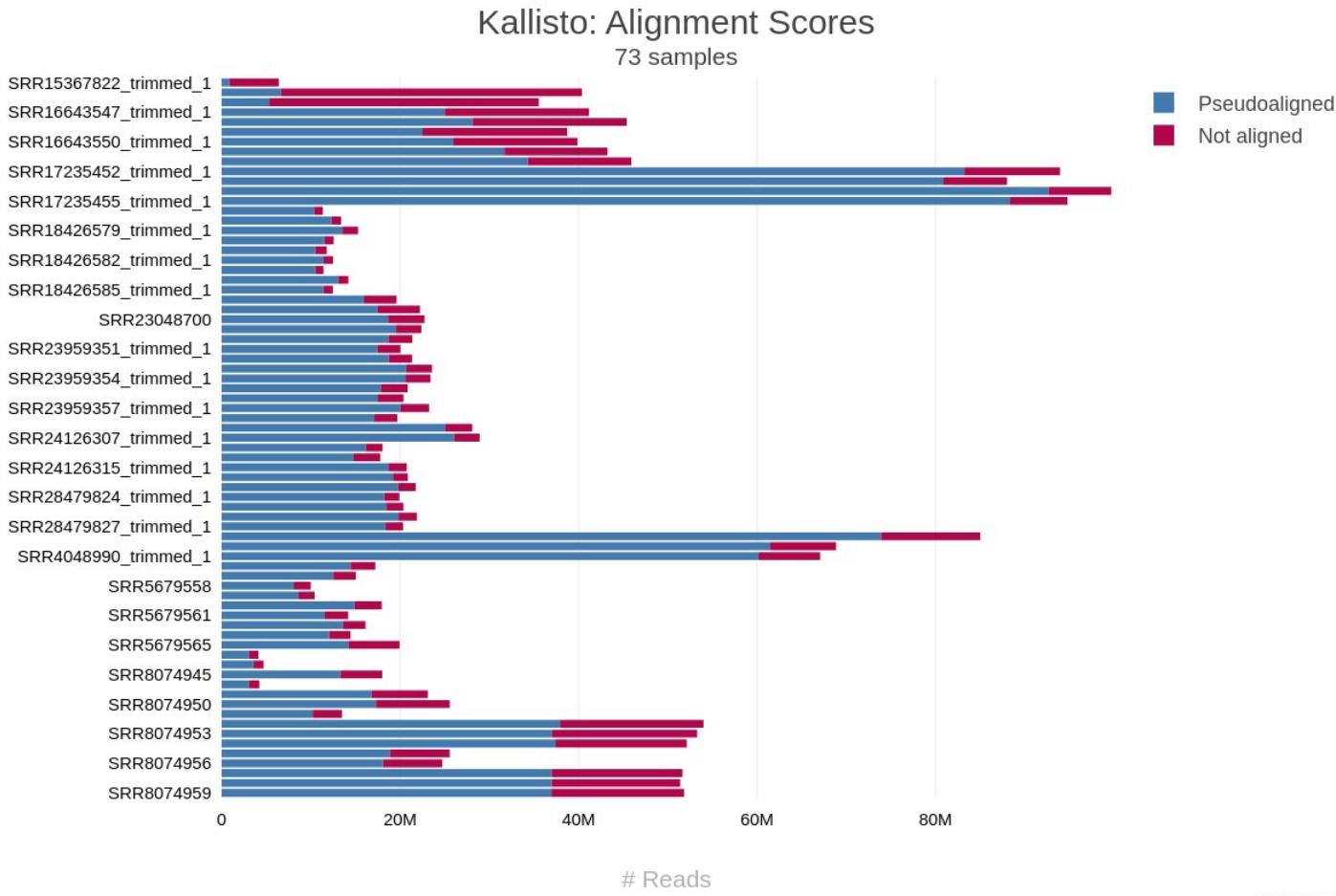


Figure 4 Kallisto alignment summary of 73 FASTQ files.

Figure 4 description: The y-axis lists the sample identifiers (SRR IDs), x-axis shows the number of reads (in millions). Blue bar indicates pseudoaligned reads, and red bars represents unaligned reads.

For the finalized 73 FASTQ files, abundance.h5 files from individual sample directory was processed using tximport (v1.36.0) R package to generate gene-level transcripts per million (TPM) and raw count matrices. To aggregate transcript-level estimates to the gene level, a transcript-to-gene (tx2gene) mapping was generated from the *Solanum lycopersicum* genome annotation file (ITAG4_gene_models.gff). The GFF file was imported using rtracklayer (v1.68.0) R package which uses `rtracklayer::import()` function. Transcripts of type “mRNA” were matched to the corresponding gene (parent attribute) to create tx2gene mapping. The extracted TPM values from the abundance was saved as `tpm_matrix.csv` and the counts were saved as `counts_matrix.csv` (files are saved to GIT).

The TPM (labelled as `tpm_matrix.csv`) was generated to represent normalized expression values to account for both transcript length and sequencing depth, enabling comparison across genes and samples. These TPM values were used for exploratory analysis via Principal Component Analysis (PCA) to visualize sample clustering, tissue-specific expression patterns, and overall variance structure. To reduce noise and improve biological interpretability in PCA, a filtering step was applied to retain genes with $\text{TPM} \geq 0.50$, while also retaining the unfiltered version for

comparison. The extracted raw counts matrix was used for downstream analysis. Counts were normalized through median-of-ratios method followed by variance stabilizing transformation (VST).

d. Normalization

i. DESeq2 & VST

Raw RNA-seq counts represent the number of sequencing reads mapped to each gene, but these raw numbers are influenced by multiple technical factors unrelated to true biological differences. These technical biases in RNA seq data can arise from library prep conditions such as weather, technician, sample quality (has it been kept in a cool logistic chain), sequencing depth, library preparation, gene length (Risso et al., 2011). Normalization corrects these biases (when we have enough samples to still get a signal from the remaining samples) and so the observed difference in count, hopefully, reflects the true biological variation.

DESeq2 (v1.48.1) normalization uses the median-of-ratios method to normalize the raw RNA-seq counts, where each sample's size factor is the median of ratios of gene counts to the geometric mean across samples. This ensures that the sequencing depth and composition bias are corrected. After DESeq2 normalization, variance stabilizing transformation (VST) is applied to make the variance roughly constant across genes(Love et al., 2014).

The resulting VST matrix was used for:

- Principal component analysis (PCA) to explore sample clustering by treatment or interaction type.
- Weighted gene co-expression network analysis (WGCNA) for module detection.
- And GENIE3 for regulatory network inference.

e. Co-expression Network analysis

i. WGCNA

WGCNA is used to model a scale-free network based on gene expression correlations. It is used to find clusters called modules of highly correlated genes, for summarizing such clusters using the module eigengene, for relating modules to one another and to external traits (using the eigengene network), and for calculating module membership measures(Ghafouri-Fard et al., 2023b)(Langfelder & Horvath, 2008).

The file vst_data.csv obtained after normalization was the input data for WGCNA (v1.73). The WGCNA data was then subjected to **goodSamplesGenes()** function which is an inbuilt function of WGCNA, which checks for missing entries in the data, entries with weights below a threshold, and zero-variance genes, and returns a list of samples and genes that pass criteria on maximum number of missing or low weight values(*goodSamplesGenesMS Function - RDocumentation*, n.d.)(Langfelder & Horvath, 2012). For the current dataset, this function filtered out 2085 genes

out of the total 34075 genes. Hence for further analysis of WGCNA only 31990 genes were considered.

ii. Signed hybrid & Soft threshold

The function **pickSoftThreshold()** evaluates and selects appropriate β (refer Equation 1) and computes a scale-free topology fit index (R^2) through log-log plots of node degree vs frequency. It selects the lowest β where R^2 exceeds a threshold of 0.85 (refer Figure 11), (*pickSoftThreshold*, n.d.) (Langfelder & Horvath, 2008) and the network type was specified as signed hybrid. Signed hybrid selects and groups those genes into modules that are positively correlated with each other; all negative correlations are excluded from this type of network (Langfelder, 2018). For the current dataset, a soft threshold of 6 was selected using this function. Then the adjacency matrix was computed with the soft-threshold determined and the type specified as signed-hybrid.

iii. Module detection using hierarchical clustering

To represent the intensity of the connection between the genes, the adjacency matrix is transformed into a topological overlap measure (TOM). Genes were analyzed via hierarchical clustering using TOM as input and network modules were detected using **DynamicTreeCut()** function (Ghafouri-Fard et al., 2023c) where identified modules are assigned different colors. The module detection method clusters a group of genes whose expression profiles are highly correlated. To summarize such modules, the function **moduleEigengene()** is used. This represents the module expression of the q-th module by the module eigengene $E^{(q)}$, which is defined as the first principal component of the expression matrix. Once the module eigengene (ME) is computed, eigengene dissimilarity is calculated as 1 minus the correlation between MEs which is used to build dendograms of modules. Once the modules are defined and visualized, an appropriate cut off is determined (Langfelder & Horvath, 2008). For this study, the modules were merged using **mergeCloseModules()** function in R specifying a **cutHeight** of 0.40. The merged modules and merged colors are used as input to plot the Module-Trait Relation (MTR) between the different interactions and microbes for further analysis.

f. Regulatory network analysis

i. GENIE3

GENIE3 (v1.28.0) is a tree based method to reconstruct GRNs (Huynh-Thu et al., 2010). The assumption of GENIE3 is that the expression of each gene could be described as a function of the expression of some TFs, which means the selected TFs could regulate the target gene (Hu et al., 2020). Since GENIE3 uses tree-based approaches to predict the expression profiles, **randomForest** function in R is implemented. For this analysis, TF list was obtained from sol genomics database which follows ITAG version 4 (same version as the reference FASTA file). Both `vst_data.csv` and TF list are loaded as input for GENIE3. GENIE3 computing results in a link list in the form of csv file which is used as the input for Louvain clustering.

ii. Louvain clustering & Resolution

Louvain clustering is a graph-based community detection algorithm that identifies modules or clusters by optimizing modularity. Louvain clustering is an inbuilt package taken from igraph (v1.5.1). Since **igraph:cluster_louvain()** requires undirected input, the directed regulatory link list is initially converted from directed to undirected. To compare the WGCNA and GENIE3 analysis, having equal or slightly greater or less modules as that of WGCNA was needed for it to be comparable and to conduct robust statistics with less imbalances. To achieve this “**resolution**” is specified while the louvain cluster is computed. For this study a resolution of 2.1 was specified.

iii. Cosine similarity & Cosine threshold

To evaluate the similarity between gene modules identified via two different clustering methods: louvain clustering on the GENIE3 network and WGCNA module detection, a cosine similarity is computed between the binary module membership vectors of genes. As the first step a binary matrix is constructed where for each gene, a 1 indicates the gene is in the that module (WGCNA) and cluster (Louvain-GENIE3) whereas 0 means that gene is absent in that module and cluster. The resulting binary membership is then subjected to cosine similarity where cosine similarity of 1 indicates that the cluster and module are identical in direction, 0 indicates they are orthogonal and in-between values reflects partial overlap.

Cosine similarity is calculated as following in R:

```
cosine_similarity <- function(x, y) {
  sum(x * y) / (sqrt(sum(x^2)) * sqrt(sum(y^2)))
}
```

Equation 2 Cosine similarity

where row is WGCNA modules and column is Louvain clusters. The overlap between some Louvain clusters and WGCNA modules are not strong hence a cosine threshold of 0.03 is set to filter and gather the stronger overlap. Lower than 0.03 cosine similarity between WGCNA and Louvain is omitted for this study.

g. Functional enrichment

i. Mercator and clusterProfiler

To identify overrepresented functional categories among genes, a functional enrichment analysis was performed using annotations from the Mercator tool (v.7) and the cluster profiler (v3.8.1) R package. The analysis focused on genes belonging to WGCNA modules and GENIE3 clusters.

Mercator4(Bolger et al., 2021)tool is used for protein annotation. In the online platform, the reference FASTA file is provided as an input to generate the mapman mapping file which is named as **solanum_proteins.results.txt**. WGCNA module of interest is loaded and along with the mapping file, **enricher()** function from the clusterProfiler package is performed. This function

performs a hypergeometric test to assess whether Mercator functional terms are overrepresented among the module genes compared to the background of all annotated genes. To visualize the enrichment results. Both dot plots and bar plots are generated using **enrichplot()** function.

To explore the biological relevance of gene regulatory clusters identified through GENIE3 clustering is required and that was achieved using Louvain clustering. The annotation file (*solanum_proteins.results.txt*) was processed to create a mapping of genes to functional terms (TERM2GENE). Each Louvain cluster was matched with its corresponding WGCNA module, and the intersection of genes between the two sets was used as the input for enrichment analysis. The background set for each test included all the genes from the corresponding WGCNA module. Functional enrichment was performed using **enricher()** function and plots (dot plot and bar plot) were created using **enrichplot()**.

5. Results

a. Principal Component Analysis (PCA)

To assess the structure and quality of the gene expression data prior to network construction principal component analysis (PCA) was performed. PCA reduces high dimensional transcriptomic data into a small number of components that capture the major sources of variation. This allows visualization of sample clustering patterns and help identify technical noise, potential outliers and biological grouping such as tissue type or microbial interaction.

In this study PCA was applied to both TPM (refer Section 4.c.i) and VST normalized data (refer Section 4.d.i) to explore how different normalization strategies influence the separation of sample groups. Clear clustering by interaction type or tissue, and increased variance explained by VST PCA, provides confidence in the data quality, and supports the use of downstream co-expression and regulatory network analysis.

i. PCA plotted using TPM values

Using the TPM values extracted, a PCA plot is generated for both un-filtered (refer Figure 5) and filtered data (refer Figure 6), by setting a threshold of 0.50 TPM for the filtered dataset (refer Section 4.c.i). The variations within the entire dataset are examined both before and after filtering to remove noise, in order to determine if the differences observed in gene expression primarily reflect true biological variation. By setting the threshold of 0.50 TPM, the focus is on the genes that have enough expression to be meaningful and relevant thereby reducing the noise and improving the reliability for downstream analyses. To highlight the impact of filtering and to assess the overall data quality, PCA plots were generated for both filtered and unfiltered TPM values, providing a visual check before proceeding with count-based analysis.

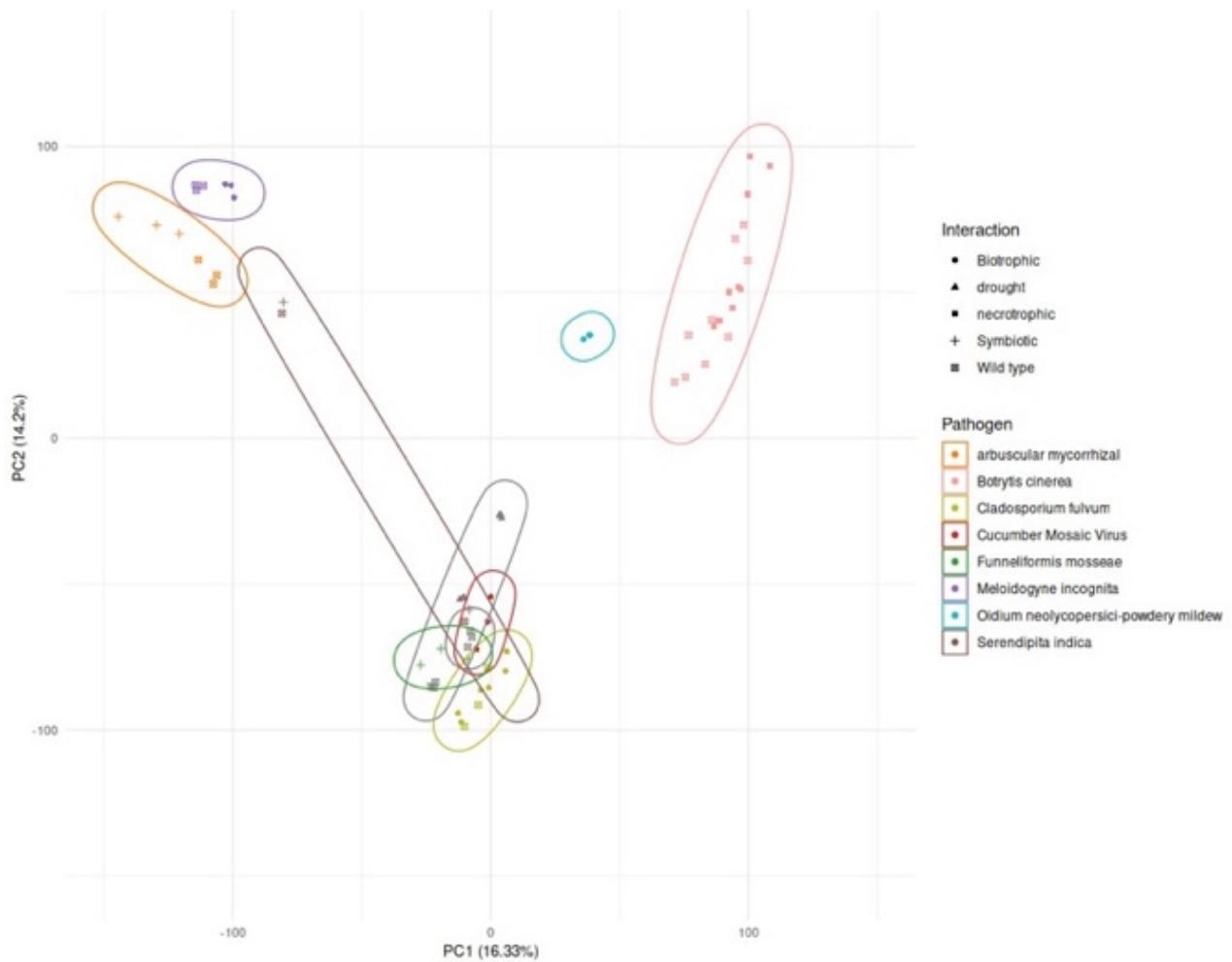


Figure 5 PCA of TPM unfiltered.

Figure 5 description: The x axis represents the principal component (PC)1 where the variance represented is 16.33%, the y axis represents PC 2 where the variance is 14.2%. The overall variance represented here is 30.33%.

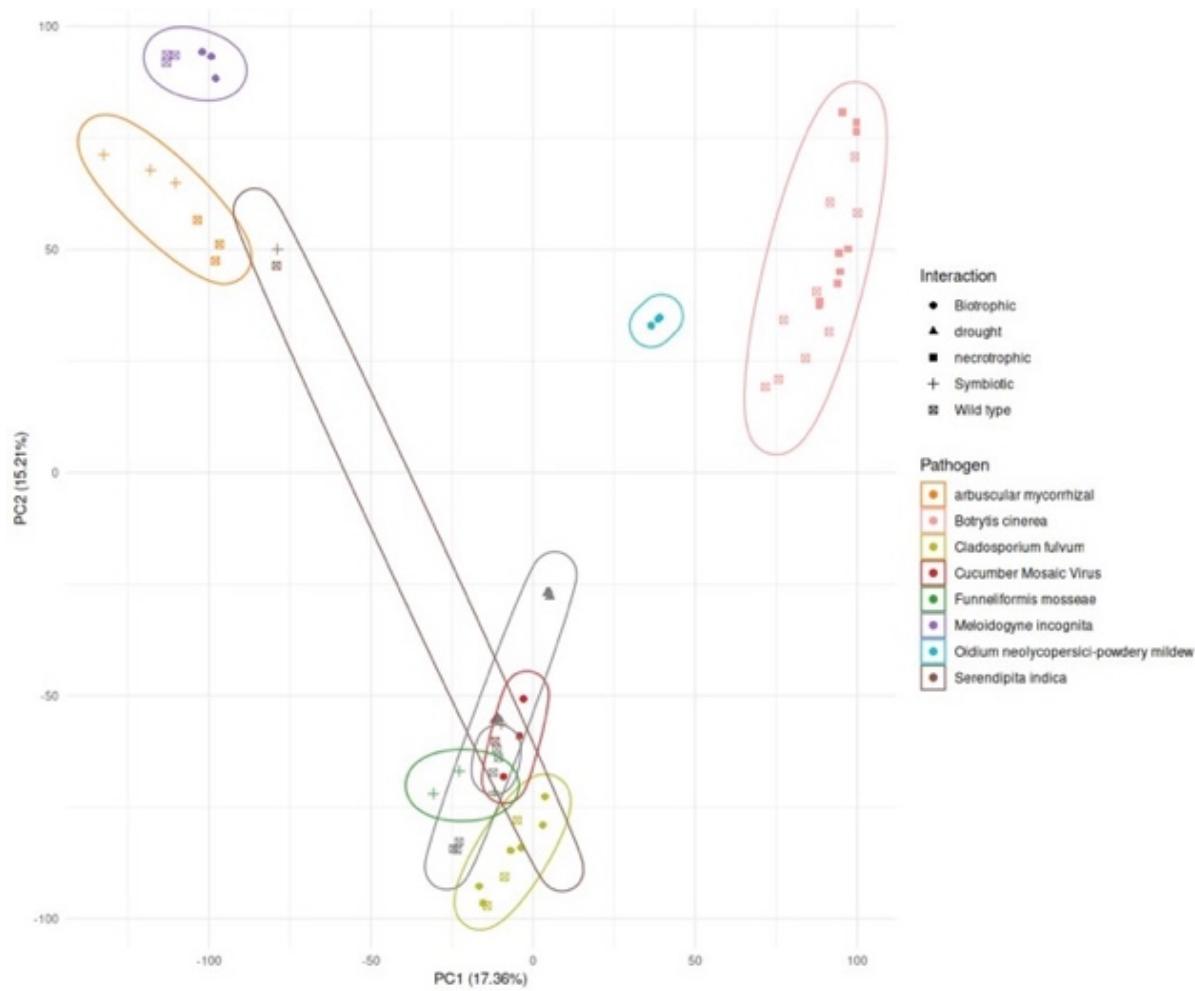


Figure 6 PCA of TPM filtered.

Figure 6 description: The x axis represents the PC1 where the variance represented is 17.36%, the y axis represents PC 2 where the variance is 15.21%. Together, these two components capture variance of 32.57%

Since the cumulative variance explained by the first two principal components doesn't exceed 50% in both PCA plots, higher principal components were also examined to ensure that additional biological variation is not overlooked. Indeed, while principal component one captures the largest portion of variance and reflects the main biological effect, it doesn't fully account for all variation observed between samples. Hence PC3 and PC4 were examined for both unfiltered and filtered TPM values, which also didn't exceed 30% as compared to PC 1 and PC2 (refer GitLab > Pre-analysis_plot > PCA3and4_filteredTPM.jpg & PCA3and4_unfilteredTPM.jpg).

ii. PCA with tissue labels

This PCA plot visualizes the distribution of samples based on tissue type which are leaf and root (refer Figure 7 and 8). Each point in the PCA represents an individual sample, distinguished by tissue type- circles for leaf and triangle for root. Clear clustering patterns suggest that tissue type strongly influences global gene expression profiles.

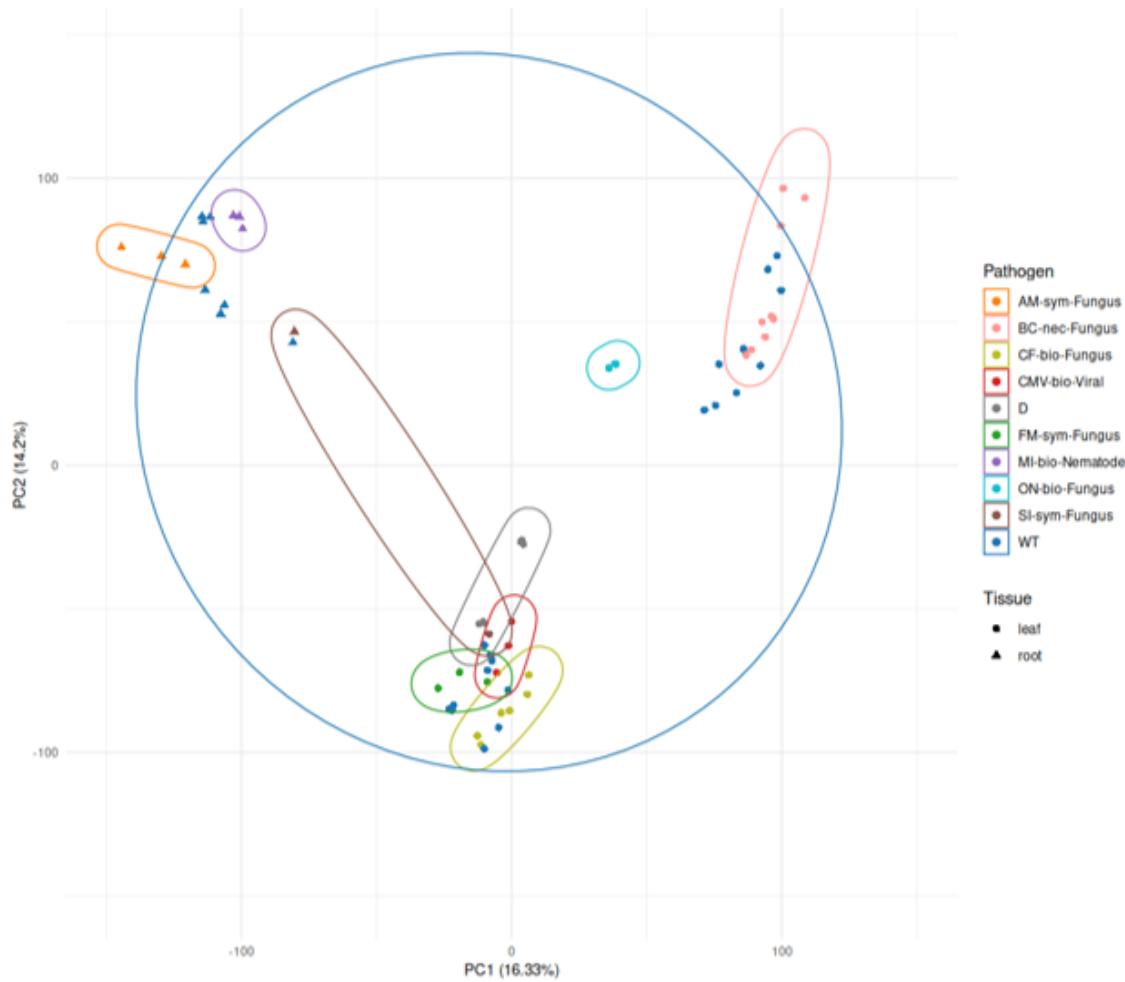


Figure 7 PCA of Tissue sample (PC1 and 2).

Figure 7 description: The x axis represents the PC1 where the variance represented is 16.33%, the y axis represents PC 2 where the variance is 14.20%. Together, these two components capture variance of 30.53%.

Together, PC 1 and 2 account for 30.53% of the total variance, indicating that the main biological differences are not fully captured by just the first two components. As the cumulative variance doesn't exceed 50 % in Figure 7 PCA plot, higher principal components (PC 3 & 4) are also explored to capture potential biological variation that may not be represented by PC 1 and 2 (refer Figure 8). Interestingly, PC 3 and 4 also explain a comparable proportion of variance, suggesting that additional biological signals are distributed across multiple dimensions.

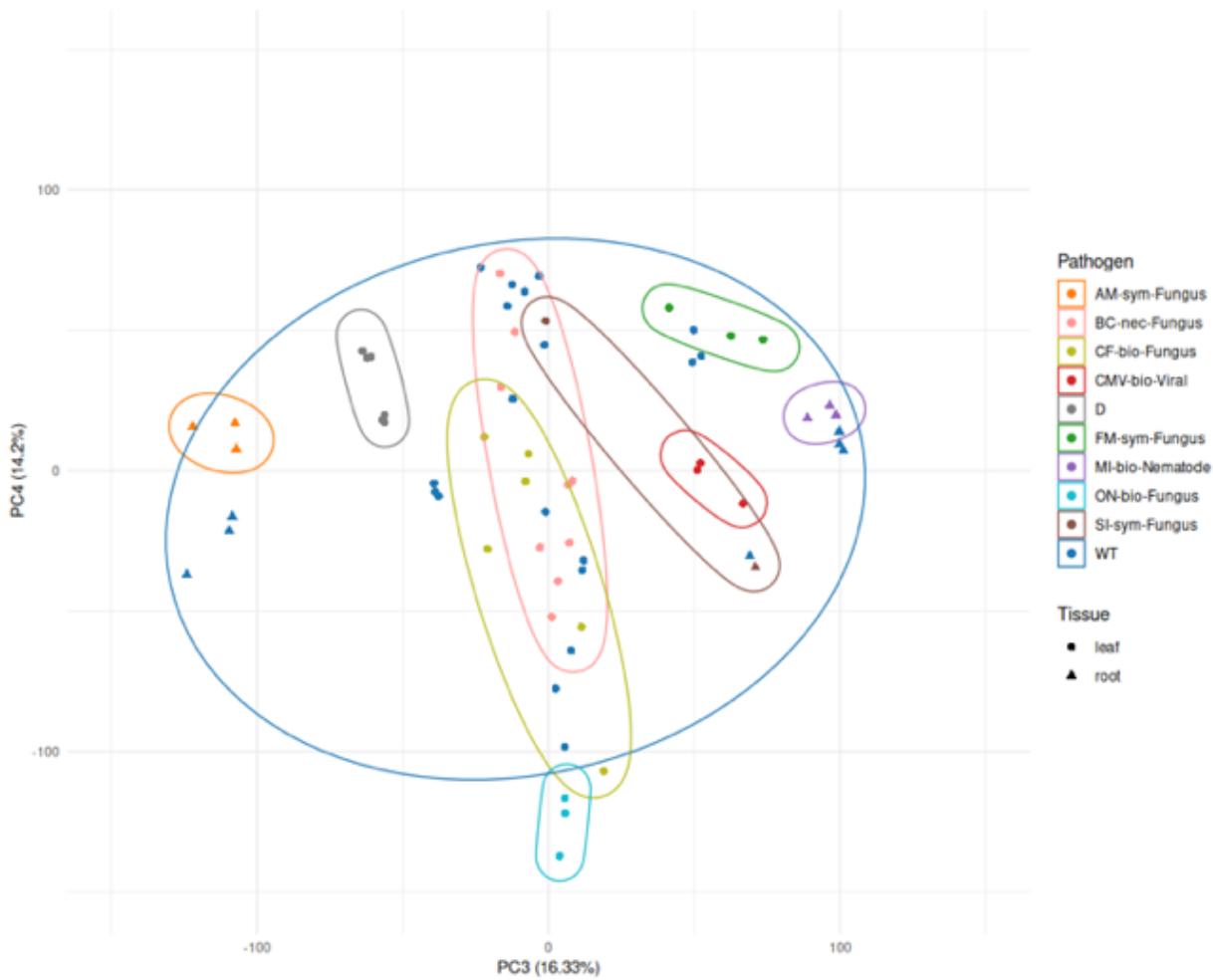


Figure 8 PCA of Tissue sample (PC 3 and 4).

Figure 8 description: The x axis represents the PC3 where the variance represented is 16.33%, the y axis represents PC 4 where the variance is 14.20%. Together, these two components capture variance of 30.53%.

iii. PCA plotted using VST data

This PCA plot is generated using raw count data processed through the DESeq2 pipeline (refer Figure 9). Sample clusters according to the different type (biotrophic, necrotrophic, symbiotic, drought and wild type) and are further grouped by different pathogens. The increased variance explained by the 1st 2 components suggest that VST effectively compresses noise and highlights biologically meaningful differences.

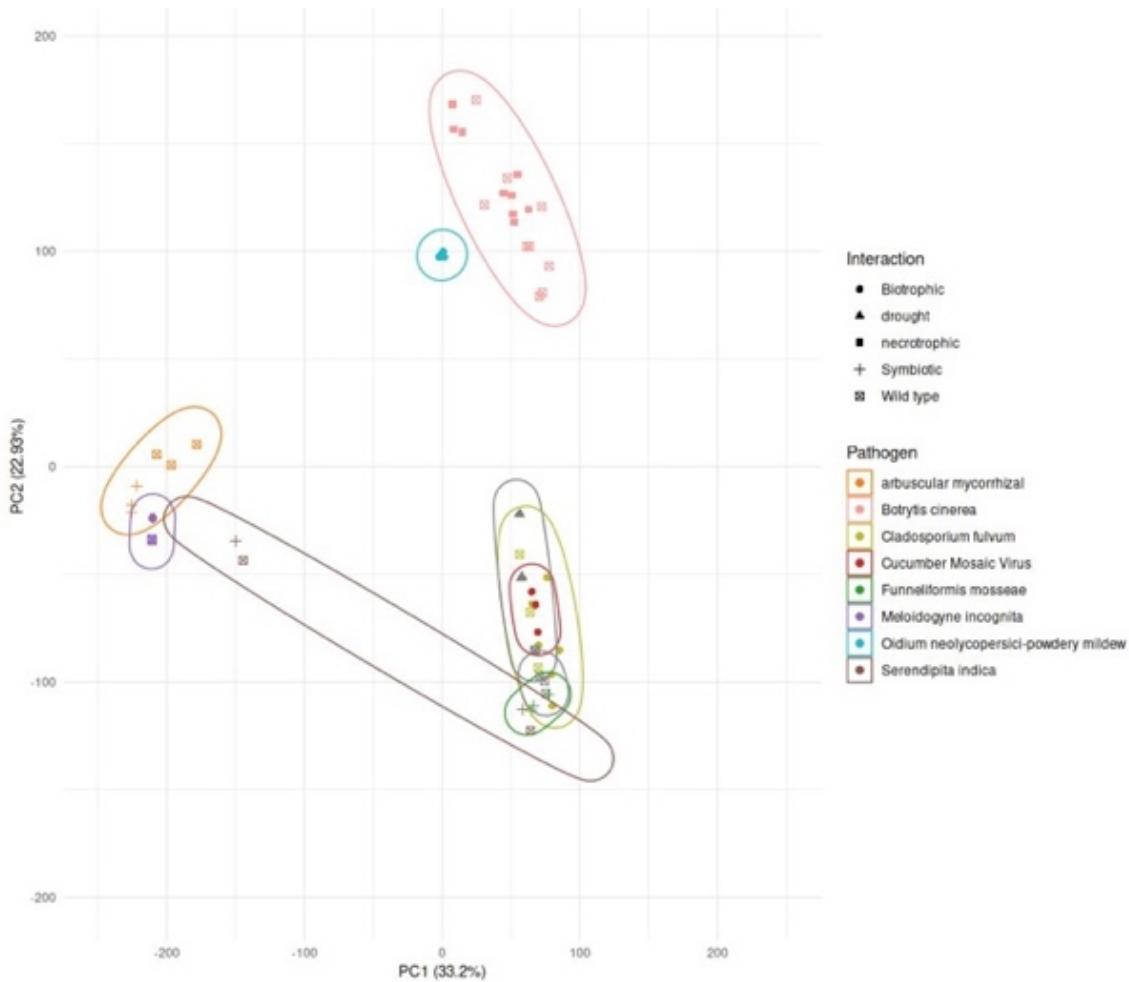


Figure 9 PCA of VST data.

Figure 9 description: The x axis represents the PC1 where the variance represented is 33.20%, the y axis represents PC 2 where the variance is 22.93%. Together, these two components capture variance of 56.13%.

b. Co-expression modules

i. Overview of all the plots generated through WGCNA

WGCNA was performed to construct correlation network and identify modules of co-expressed genes associated with different biological conditions. The analysis began with hierarchical clustering of samples (refer Figure 10), confirming the data quality and excluding potential outliers. A soft thresholding of power 6 was selected (refer Figure 11 and 12) based on the scale-free topology criterion to ensure robust network construction. Modules were identified using dynamic tree cutting, and similar modules were merged based on eigengene correlations (refer Figure 13). In total, several distinct modules each with unique colors were detected. The correlation of module eigengenes with experimental traits such as interaction type (Figure 14& 15) and pathogen treatment (Figure 26 & 27) revealed specific modules significantly associated with these biological factors. These outputs collectively enabled the identification of biologically meaningful modules and their functional relevance in context of plant-microbe interaction.

ii. Sample clustering

This figure (refer Figure 10) presents a hierarchical clustering dendrogram of all the samples, constructed using Euclidean distance on normalized gene expression values as the measure of dissimilarity. The main objective of this analysis was to perform an initial quality check by identifying any potential outlier samples that might compromise the integrity of the downstream analysis.

In the dendrogram, samples that are more similar in their overall gene expression profiles are clustered together and joined by branches lower on the vertical axis. Conversely, samples that are more dissimilar merge higher up. A key point to look for in this plot is whether any sample branches off at a significantly higher height than rest of the samples.

In this case, the clustering reveals well-defined groups of samples with relatively uniform heights at which branches merge. There are no obvious outlier or aberrant samples that deviate significantly from the rest of the dataset. This outcome supports the overall quality and homogeneity of the dataset and confirms that it is suitable for subsequent steps in WGCNA pipeline. By ensuring that outliers are absent at this stage, we reduce the risk of the distortions in the co-expression network and improve the reliability of the downstream biological interpretations.

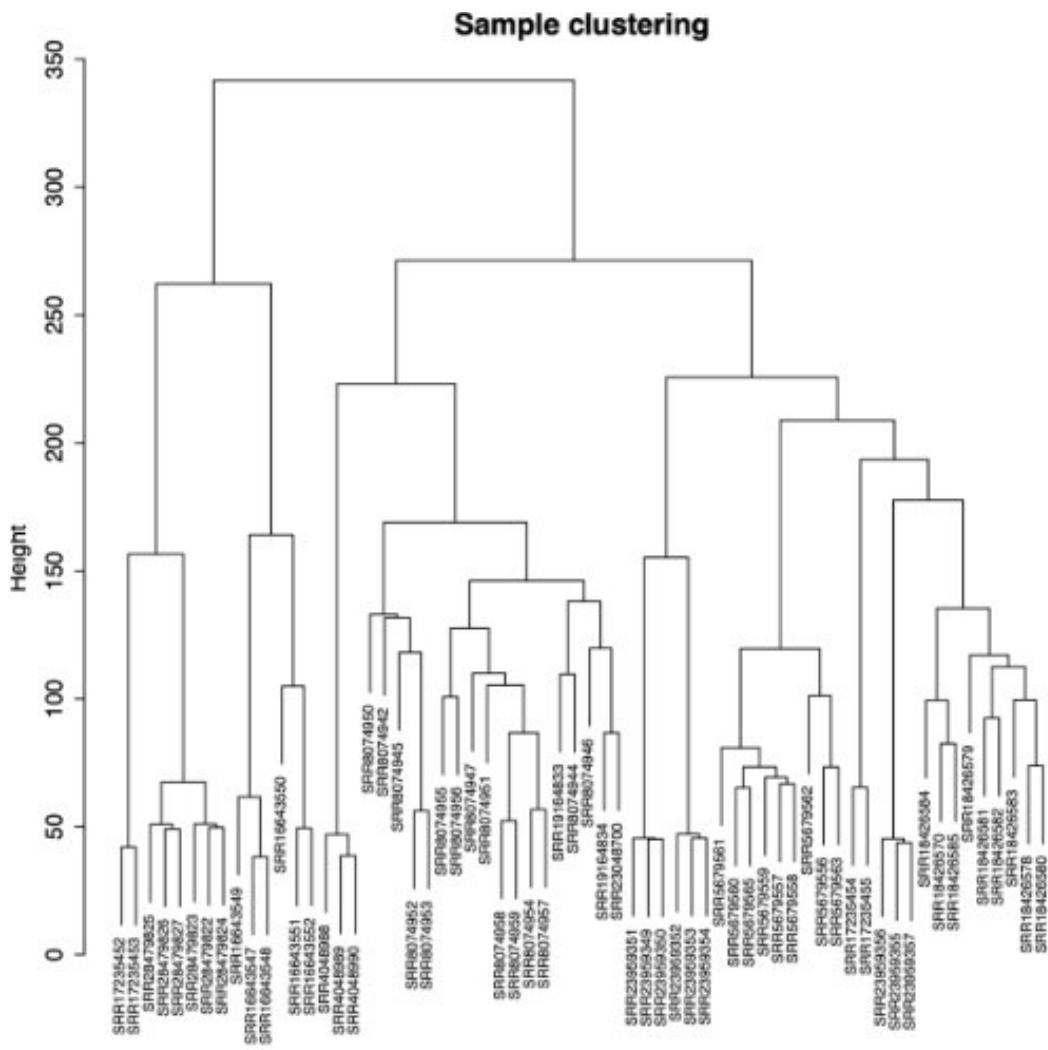


Figure 10 Sample clustering dendrogram.

Figure 10 description: The Y axis represents the height and each branch represented on the plot is a sample identified using SRR id.

iii. Soft-thresholding power selection

To determine the most suitable soft-thresholding power to construct a biologically meaningful network, we referred to the scale free topology criterion illustrated in Figure 11. This plot shows how well the network fits a scale free model at different power values. As the power increases, the model fit (signed R^2) also increases, meaning power 6 or higher yields a network that meets the scale free topology criterion.

To complement this, Figure 12 shows mean connectivity. Mean connectivity refers to the average number of connections each gene has and this changes with increasing power. Mean connectivity drops as the power increases, reflecting a sparser network. Beyond a certain point, a saturation effect is evident where further increase in power doesn't significantly change the connectivity.

By integrating the information from both plots, we can conclude that a soft-thresholding power of 6 strikes a good balance between scale-free topology while maintaining sufficient connectivity to retain biologically relevant interactions, without introducing excessive noise.

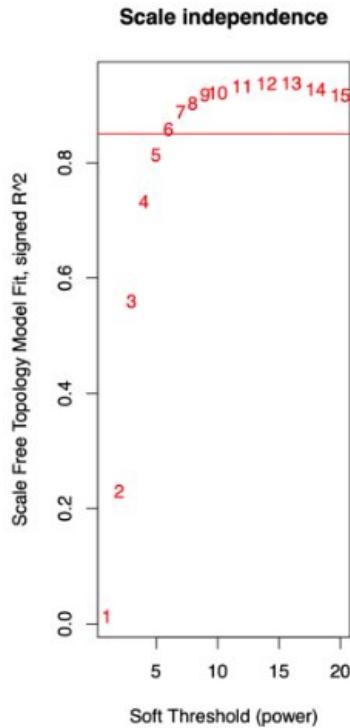


Figure 11 Scale independence plot.

Figure 11 description: Y axis represents scale free topology model fit, signed R^2 and the X axis represents the different candidate values of power parameter represented as soft threshold (power).

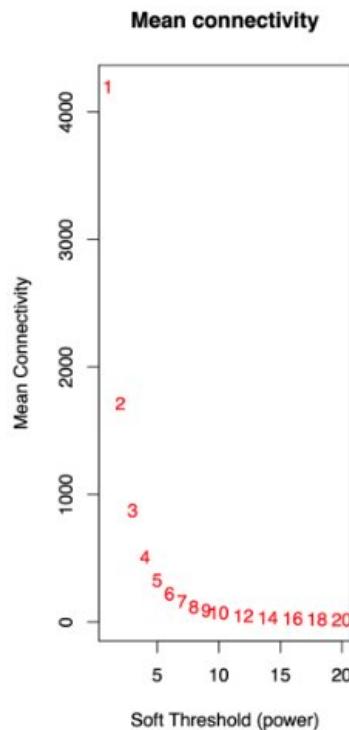


Figure 12 Mean connectivity.

Figure 12 description: Y axis represents mean connectivity and X axis represents soft threshold (power).

iv. Merged modules

This is a hierarchical clustering tree (dendrogram) of all the genes included in the analysis, constructed using topological overlap-based dissimilarity. Genes with similar co-expression profiles cluster together and appear on the same branches. Each vertical line in the dendrogram (refer Figure 13) represented in the Y axis corresponds to an individual gene, and branches represent clusters of genes that share similar expression pattern.

Bottom panel represents two rows of module color bars, the first bar is labeled “original modules”, where each color represents a preliminary module, which is a group of genes with similar expression patterns. These modules were defined using the dynamic tree cutting. The second bar represented as “merged modules”, displays the final set of modules after a merging step. In this step, the eigengenes (the first principal component summarizing the expression profiles of each module) are computed, and modules with highly correlated eigengenes are merged. By carrying out this process redundancy is reduced and enhances biological interpretability, as highly similar modules likely represent the same underlying biological signal. Hence, several colors from the original modules row may have been unified into one. Based on the eigengene correlation, a cut height of 0.40 was selected. Following the merging process, the original 77 modules were consolidated, resulting in a new total of 39 merged modules.

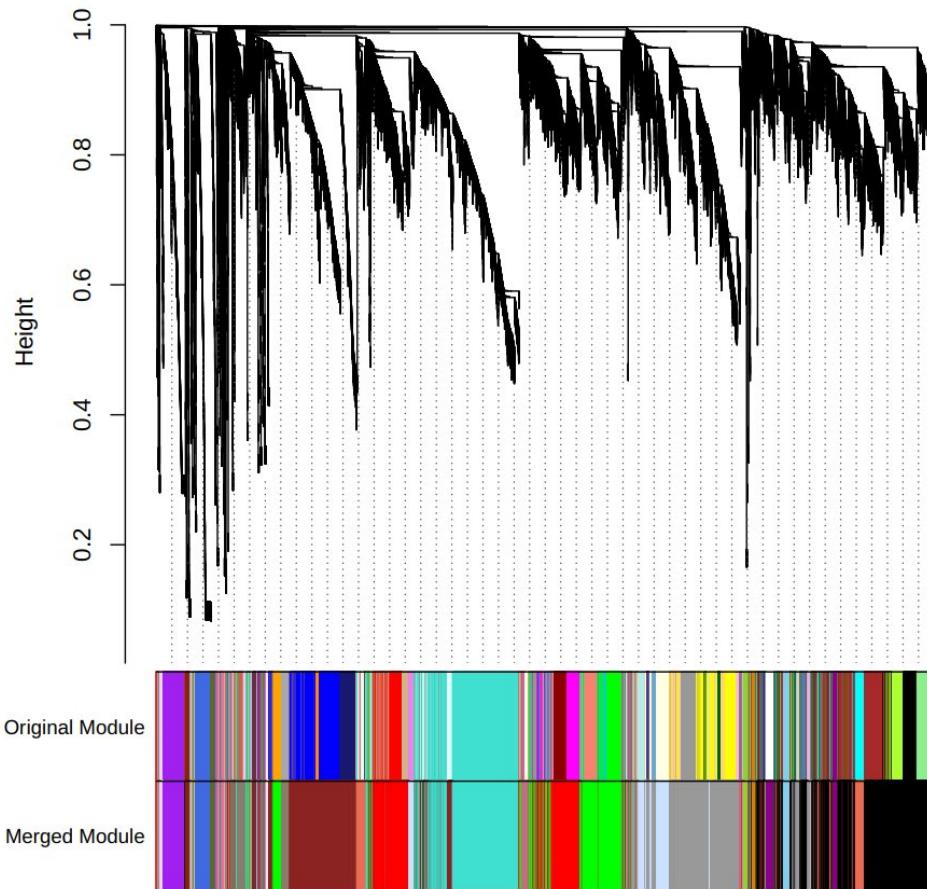


Figure 13 Gene dendrogram and module colors.

Figure 13 description: The Y axis labeled “height” represents the dissimilarity between genes, typically derived from the TOM. The X axis shows the modules (original and merged modules).

v. Module-trait associations and eigengene expression across interaction types

Following the identification and merging of modules, the next step was to explore the biological relevance of these gene clusters by examining their associations with experimental traits. To achieve this, a module-trait relationship (refer Figure 14) analysis conducted, where the correlation between each module eigengene and specific interaction types (biotrophic, symbiotic, necrotrophic, drought and wild-type) were calculated. The resulting heatmap visualizes these correlations and their associated p-value, allowing for the identification of modules most relevant to each condition. Each row corresponds to a module eigengene, which summarizes the expression pattern of all genes within that module. Each cell shows the Pearson correlation coefficient between the modules eigengene and the trait, and its corresponding p-value (in each cell represented in parentheses). The color gradient represents the strength and direction of the correlation, where red is for positive correlation and blue is for negative correlation, with deeper shades indicating stronger correlation.

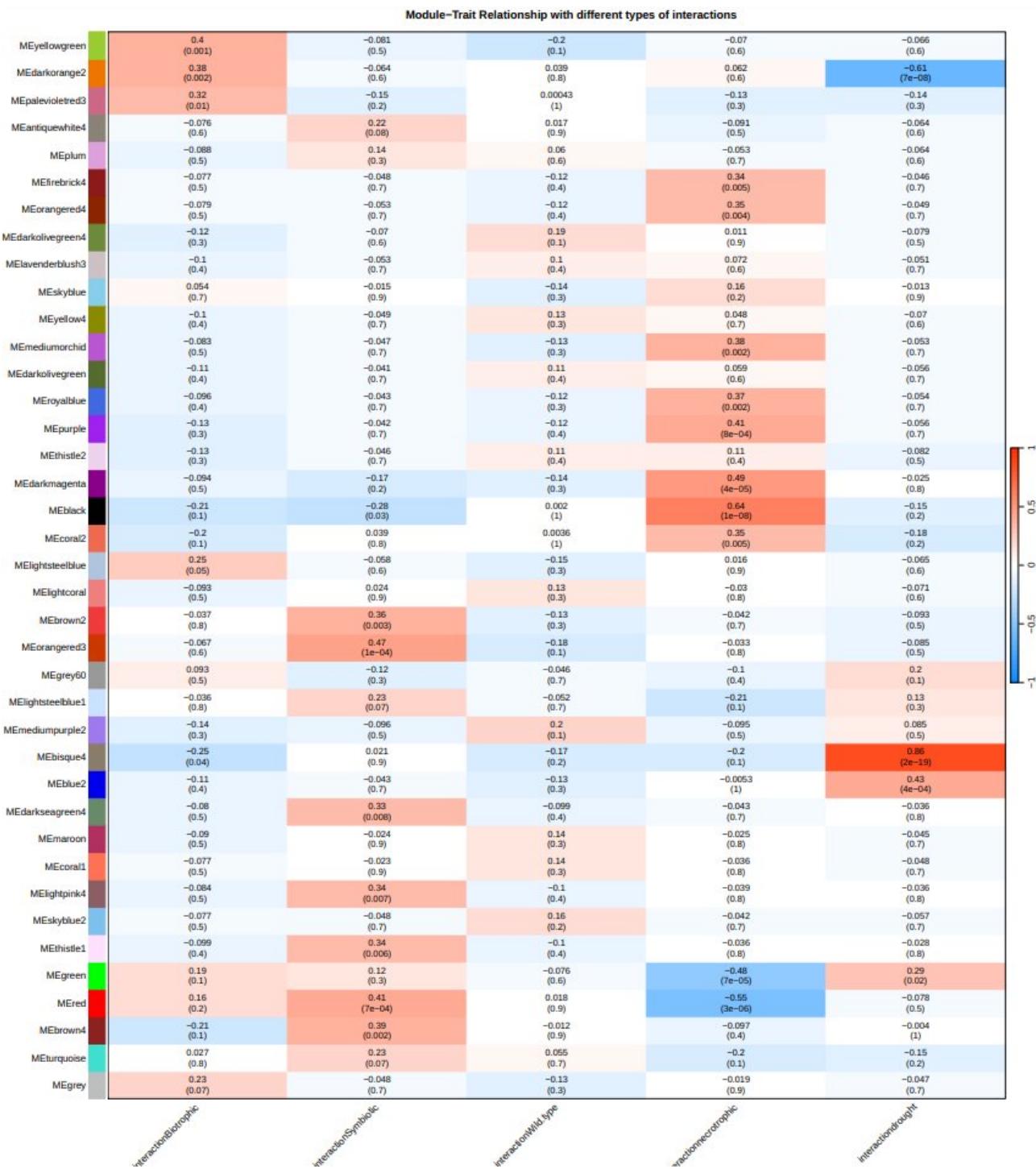


Figure 14 Module-Trait relationship with different types of interactions.

Figure 14 description: The x axis represents different interactions such as biotrophic, symbiotic, wild type, drought, and necrotrophic. The y axis represents 39 modules obtained after module merging.

Among the 39 merged modules, we observe that there are several modules that show correlations both positive and negative for specific interaction type. MEred and MEGreen modules show shared positive correlation for biotrophic and symbiotic interaction which having negative correlation (MEred r = -0.55 and MEGreen r = -0.48 and p = 3e-06 and 7e-05 respectively) with necrotrophic interaction where the positive correlation values range between 0.12 -0.4. This suggests that the genes in these modules are involved in general host responses to mutualistic and pathogenic stimuli, while being suppressed during necrotrophic stress. MEyellowgreen showed

a strong positive correlation with biotrophic interaction ($r = 0.4$, $p = 0.001$), while being negatively correlated with the remaining interaction types, suggesting that this module maybe specifically upregulated in biotrophic context.

The ME brown4 modules was positively correlated with the symbiotic interaction ($r = 0.39$, $p = 0.002$), indicating its potential involvement in the gene network associated with beneficial plant-microbe associations. Likewise, MEorangered3 showed notable positive correlation with the symbiotic condition ($r = 0.47$, $p = 1e-04$), highlighting this module as another candidate involved in mutualistic interaction responses. METurquoise also displayed a positive, though weaker, correlation with the symbiotic interaction ($r = 0.23$, $p = 0.07$), suggesting possible but less definitive relevance. With respect to the biotrophic trait, several modules showed notable correlations. MElightsteelblue was positively correlated with biotrophic interaction ($r = 0.25$, $p = 0.05$), MEpalevioletred3 also demonstrated positive correlation with biotrophic interaction ($r = 0.32$, $p = 0.01$).

To further explore the global co-expression structure among modules and their interaction types, we examined the eigengene adjacency heatmap with merged module dendrogram (refer Figure 15). This visualization provides insights into how module eigengenes relate to one another across the dataset. For instance, MERed and MEgreen cluster closer together, potentially indicating a shared or coordinated regulation of functional pathways. MEorangered3 is clustered closely with the symbiotic interaction meaning, this module might have most genes that contribute to a functional pathway leading to a symbiotic relationship. MEyellowgreen module is clustered together with the biotrophic interaction also indicating that there are potential genes from the yellowgreen module that may contribute for specific biotrophic interaction pathways.

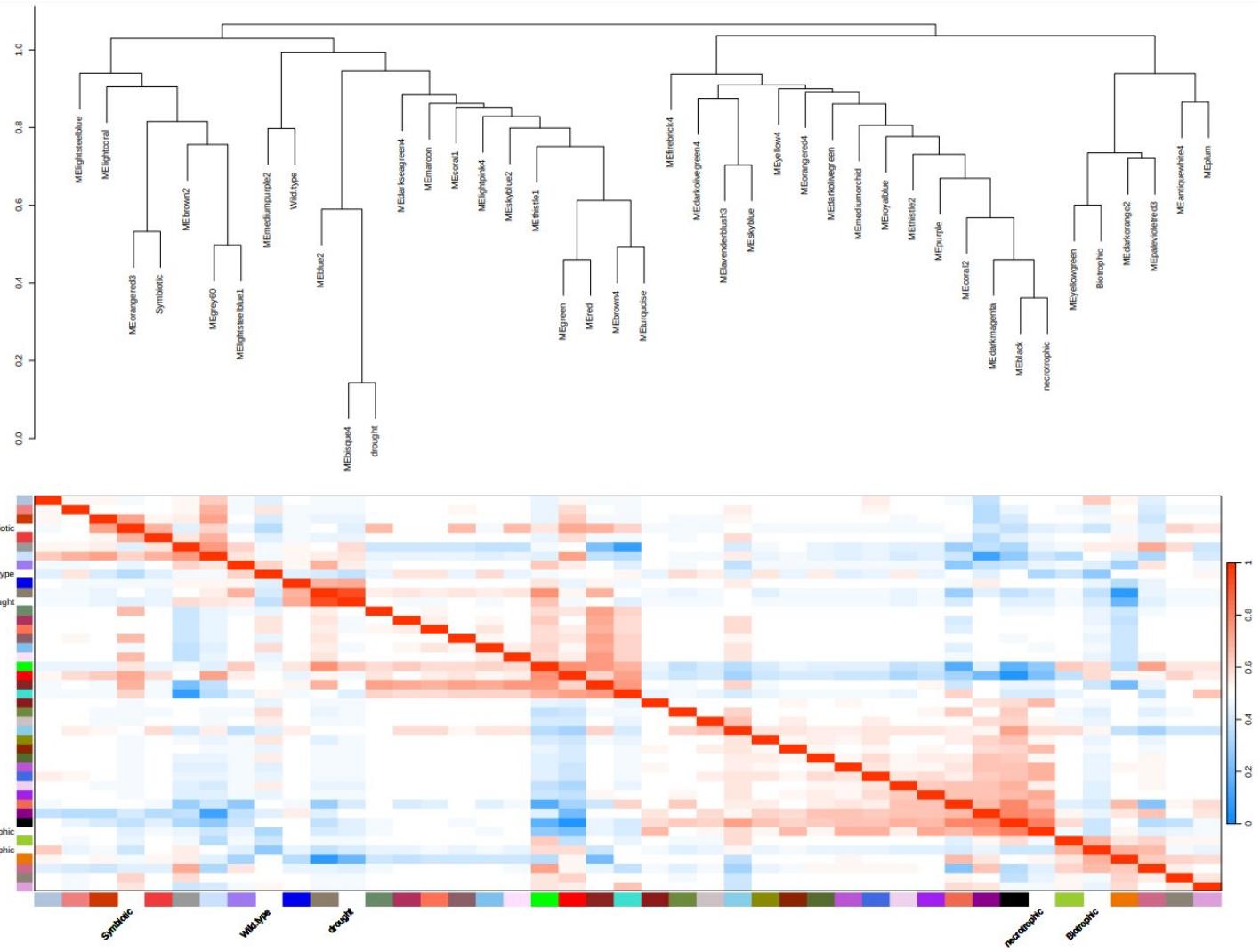


Figure 15 Dendrogram and heatmap of module eigengene and interaction type.

Figure 15 description: The top panel shows hierarchical clustering of modules eigengenes and interaction types based on expression similarity. The bottom heatmap visualizes pairwise correlations.

The control samples representing stress-related conditions displayed distinct patterns within the eigengene heatmap. MEblack module and many other modules under the same main branch clustered closely with the necrotrophic interaction, while MEbisque4 clustered with the drought condition and showed high correlation as the branch with MEbisque4 was closer to the heatmap. This clustering supports their association with stress-induced gene expression, which is likely enriched in pathways involving defense responses, hormone signaling and programmed cell death. Given that the aim of this study is to elucidate pathways specific to mutualistic plant-microbe interactions, particularly biotrophic and symbiotic associations, these stress-responsive modules can be deprioritized for further analysis. By omitting modules such as MEblack and MEbisque4, which is more likely to capture canonical stress or apoptosis related signatures, we refine our focus towards modules that may uncover more specific transcriptional programs underlying beneficial or compatible interaction. This filtering strategy enhances the specificity of our investigation, allowing for clearer insights into the molecular mechanisms governing plant-microbe communication.

Although these results provide useful information about potential module-trait relationships, they are merely the first step towards biological relevance. Examining module eigengene expression patterns in connection to individual microbial treatment is crucial to validating these correlations. Such an analysis would provide finer resolution, enabling us to determine whether the observed associations are consistent across different organisms within each interaction type.

vi. Gene significance plot

To identify biologically significant modules and genes are a major goal of co-expression analyses. Hence, to incorporate information regarding the relevance of individual genes to specific interaction types into the co-expression network, we make use of gene significance (GS) and module membership (MM) measures (refer Figures 16 – 25 (A-E)). Gene significance is defined as a function GS that assigns a non-negative number to each gene; the higher G_i the more biologically significant is the gene i (Langfelder & Horvath, 2008), this represented on the Y axis. The MM is represented on the X axis, and this indicates how strongly each gene is associated with the modules eigengene, which is essentially the representative expression profile of the module (Langfelder & Horvath, 2008). Each dot in the scatter plot represents a single gene.

Based on the findings from module-trait relationship and eigengene heatmap for different interaction types, both red and green modules showed slight positive correlations with biotrophic and symbiotic interactions. To further investigate this, GS vs MM plots are generated for each interaction type. These plots showed a clear difference in how all genes relate within each module for a specific interaction. In the red module (refer Figure 16 A-E), a moderate level positive correlation was observed for symbiotic ($\text{cor} = 0.48, p = 5\text{e-}194$) interaction, indicating that genes are highly connected within this module tend to be biologically relevant to symbiosis. In contrast, slight negative correlation was found for biotrophic interaction ($\text{cor} = -0.0055, p = 0.75$), indicating limited evidence of the number of genes correlating with the red module eigengene vector.

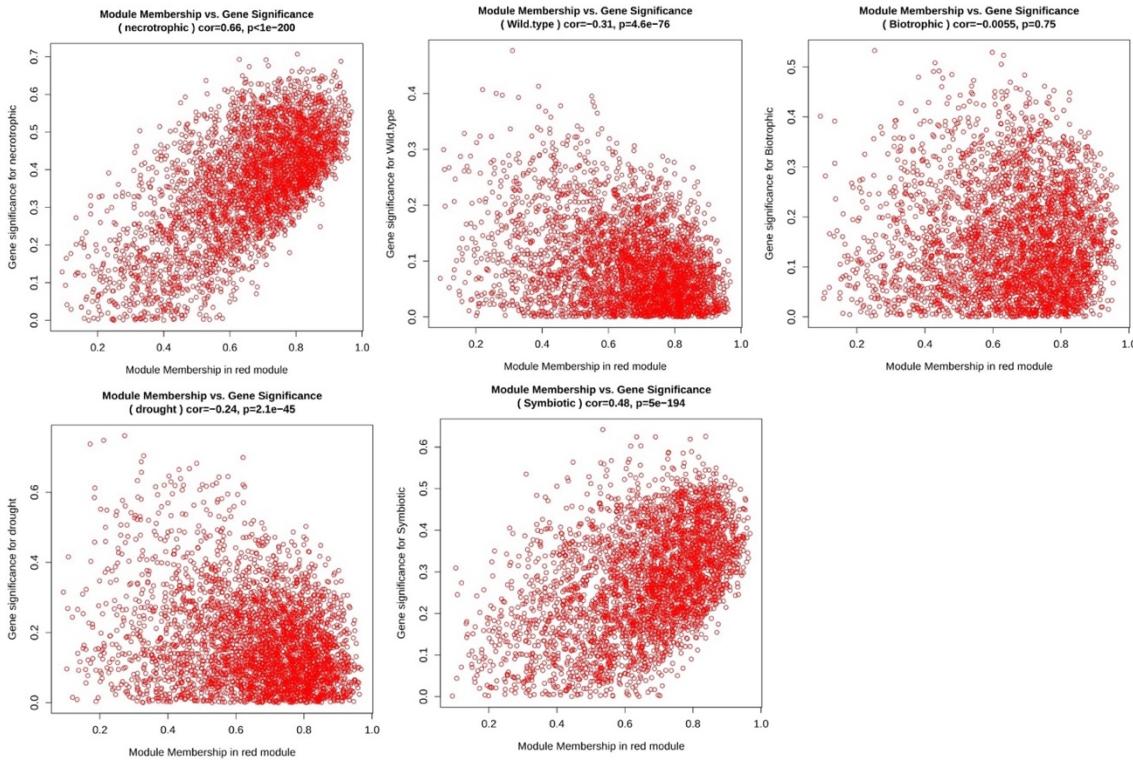


Figure 16 Scatter plot of MM vs GS in MEred module.

Figure 16 (A-E) description: MM presents the correlation between gene expression and each module eigengene. GS represents the association between gene expression and each interaction type. A) there is positive correlation with the red module and necrotrophic interaction. B) Negative correlation with all the wild types. C) Slight negative correlation with biotrophic interaction but no significant p-value. D) Negative correlation with drought. E) Positive correlation with symbiotic interaction.

The green module (refer Figure 17 A-E) showed a weak but statistically significant positive correlation with biotrophic interaction ($\text{cor} = 0.059, p = 0.0023$) and a non-significant correlation with symbiotic interaction ($\text{cor} = 0.022, p = 0.26$). The results suggest that the red module, unlike the green module, shows a distinct gene-level association with biotrophic and symbiotic interactions. Through module trait relationship with different microbes and functional enrichment can provide better understanding of the biological relevance of genes within red module.

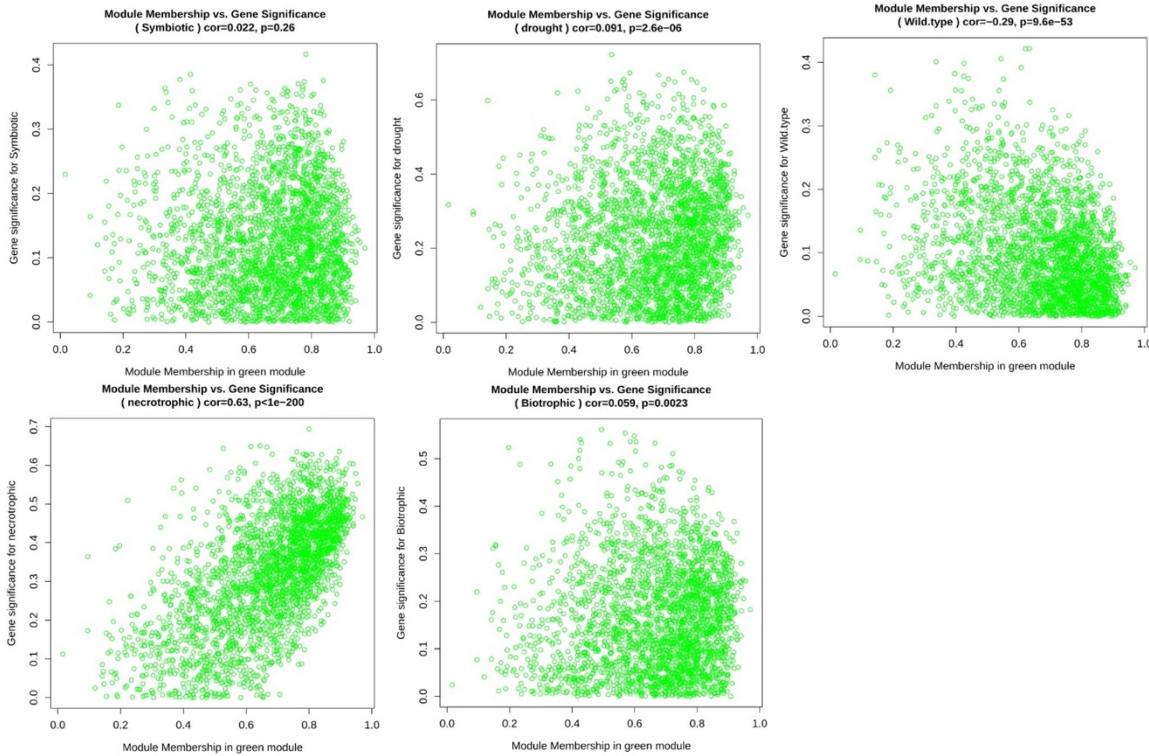


Figure 17 Scatter plot of MM vs GS in MEgreen module.

Figure 17 (A-E) description: MM presents the correlation between gene expression and each module eigengene. GS represents the association between gene expression and each interaction type. A) there is slight positive correlation with the green module and symbiotic interaction. B) Positive correlation with drought. C) Negative correlation with all wild types. D) Positive correlation with necrotrophic interaction. E) Slight positive correlation with biotrophic interaction.

Based on the eigengene heatmap (refer Figure 14), it was evident that the MEyellowgreen module shared the same branch in the dendrogram as biotrophic interaction. The GS vs MM shows that the MEyellowgreen module (refer Figure 18 A-E) has the strongest positive correlation with biotrophic interaction ($\text{cor} = 0.49, p = 5.5\text{e-}13$) while having the strongest negative correlation with symbiotic interaction ($\text{cor} = -0.2, p = 0.0054$), indicating the correlations shown are also biologically relevant.

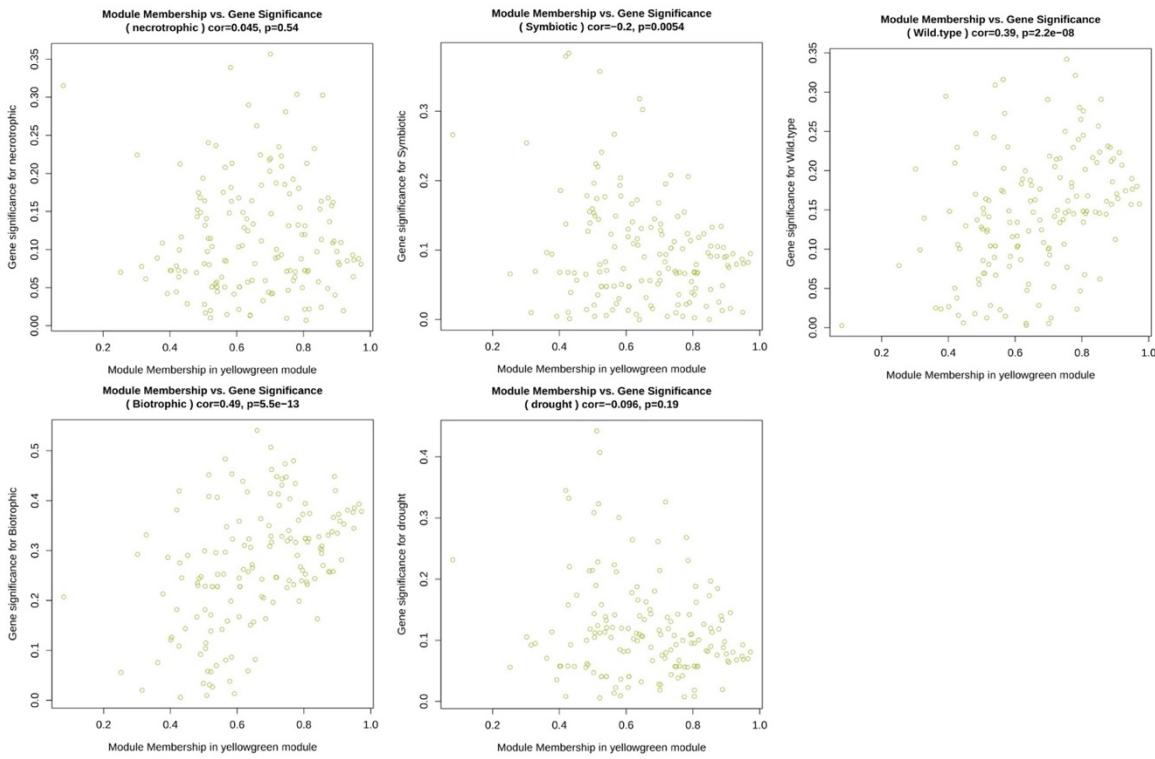


Figure 18 Scatter plot of MM vs GS in MEyellowgreen module.

Figure 18 (A-E) description: MM presents the correlation between gene expression and each module eigengene. GS represents the association between gene expression and each interaction type. A) there is slight positive correlation with the yellowgreen module and necrotrophic interaction. B) Negative correlation with symbiotic interaction. C) Positive correlation with all wild types. D) Positive correlation with biotrophic interaction. E) Slight negative correlation with drought stress.

In the eigengene heatmap (refer Figure 14) of modules with respect to interaction type, MEbrown4 module and METurquoise were closely clustered together. Based on GS vs MM (refer Figure 19 & 20 A-E), some of the interactions show similar kind of correlation. With respect to symbiotic interaction, MEbrown4 (cor = 0.46, $p < 1e-200$) and METurquoise (cor = 0.27, $p = 1.8e-68$) show strong positive statistically relevant correlation, while for biotrophic interaction MEbrown4 and METurquoise show slight positive correlation (cor = 0.0059, $p = 0.71$) yet not statistically relevant and negative correlation (cor = -0.39, $p = 5.8e-147$) which is statistically relevant respectively.

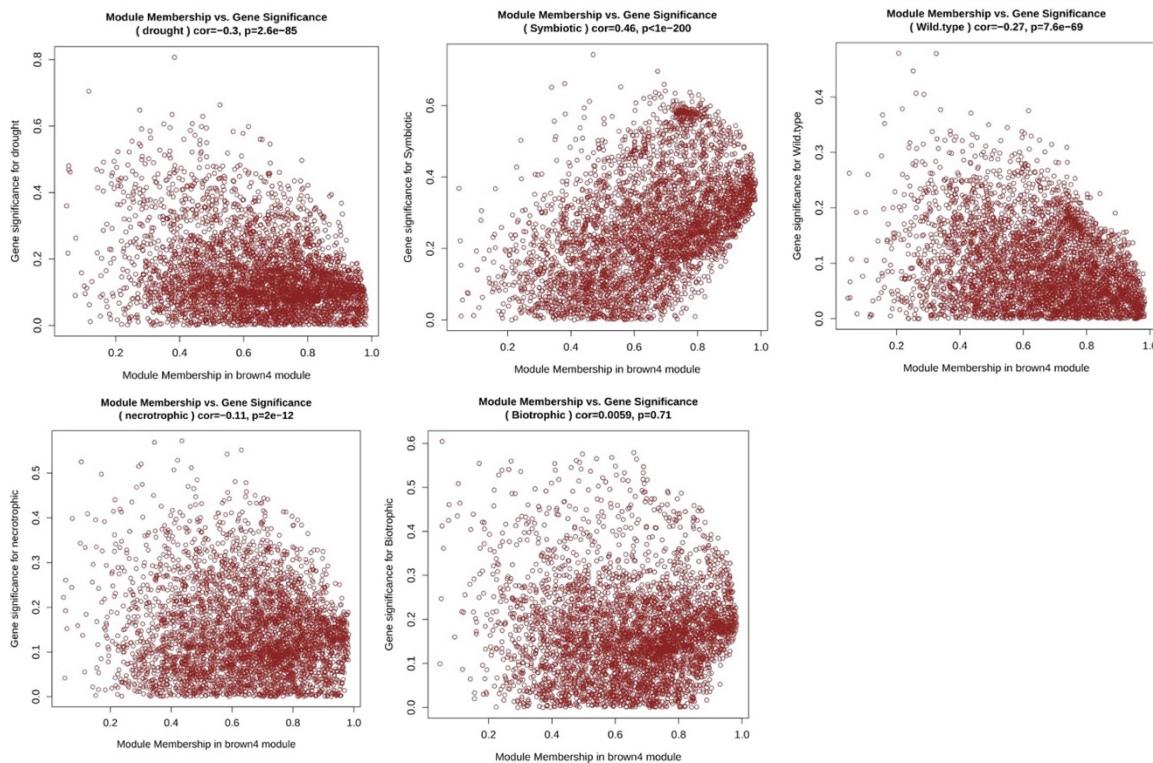


Figure 19 Scatter plot of MM vs GS in MEBrown4 module.

Figure 19 (A-E) description: MM presents the correlation between gene expression and each module eigengene. GS represents the association between gene expression and each interaction type. A) Negative correlation with the brown4 module and drought stress. B) Positive correlation with symbiotic interaction. C) Negative correlation with all wild types. D) Negative correlation with necrotrophic interaction. E) Slight positive correlation with biotrophic interaction.

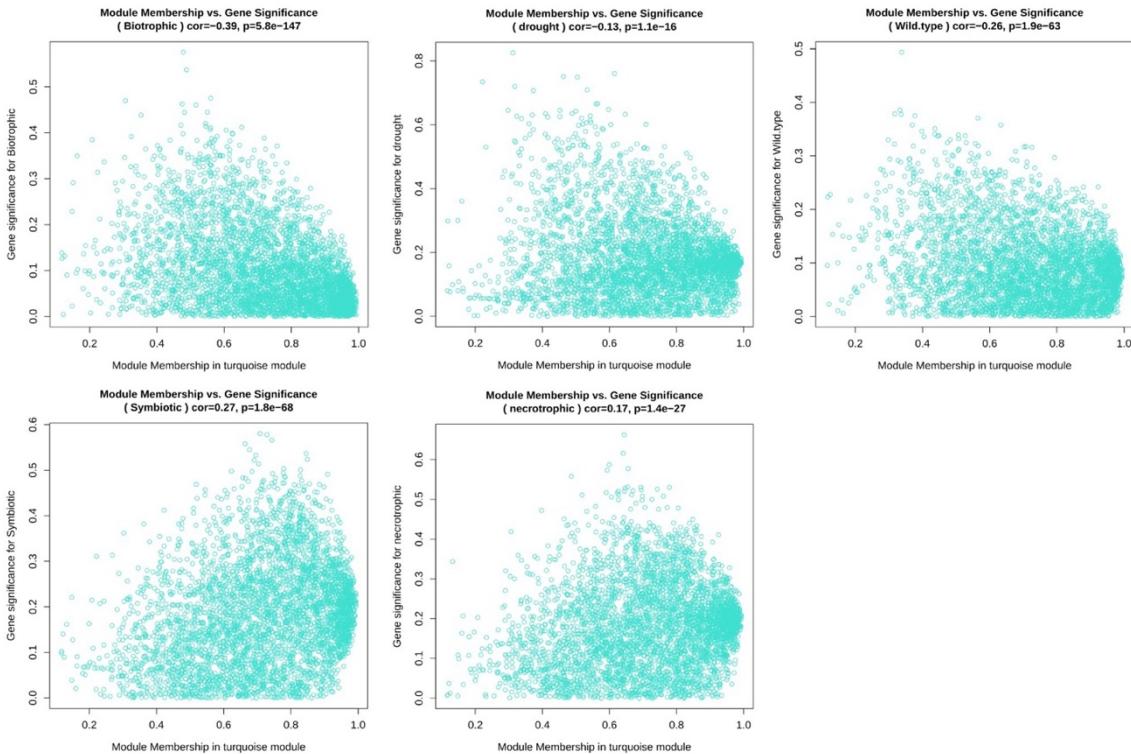


Figure 20 Scatter plot of MM vs GS in METurquoise module.

Figure 20 (A-E) description: MM presents the correlation between gene expression and each module eigengene. GS represents the association between gene expression and each interaction type. A) Negative correlation with the turquoise module and biotrophic interaction. B) Negative correlation with drought stress. C) Negative correlation with all wild types. D) Positive correlation with symbiotic interaction. E) Negative correlation with necrotrophic interaction.

The MEblack modules was closely cluster together with the necrotrophic interaction. In this study, the necrotrophic interaction is considered as a control interaction to elucidate the genes that maybe responsible for plant cell death or stress hormone. The GS shows strong positive correlation ($\text{cor} = 0.73, p < 1e-200$) which is statistically relevant between the MEblack module (refer Figure 21 A-E) and the necrotrophic interaction, indeed suggesting that the genes contributing to this interaction is grouped as one module named as black. Although there is positive correlation for symbiotic ($\text{cor} = 0.3, p = 2.2e-110$) and biotrophic ($\text{cor} = 0.12, p = 2.1e-18$), they are not as strong as the necrotrophic interaction.

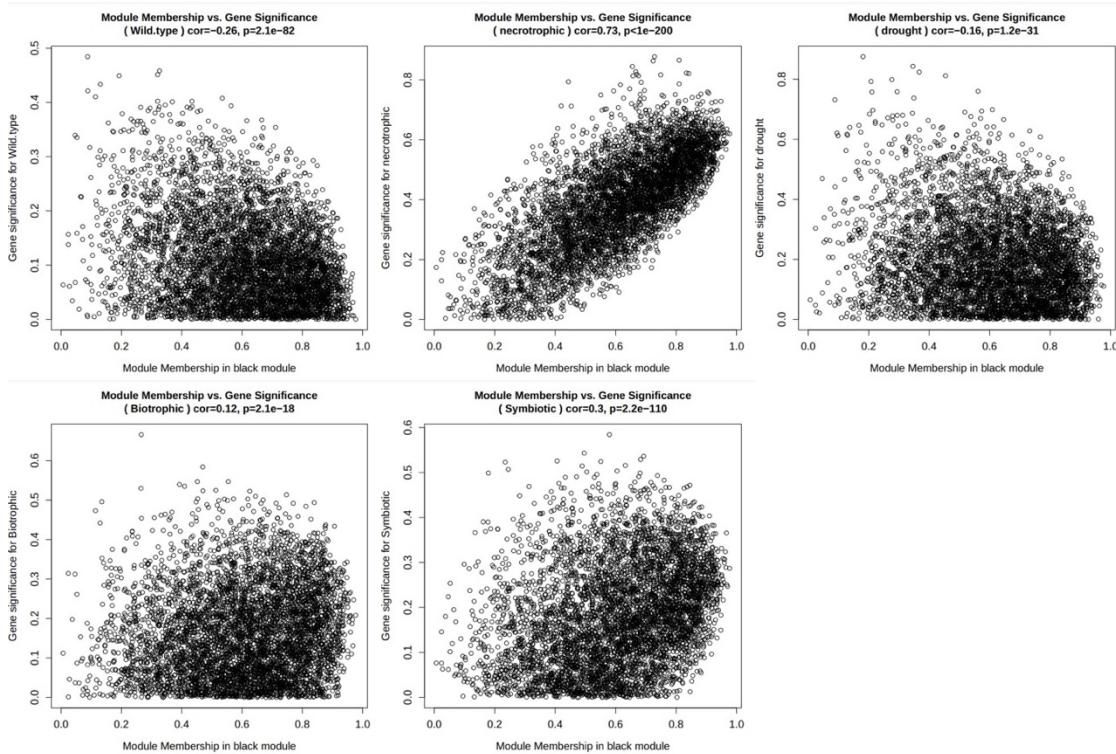


Figure 21 Scatter plot of MM vs GS in MEblack module.

Figure 21 (A-E) description: MM presents the correlation between gene expression and each module eigengene. GS represents the association between gene expression and each interaction type. A) Negative correlation with the black module and all wild type. B) Strong positive correlation with necrotrophic interaction. C) Negative correlation with drought stress. D) Slight positive correlation with biotrophic interaction. E) Slight positive correlation with symbiotic interaction.

Although MElightsteelblue is placed on the same side as the symbiotic interaction on the eigengene dendrogram but not closely clustered. From the GS vs MM, that symbiotic interaction is negatively correlated ($\text{cor} = -0.54, p = 6.2e-05$) with this module (refer Figure 22 A-E) which is statistically relevant but slightly positively correlated ($\text{cor} = 0.17, p = 0.24$) with the biotrophic interaction but this is not statistically relevant observation.

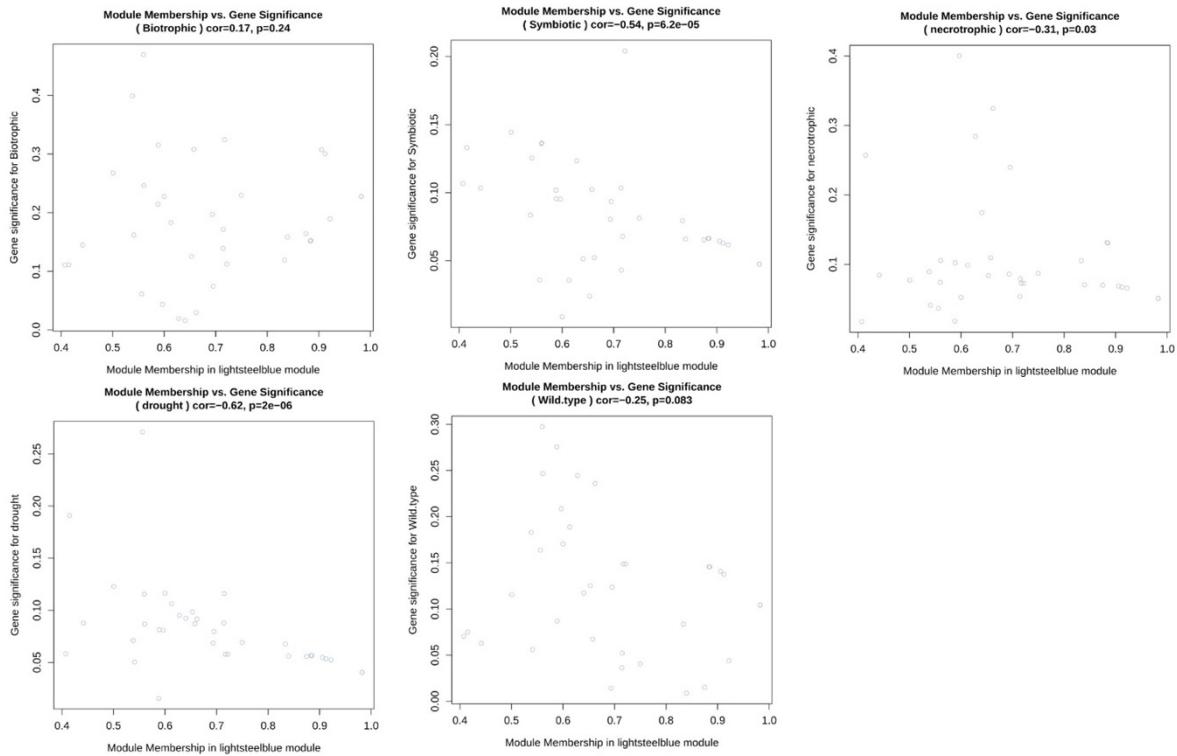


Figure 22 Scatter plot of MM vs GS in MElightsteelblue module.

Figure 22 (A-E) description: MM presents the correlation between gene expression and each module eigengene. GS represents the association between gene expression and each interaction type. A) Positive correlation with the lightsteelblue module and biotrophic interaction. B) Negative correlation with symbiotic interaction. C) Negative correlation with necrotrophic interaction. D) Negative correlation with drought stress. E) Negative correlation with all wild types.

The MEpalevioletred3 was placed closer to the biotrophic interaction on the eigengene heatmap for different interaction types. It is evident from GS vs MM, this module (refer Figure 23 A-E) is positively correlated with biotrophic interaction ($\text{cor} = 0.27, p = 4\text{e-}08$), the p value also indicates that this correlation is statistically significant. With respect to symbiotic interaction, it is negatively correlated ($\text{cor} = -0.096, p = 0.055$). This p value exceeds the used significance threshold (that is 0.05), suggesting that negative correlation is not statistically significant and may not reflect a meaningful relationship. Therefore, these results suggest that MEpalevioletred3 is more relevant to biotrophic interaction than to symbiotic interaction.

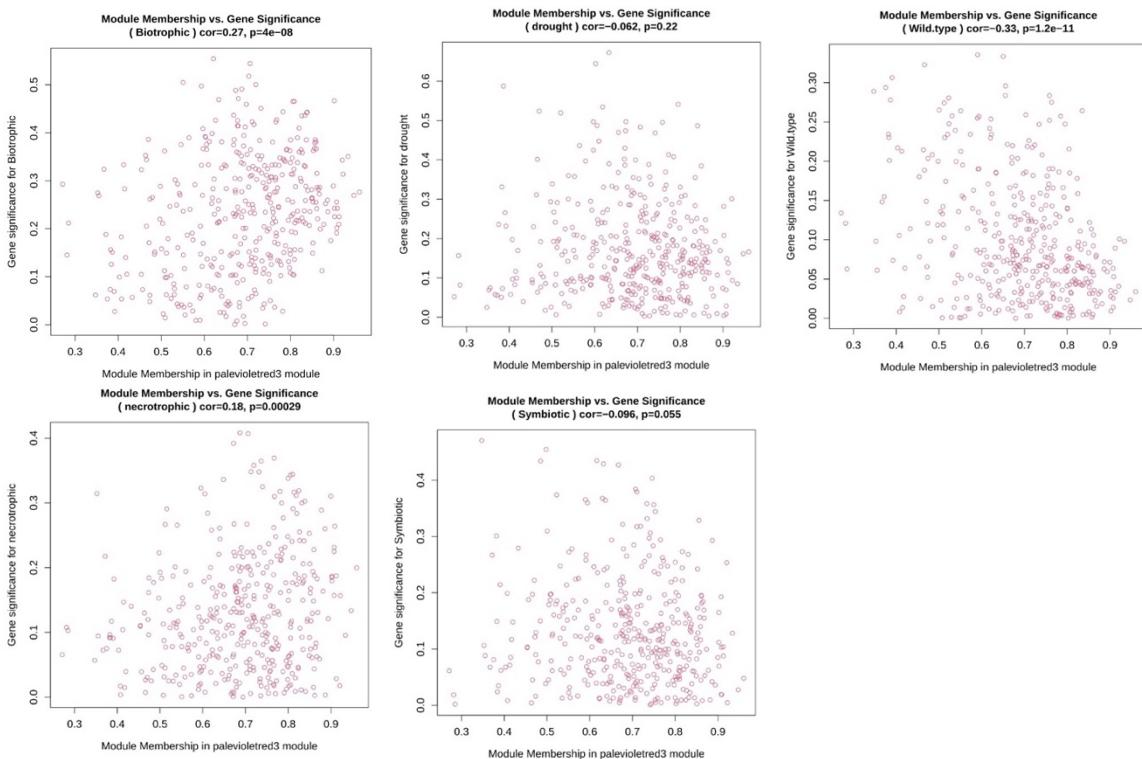


Figure 23 Scatter plot of MM vs GS in MEpalevioletred3 module.

Figure 23 (A-E) description: MM presents the correlation between gene expression and each module eigengene. GS represents the association between gene expression and each interaction type. A) Positive correlation with the palevioletred3 module and biotrophic interaction. B) Negative correlation with drought stress. C) Negative correlation with all wild types. D) Slight positive correlation with necrotrophic interaction. E) Slight negative correlation with symbiotic interaction.

MEorangered3 closely clustered with the symbiotic interaction in the eigengene heatmap dendrogram with interaction types. Based on the GS vs MM it is clearly evident that this module (refer Figure 24 A-E) is only positively correlated with symbiotic interaction ($\text{cor} = 0.57, p = 3\text{e-}06$) which is also statistically significant while being negatively correlated with all other types of interaction.

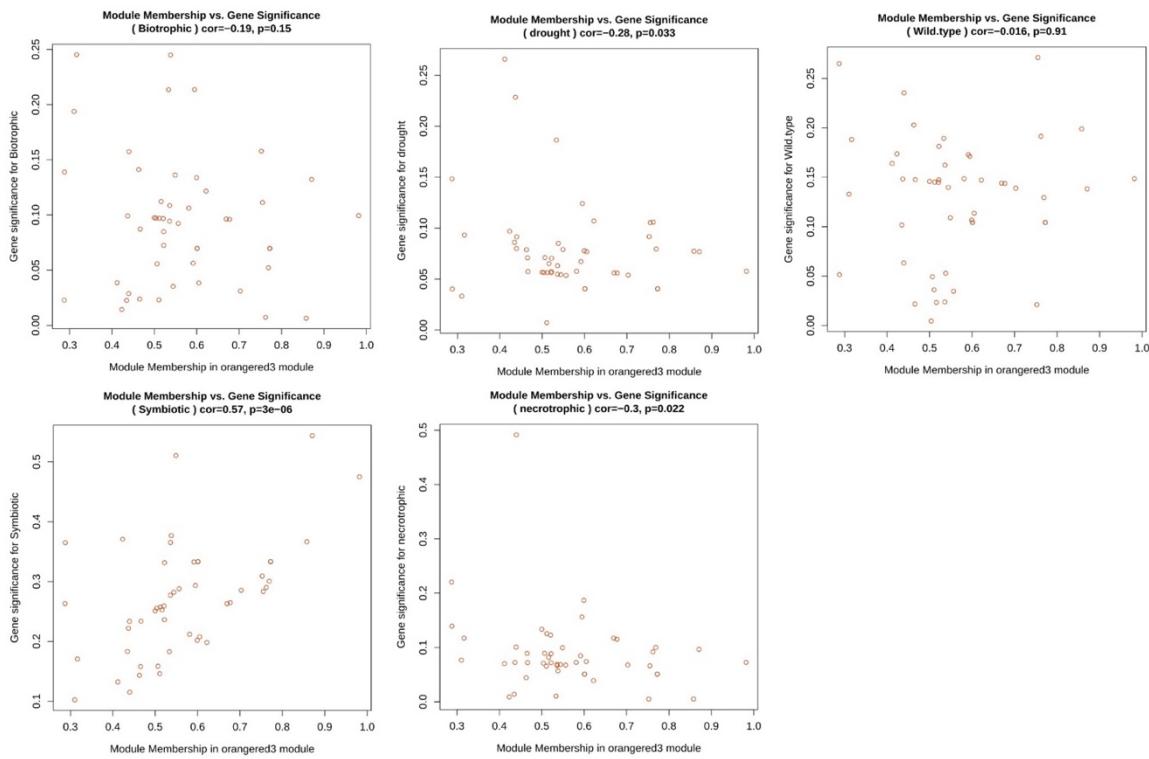


Figure 24 Scatter plot of MM vs GS in MEorangered3 module.

Figure 24 (A-E) description: MM presents the correlation between gene expression and each module eigengene. GS represents the association between gene expression and each interaction type. A) Negative correlation with the orangered3 module and biotrophic interaction. B) Negative correlation with drought stress. C) Negative correlation with wild type. D) Positive correlation with symbiotic interaction. E) Negative correlation with necrotrophic interaction.

MEbisque4 clustered closely with the drought stress in the eigengene heatmap dendrogram with interaction types. This is also one of the control samples included in the study. From GS vs MM it is evident that the genes in this module (refer Figure 25 A-E) are highly positively correlated ($\text{cor} = 0.67, \text{p} = 6.3\text{e-}125$) with drought which is statistically relevant. Biotrophic interaction ($\text{cor} = 0.41, \text{p} = 7.4\text{e-}40$) and symbiotic interaction ($\text{cor} = 0.027, \text{p} = 0.41$) also show positive correlation with this module but are not statistically significant.

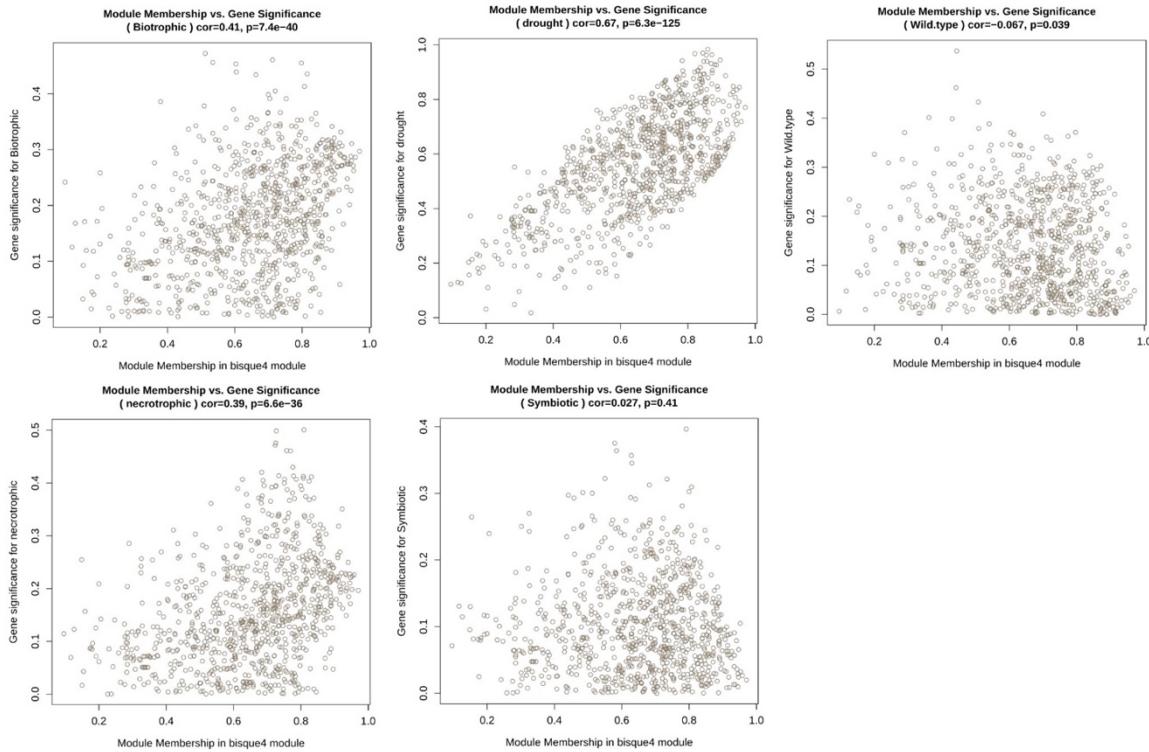


Figure 25 Scatter plot of MM vs GS in MEbisque4 module.

Figure 25 (A-E) description: MM presents the correlation between gene expression and each module eigengene. GS represents the association between gene expression and each interaction type. A) Positive correlation with the bisque4 module and biotrophic interaction. B) Strong positive correlation with drought stress. C) Negative correlation with all wild type. D) Positive correlation with necrotrophic interaction. E) Slight positive correlation with symbiotic interaction.

vii. Module-trait relationships and module eigengene expression across microbes

The module trait relationship (refer Figure 26) illustrates the correlation between the gene co-expression modules and distinct microbes (refer Table 1). Each cell represents a Pearson correlation coefficient with its associated p value, where red shades indicate positive correlations and blue shades indicate negative correlation of a module eigengene to specific microbes listed in Table 2.

Table 2 Different modules that are correlated with their corresponding microbes.

Module	Microbe	Interaction
Brown4	Arbuscular Mycorrhizal (AM)	Symbiotic
Antiquewhite4	Serendipita Indica (SI)	Symbiotic
Orangered3	Funneliformis Mosseae (FM)	Symbiotic
Lightsteelblue	Cucumber Mosaic Virus (CMV)	Biotrophic
Palevioletred3	Cladosporium Fulvum (CF)	Biotrophic
Turquoise	Meloidogyne Incognita (MI)	Biotrophic
Yellowgreen	Oidium Neolycopersici-powdery mildew (ON)	Biotrophic
Black	Botrytis Cinerea (control) (BC)	Necrotrophic
Bisque4	Drought stress (control) (D)	Drought

Table 2 description: Different modules that are correlated with their corresponding microbes based on the eigengene heatmap and dendrogram (refer figure 27).

Based on figure 26 the MEred module showed slight positive correlations with a range of microbes including symbiotic interactions such as AM (cor = 0.21, p = 0.1), SI (cor = 0.17, p = 0.2), FM (cor = 0.3, p = 0.02), and biotrophs such as CMV (cor = 0.15, p = 0.2), CF (cor = 0.16, p = 0.2) and MI (cor = 0.22, p = 0.08), suggesting that this module has genes within this module may be part of a common signaling or defense-related pathway that is broadly activated upon contact with a range of microbial organisms, regardless of their interaction type.

The MEgreen module showed positive correlations with multiple microbes, including symbionts such as AM (cor = 0.26, p = 0.04), SI (cor = 0.091, p = 0.5), CF (cor = 0.2, p = 0.1), MI (cor = 0.047, p = 0.7) and ON (cor = 0.16, p = 0.2), as well as with drought stress (D), while displaying negative correlations with biotrophs such as FM (cor = -0.14, p = 0.3) and CMV (cor = -0.11, p = 0.4). The nature of these associations suggests that the MEgreen may contain genes involved in crosstalk between biotic and abiotic stress responses, potentially through shared signaling mechanisms such as hormone regulation or general stress responsive factors.

The MEyellowgreen modules was positively correlated specifically with ON (cor = 0.98, p = 4e-47), a powdery mildew pathogen. This specificity implies that the module may harbor genes that are particularly responsive to specific fungal infections, possibly contributing to pathogen recognition or localized defense responses such as the hypersensitive response.

The MEbrown4 modules is positively correlated with the AM (cor = 0.74, p = 2e-12) and negatively correlated with rest of the microbes, indicating a high level of interaction type specificity. This suggests that this module likely has genes that are selectively activated during symbiotic associations, and may be suppressed or inactive during pathogenic interaction, potentially contributing to mutualism specific signaling or metabolic process. ME orangered3 module showed strong positive correlation with FM. This response suggests that this module may contain genes that are involved in gene in AM symbiosis pathway.

ME turquoise was strongly positively correlated with MI ($\text{cor} = 0.5$, $p = 2\text{e-}05$) and slightly positively associated with AM and SI. This may suggest that the modules may have genes that are involved in responses to root associated organisms, whether pathogenic (MI) or symbiotic (AM and SI). The strong MI signal implies that the module may play a role in root invasive defense, while AM and SI associations could reflect overlapping root signaling pathways.

MEblack module showed a positive correlation with *Botrytis cinerea* (BC) ($\text{cor} = 0.64$, $p = 1\text{e-}08$) and a slight positive correlation with ON ($\text{cor} = 0.2$, $p = 0.1$), both of which are fungal pathogens. The strong positive correlation with BC indicates that the expression of genes in this module is likely upregulated or significantly associated with infection by this pathogen.

MEbisque4 module was positively correlated with drought stress (D) ($\text{cor} = 0.86$, $p = 2\text{e-}19$), as also supported by GS vs MM scatter plot. This module likely contains genes involved in abiotic stress adaptation. Additionally, the modules showed a very weak positive correlations with AM interaction, though not statistically significant. This slight overlap may reflect a shared set of genes that respond to root associated cues, although the module appears to be primarily associated with drought adaptation rather than microbial infection.

MElightsteelblue module showed positive correlation with CMV($\text{cor} = 0.62$, $p = 5\text{e-}08$), suggesting a role in the plant's transcriptional response to viral infection. In contrast, it exhibited negative correlation with all other microbial infections, indicating a potentials virus specific expression profile. This pattern implies that genes within this module may be involved in CMV triggered defense responses, while being suppressed or inactive during other microbial infection.

The MEpalevioletred3 module shows a positive correlation with CF ($\text{cor} = 0.65$, $p = 5\text{e-}09$) a known biotrophic fungal pathogen. With respect to the GS vs MM, it is evident that this module has genes that are activated during a biotrophic fungal infection.

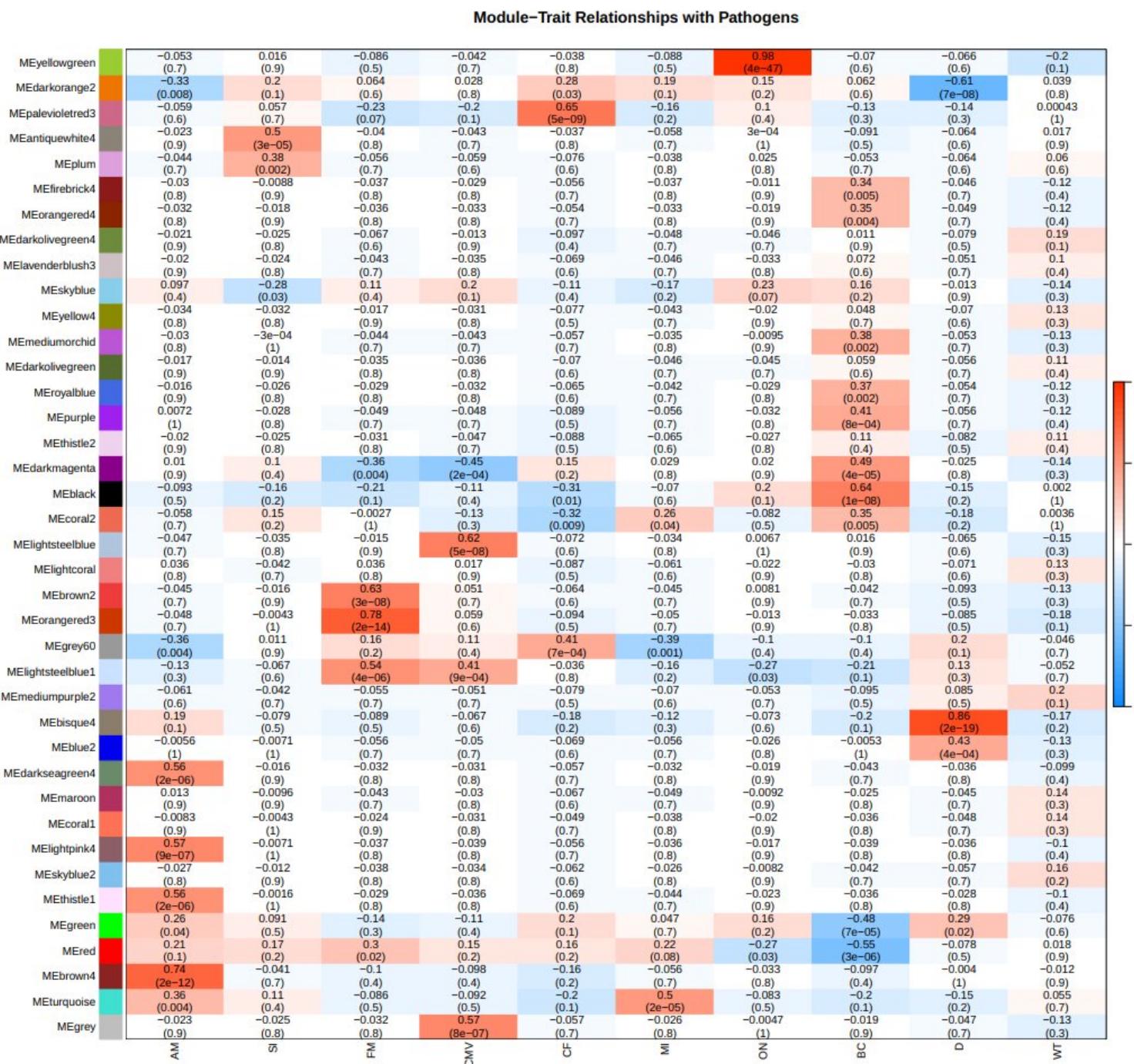


Figure 26 A module trait relationship plot for different microbes.

Figure 26 description: The X axis represents the microbes along with the wild type and controls which as BC (necrotroph) and D which is Drought (Abiotic stress). The Y axis represents the 39 modules obtained after module merging.

The hierarchical clustering dendrogram and associated heatmap (refer Figure 27) represents the relationships between module eigengene and various microbial interaction. From this heatmap, we get an even more clear understanding of the different modules that is associated with each of the microbes (refer Table 2). Although some of the modules were identified as specific for the microbes from the module trait relationship across microbes, it is much more evident from the dendrogram and heatmap.

The eigengene heatmap for microbes revealed that the MEred and MEGreen modules are closely associated, suggesting that the genes may contribute to the first line of defense response when a plant is in contact with the microbes. This association is also noticed in the eigengene heatmap and dendrogram plotted for the different interaction types. MEyellowgreen module showed positive correlation with biotrophic interaction in GS vs MM scatter plot and in the eigengene dendrogram we notice that the module is closely linked to ON, another indication of strong correlation was the branch of the dendrogram being closer to the heatmap. Based on the GS vs MM scatter plot the MEorangered3 showed strong positive correlation with symbiotic interaction and through figure 26 and 27 it is evident that the module has genes that contribute for the FM interaction with tomato. MElightsteelblue is closely correlated with CMV, and this was evident also through the GS vs MM scatter plot showing correlation with biotrophic interaction while being negatively correlated with symbiotic interaction. The brown4 module is linked to AM fungus. GS vs MM scatter plot also showed positive correlation with statistical significance for brown4 module with symbiotic interaction. Black and bisque4 modules was also associated closely with BC and D in the eigengene dendrogram (refer Figure 27) and is also illustrated in previous eigengene interaction dendrogram (refer Figure 15).

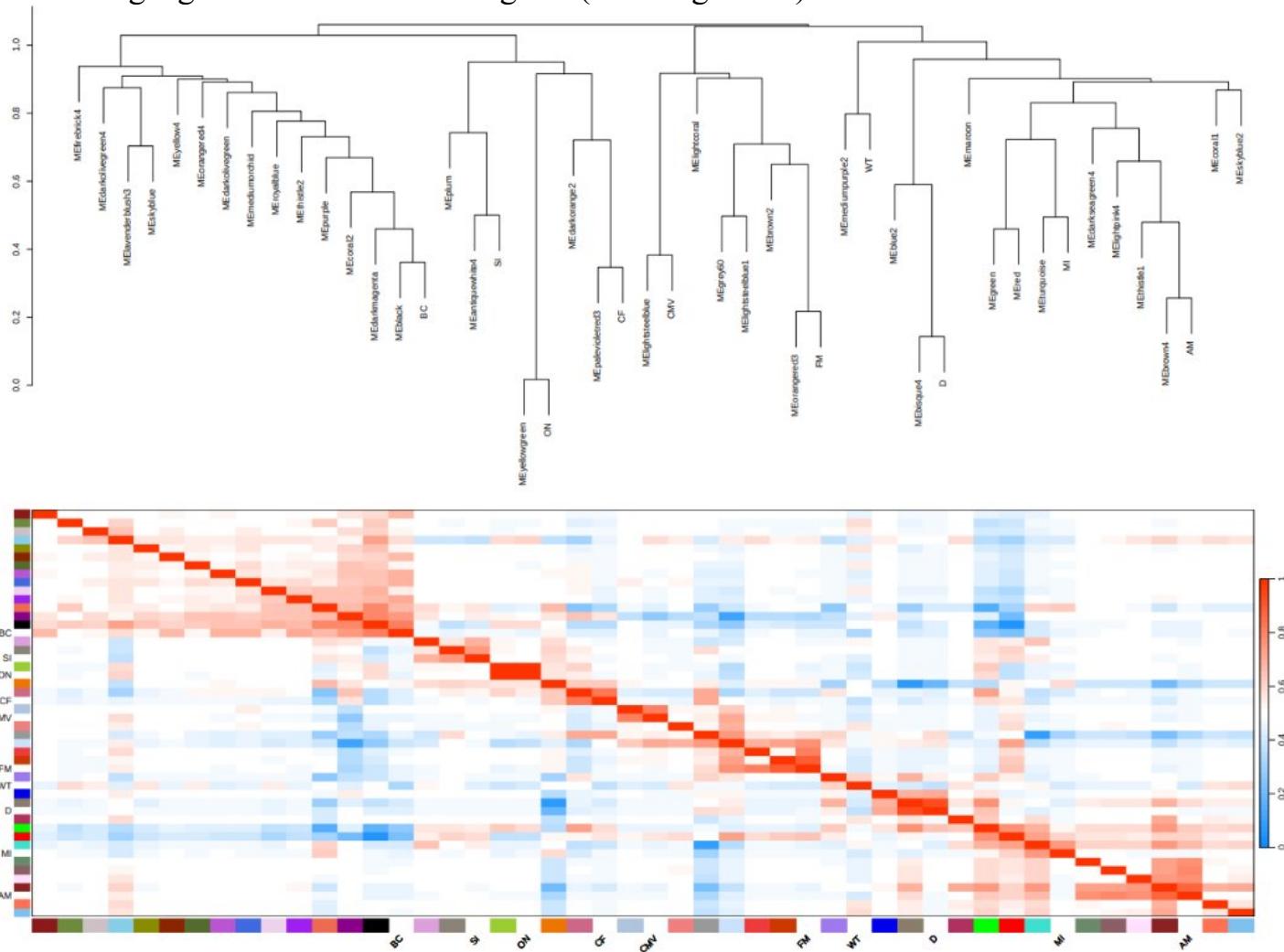


Figure 27 Dendrogram and heatmap of module eigengene and microbes.

Figure 27 description: The top panel shows hierarchical clustering of modules eigengenes and different microbes based on expression similarity. The bottom heatmap visualizes pairwise correlations, highlighting distinct clustering of micro-organisms associated modules from control organisms.

viii. Functional enrichment of WGCNA

Functional enrichment analysis is essential for interpreting large gene expression datasets by identifying biological processes, pathways, or molecular functions that are statistically overrepresented among a set of genes(Lohse et al., 2014)(Mi et al., 2017). One key metric of functional enrichment analysis is the gene ratio which represents the proportion of genes in the input list that are associated with a specific functional category relative to the total number of genes annotated with that function in the reference background(Blüthgen et al., 2005). A higher gene ratio suggests that a large fraction of the genes in the module are involved in that specific function indicating its potential biological importance in simple terms higher gene ratio suggests stronger enrichment(Subramanian et al., 2005).

The functional enrichment analysis of the red module (refer Figure 28) which is shared positive correlation among AM, SI, FM, CMV, CF and MI it reveals a coordinated transcriptional program which is involved in key cellular and structural processes, especially those underpinning cell division, cytoskeleton dynamics, and cell wall biosynthesis. The dot plot demonstrates the enrichment of several gene ontology (GO) terms based on their gene ratios represented on the X axis and statistical significance, which is represented with color intensity, with larger dot sizes indicating a larger number of genes contributing to each term. The most significantly enriched term “cell division, cell cycle organization, cell cycle control” involving CYCLIN regulatory protein CYCB, emphasizing the module’s role in regulating mitosis and cell cycle progression which is potentially manipulated during both symbiotic and pathogenic interactions (Suzuki et al., 2021). The enrichment of “protein modification via phosphorylation” involving LRR-XI protein kinase points to signal transduction as a central feature, possibly mediating the plant’s response to external stimuli from the diverse microbes (Shiu & Bleecker, 2003).

Another prominent cluster of enriched functions involves cytoskeleton organization, particularly microtubular motor proteins like Kinesin-12, Kinesin-10, Kinesin-14B, and Kinesin-5, indicating an emphasis on vesicle trafficking, spindle assembly and intracellular transport, which are all essential during host pathogen or symbiont interactions(Y. Wu & Zhou, 2013). Terms like “cell wall organization, cellulose synthase complex” and “cell wall organization with fasciclin-like arabinogalactan proteins (FLA11)” highlighting the role of this module in cell wall reconstructing, likely to fortify the plant during nematode penetration or fungal colonization. The inclusion of chromosome segregation (via TPX2) and cytokinesis related microtubule activities further illustrates that tight control of cellular architecture and division is critical when responding to such diverse microbial pressures (Ma et al., 2022) (K. Yao et al., 2023). Overall, the red module appears to orchestrate a comprehensive cellular response such as balancing growth, division, and defense, enabling the host plant to effectively engage with both beneficial and harmful microbes across a spectrum of biotic interaction.

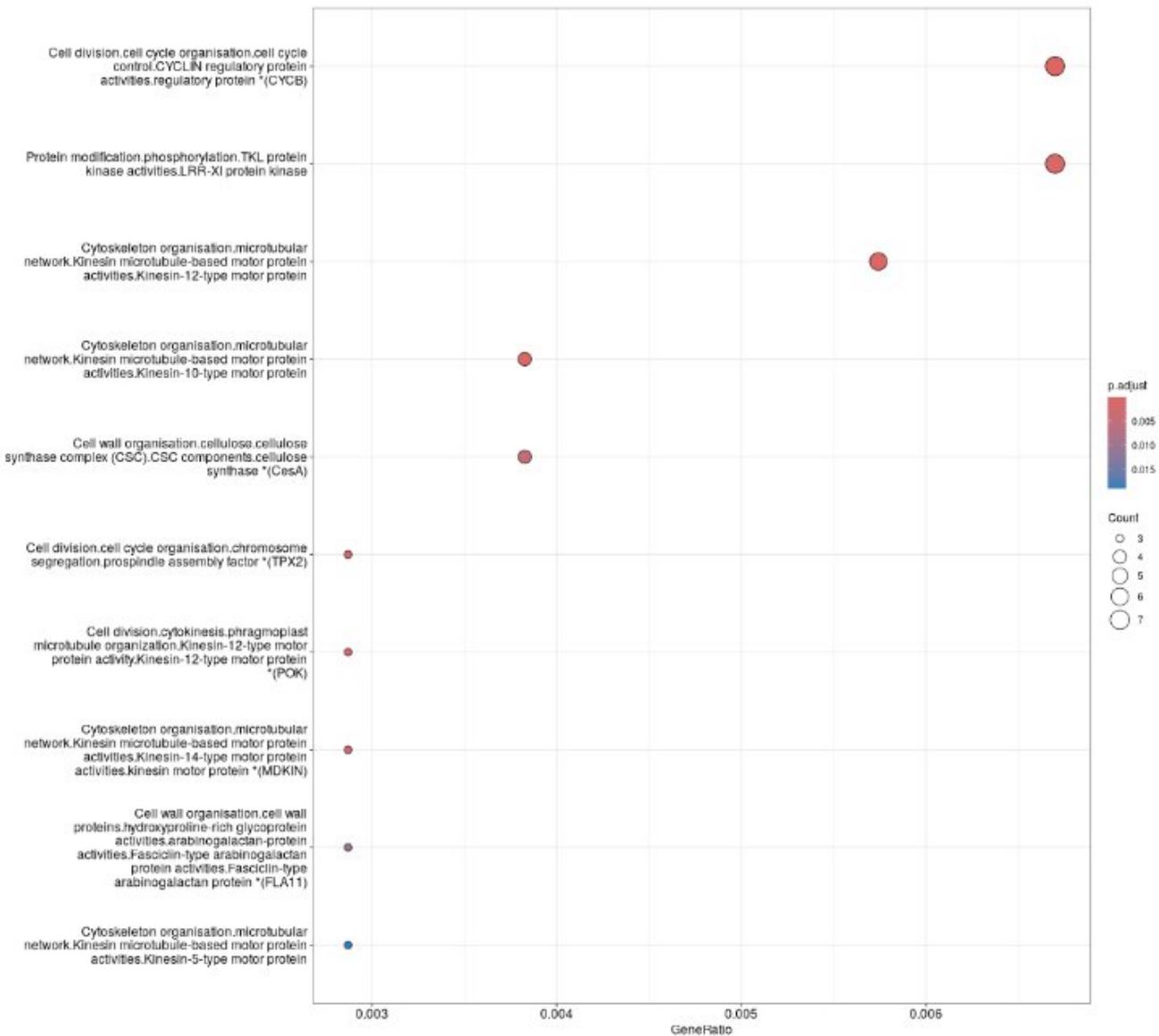


Figure 28 Functional enrichment of Red Module (WGCNA).

Figure 28 description: The Y axis represents the enriched terms in the module and X axis represents the gene ratio. Each dot size represents the number of genes contributing to that enriched term and color gradient is the p adjusted value.

The functional enrichment analysis of green module (refer Figure 29), which is shared among AM, SI, CF, MI and ON and drought stress, reveals a singular but significant functional category “External stimuli response, salinity, gibberellin-abscisic acid signaling pathways crosstalk, E3 ubiquitin ligase (XERICO)”. This enrichment indicates a highly focused transcriptional response, where only one term is significantly represented with a gene ratio near 0.01 and p adjusted value of ~0.004, emphasizing statistical confidence while also having a low count of associated genes (n = 3).

The biological implication of this result is particularly interesting because the XERICO gene encodes for E3 ubiquitin ligase known to modulate abscisic acid (ABA) accumulation, which is critical in the abiotic stress tolerance, especially drought and salinity. Its enrichment here suggests

that the green module is tightly linked to ABA mediated stress response, integrating hormonal crosstalk between ABA and gibberellin, often antagonistic in developmental and stress context(D.-E. Zeng et al., 2015)(Shu & Yang, 2017).

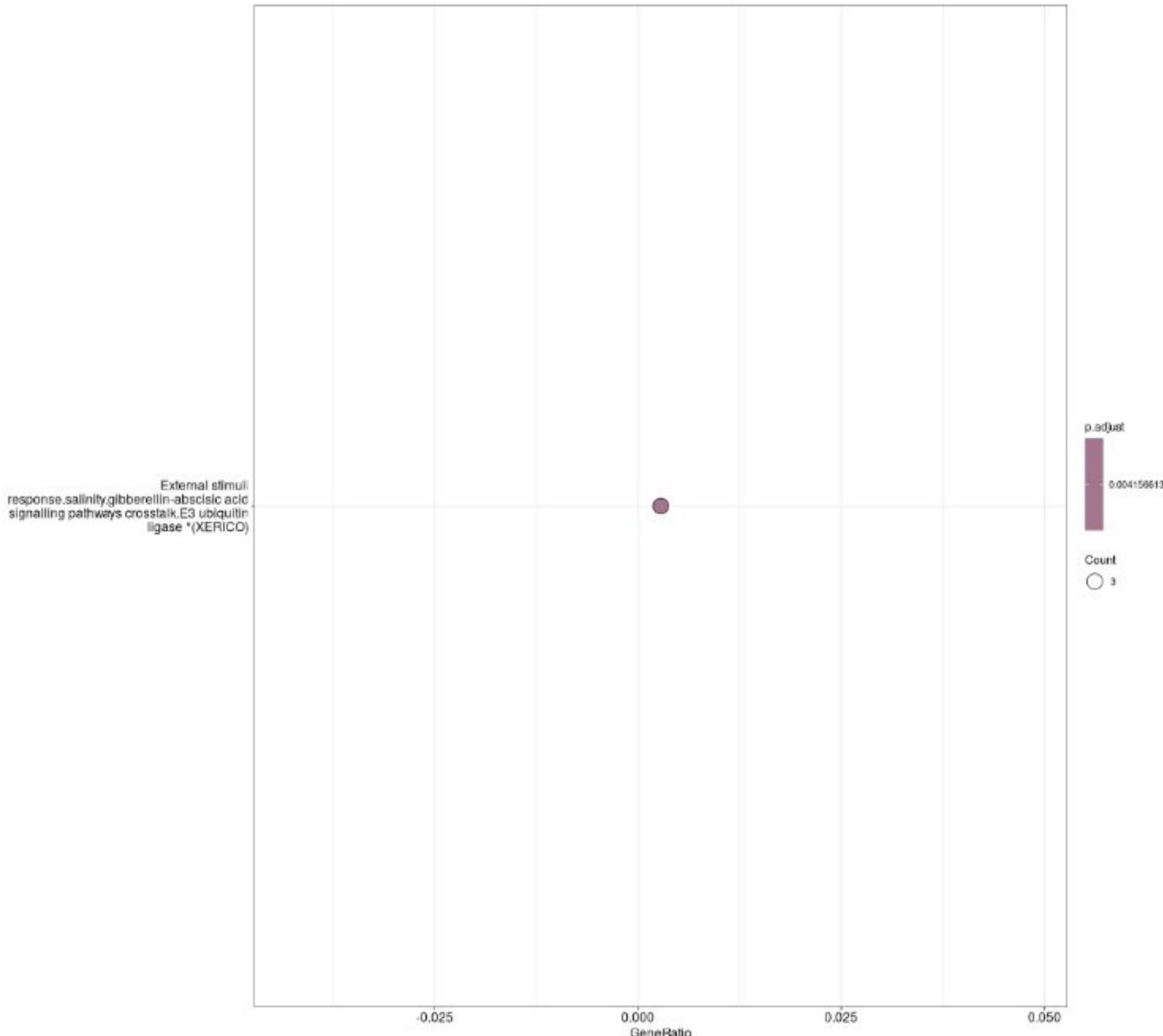


Figure 29 Functional enrichment of Green Module (WGCNA).

Figure 29 description: The Y axis represents the enriched terms in the module and X axis represents the gene ratio. Each dot size represents the number of genes contributing to that enriched term and color gradient is the p adjusted value.

The Yellowgreen module (refer Figure 30) which correlates with ON, is significantly enriched for protein S-glutathionylation/glutaredoxin activity, indicating a pronounced role in redox-based post translational regulation (Matsui et al., 2020). This enrichment is underscored by the highest ratio (~0.05) and low p adjusted value of ~0.025, reflecting a sizable proportion of module genes involved in S glutathionylation a reversible modification regulation protein function under oxidative stress. Glutaredoxins (Grxs) catalyze these thiol-dependent reversible changes and are pivotal in maintaining redox homeostasis (Kalinina & Novichkova, 2021).

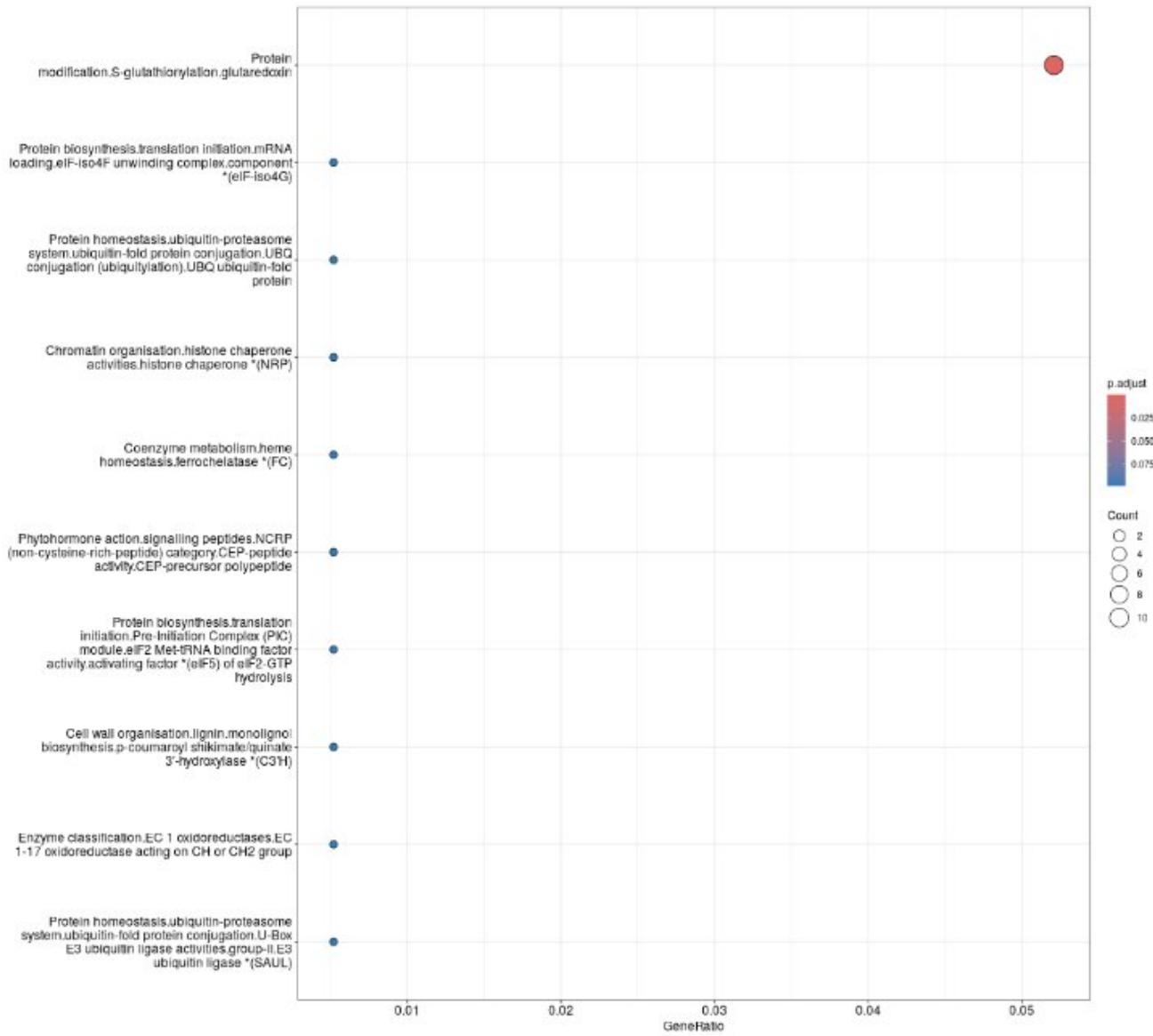


Figure 30 Functional enrichment of Yellowgreen Module (WGCNA).

Figure 30 description: The Y axis represents the enriched terms in the module and X axis represents the gene ratio. Each dot size represents the number of genes contributing to that enriched term and color gradient is the p adjusted value.

The functional enrichment of brown4 module (refer Figure 31) shows positive correlation with AM symbiosis, is notably enriched for pathways involving photosynthesis, specifically photophosphorylation and the core reaction center proteins of Photosystems I and II (e.g., D1/PsbA, PsbB, PsbC, PsbD, PsaA), as well as protein translocation into the chloroplast (TIC20/YCF1 complex) and importantly, nutrient exchange mechanisms, particularly nitrate update and transporter activity (eg., NRT1-1). The highest gene ratio (up to ~0.03) and strong statistical significance (adjusted p value of ~0.01) underscores the module's focused regulation of chloroplast assembly and photosynthetic machinery in response to AM colonization.

The upregulation of photosystem core proteins (PsbA, PsbB, PsbC, PsbD, PsaA) and chloroplast import machinery (TIC20/YCF1 complex) reflects the increased demand for energy and

metabolic capacity during symbiosis. AM fungi are known to enhance photosynthetic efficiency in host plants, not only by boosting carbon assimilation but also by regulating genes involved in chloroplast development and protein import (Moustakas et al., 2020)(Rochaix, 2022).

The enrichment for nitrate transporter activity (e.g., NRT1-1) links this module to nutrient exchange. In return for this photosynthetically derived carbon, AM fungi facilitate the acquisition of key nutrients such as nitrogen and phosphorous, and plants often upregulate nitrate and ammonium transporters in response. This dual activation of photosynthesis and nutrient uptake pathways supports a tightly integrated relationship between host plant and the fungal symbiont (Sanchez-Bel et al., 2016) (Ahmed et al., 2025).

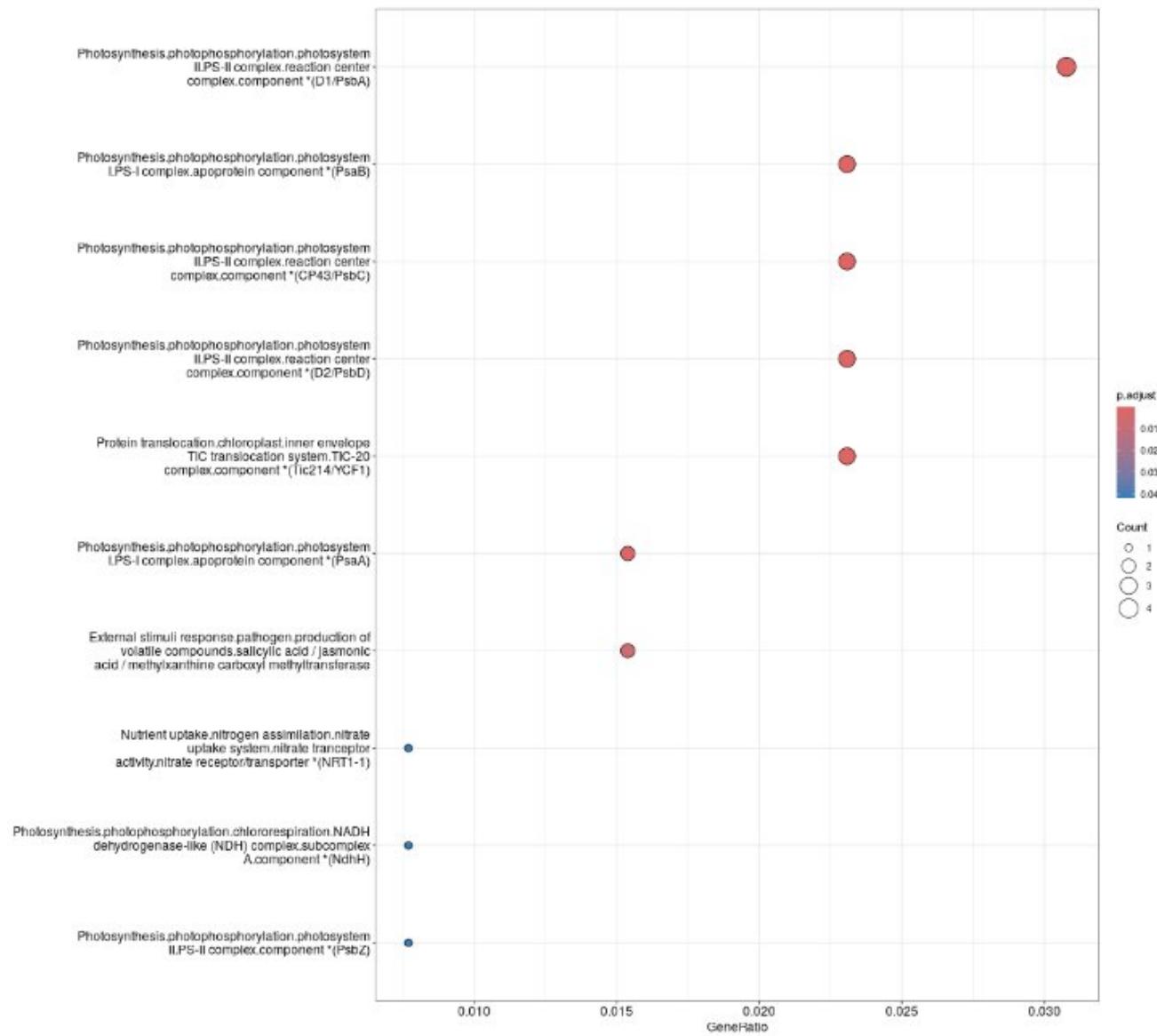


Figure 31 Functional enrichment of Brown4 Module (WGCNA).

Figure 31 description: The Y axis represents the enriched terms in the module and X axis represents the gene ratio. Each dot size represents the number of genes contributing to that enriched term and color gradient is the p adjusted value.

The black module (refer Figure 32), upregulated in response to *Botrytis cinerea* infection, is enriched for protein kinases (including CDPK and CAMK subfamilies), ERF/MYB – type transcription factors, vesicle trafficking (e.g., TOL ubiquitin adaptors), and a variety of metabolic enzymes such as oxidoreductases, hydrolases, and amino acid biosynthesis (e.g., ADH). The top functional category, transferases (EC 2), shows the highest gene ratio (~0.07) and strong statistical significance, suggesting active post translational and metabolic reprogramming. Enriched kinases likely represent botrytis induced kinases like BIK1, known to regulate defense hormone signaling during necrotrophic fungal attack (González et al., 2013)(Badmi et al., 2022). Similarly, ERF transcription factors (e.g., ERF16/MYB306 in grapevine) have been shown to amplify defense gene expression and confer resistance to *B.cinerea* (Zhu et al., 2022).

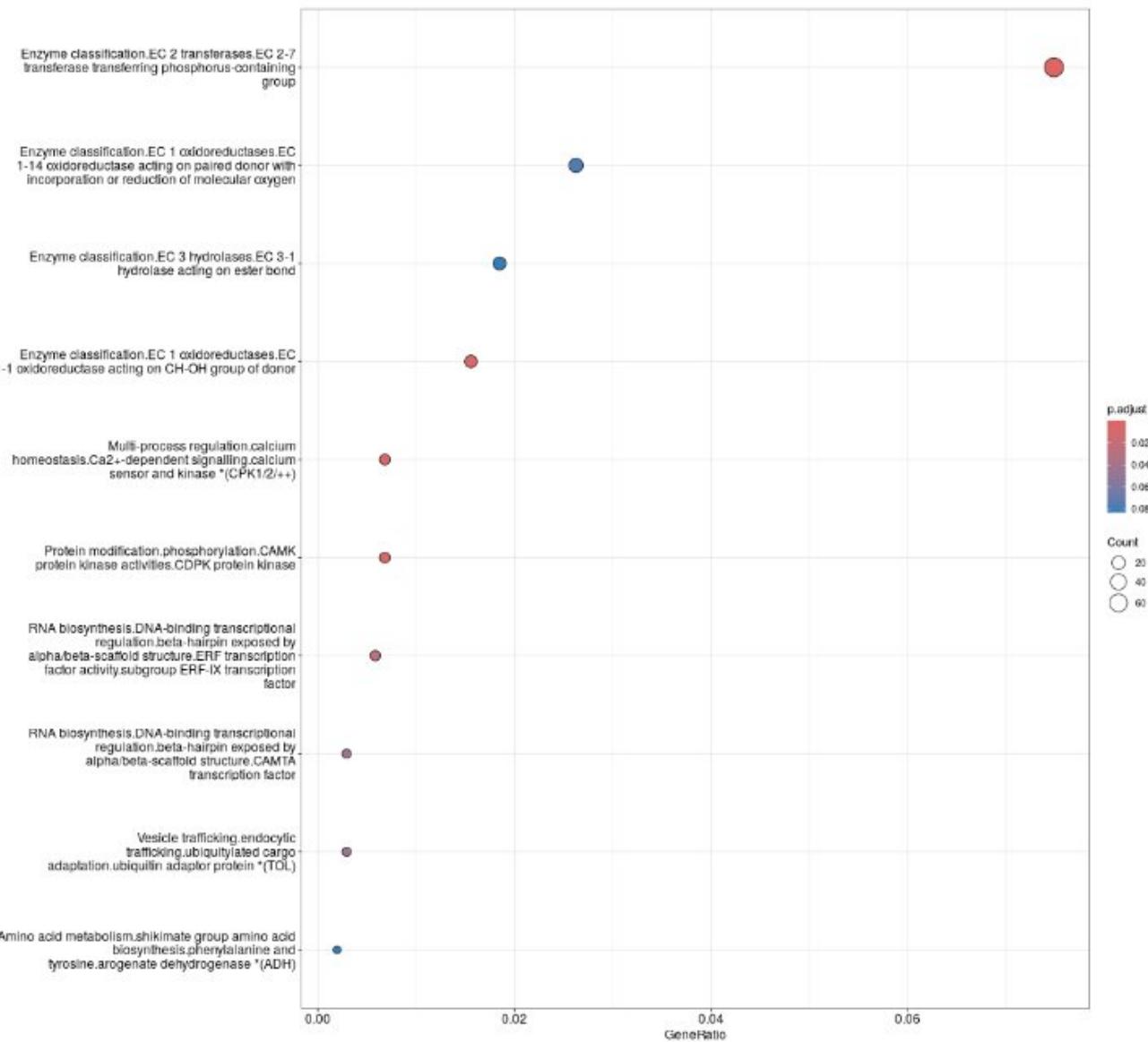


Figure 32 Functional enrichment of Black Module (WGCNA).

Figure 32 description: The Y axis represents the enriched terms in the module and X axis represents the gene ratio. Each dot size represents the number of genes contributing to that enriched term and color gradient is the p adjusted value.

The Lightsteelblue module (refer Figure 33), upregulated by CMV, is enriched for fundamental processes like ribosome biogenesis and large ribosomal subunit assemble (e.g., UL29, eIF6), protein ubiquitin proteasome regulation (CUL4-DCAF E3 ligases), hemicellulose and arabinogalactan cell wall biosynthesis (B3GALT enzymes) and ZHOUPI transcription factors implicated in seed and endosperm development. The high gene ratios (~0.03 – 0.06) and robust p adjusted values (< 0.015) suggest CMV drives a coordinated induction of translation machinery and post translational quality control likely facilitating viral replication. Ribosome and proteosome upregulation may reflect a hijacking of host protein synthesis and turnover mechanisms to favor viral protein production and to modulate immune receptors (Real et al., 2025) (Šubr et al., 2020).

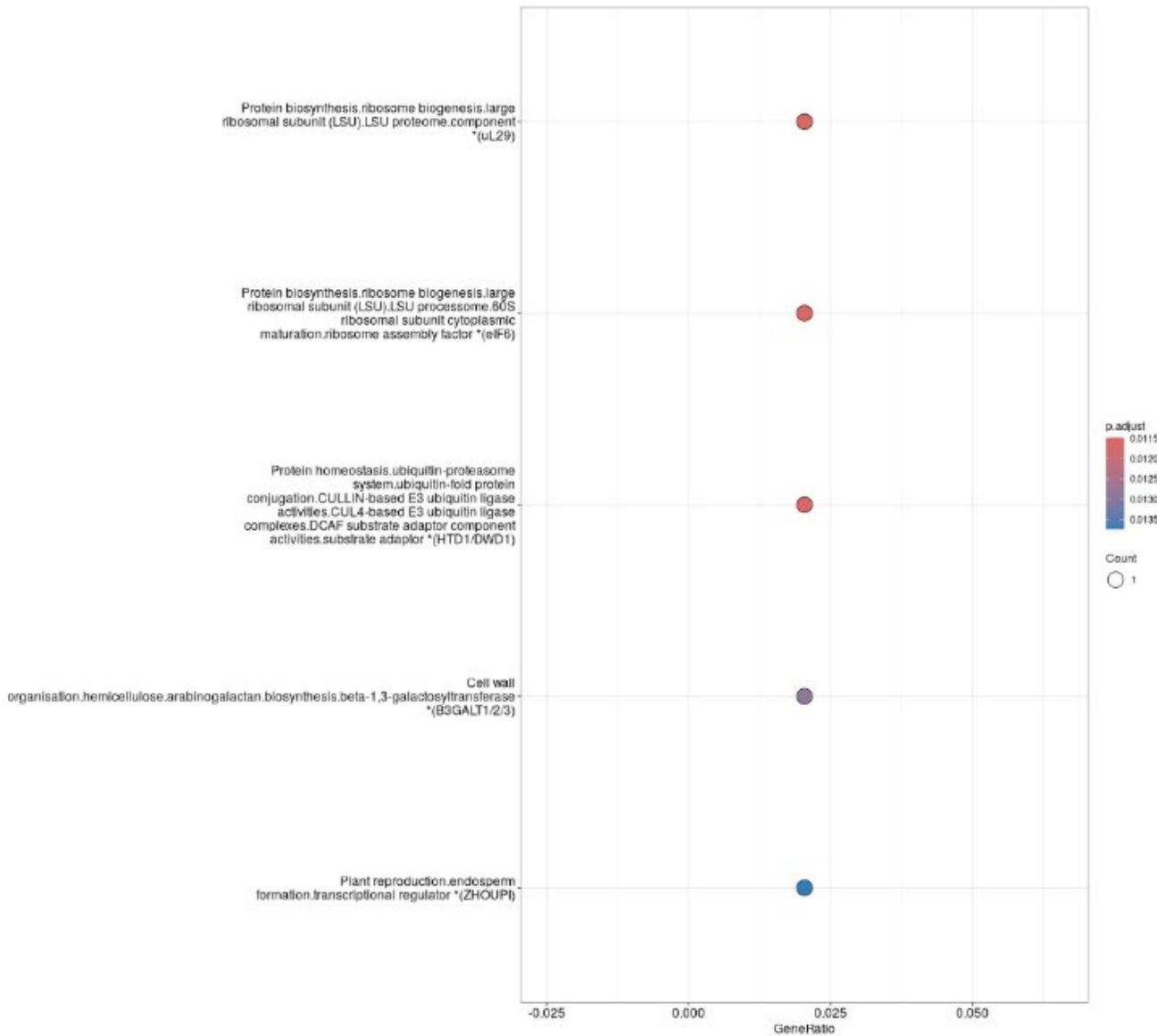


Figure 33 Functional enrichment of Lightsteelblue Module (WGCNA).

Figure 33 description: The Y axis represents the enriched terms in the module and X axis represents the gene ratio. Each dot size represents the number of genes contributing to that enriched term and color gradient is the p adjusted value.

The palevioletred3 module (refer Figure 34), which is positively correlated with CF infection, is enriched for genes encoding hydrolases acting on carbon nitrogen bonds, cell wall organization (cellulose synthase complex, CSL trafficking regulator TVA), sieve element development (pectate lyase PLL12), ribosome biogenesis (eIF6), programmed cell death (PCD) regulators like CYC LASE, vesicle trafficking (E-class Rab-GTPase), ABA biosynthesis and signaling (abscisic aldehyde oxidase, AAO), and phosphatidylinositol signaling (PI-PLC). The gene ratio (~ 0.02) and significant p adjusted value of ~ 0.065 indicate a concerted activation of both structural defense and signaling pathways during CF invasion.

The enrichment of cell wall biosynthesis and remodeling enzymes such as pectate lyases and cellulose synthase regulators supports reinforcement of the apoplast against fungal penetration (de Wit et al., 2012) (Zhao et al., 2019). The elevated expression of the PCD components and ABA biosynthesis enzymes implies the activation of host cell death and hormonal signaling, which indicates resistance mediated by Cf genes. Additionally the upregulation of vesicular transport (Rab-GTPase) and PI-PLC suggests enhanced intercellular trafficking and secondary messenger signaling triggered during effector recognition (D.-Q. Xue et al., 2017) (Zhao et al., 2019).

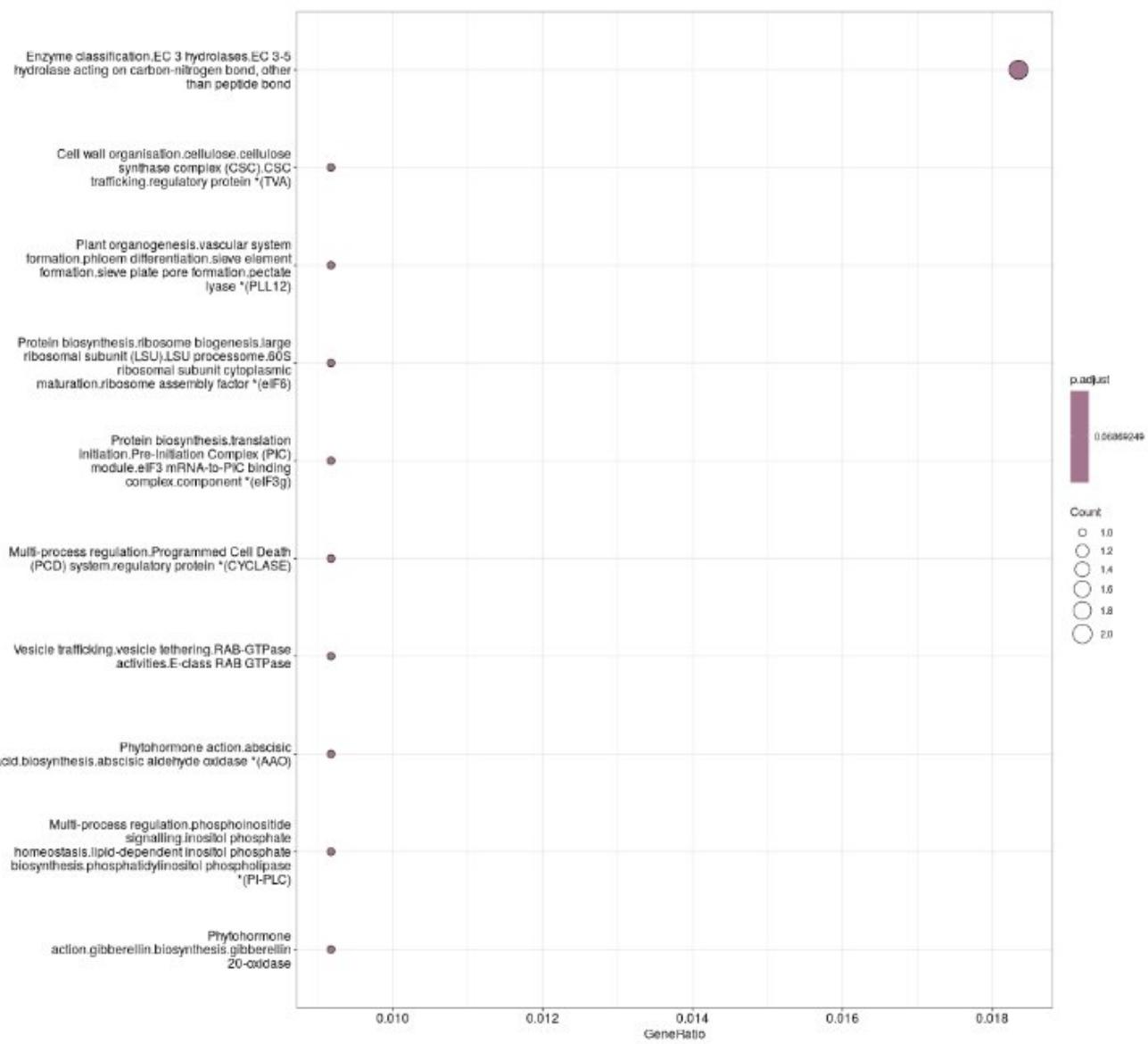


Figure 34 Functional enrichment of Palevioletred3 Module (WGCNA).

Figure 34 description: The Y axis represents the enriched terms in the module and X axis represents the gene ratio. Each dot size represents the number of genes contributing to that enriched term and color gradient is the p adjusted value.

The orangered3 module (refer Figure 35) is significantly enriched for genes involved in MADS-family transcription factors, epigenetic regulation via RNA-directed DNA methylation (RDM1), ribosome biogenesis (SSU protein eS27), fatty acid desaturation (delta-9 stearoyl – ACP desaturase, AAD), mRNA degradation (CCR4-NOT complex, NOT9), and defensin like peptides (PDF2). This module is positively correlated with FM fungal colonization. These enrichments with a gene ratio around ~0.02 and p adjusted value of ~ 0.0325 highlights a multifaceted network activated during FM fungal symbiosis. Transcriptional control (MADS-box TF + epigenetic modulators) suggests that FM colonization triggers transcriptional reprogramming, possibly to modulate root development and symbiotic interactions. Indeed, MADS-box genes have shown to participate in root nodule and mycorrhizal symbiosis signaling (Ayra et al., 2021).



Figure 35 Functional enrichment of orange3 Module (WGCNA).

Figure 35 description: The Y axis represents the enriched terms in the module and X axis represents the gene ratio. Each dot size represents the number of genes contributing to that enriched term and color gradient is the p adjusted value.

The bisque4 module (refer Figure 36), significantly upregulated under stress, is enriched in genes involved in lipid droplet-associated proteins (oleosin type structural proteins), soluble transport channels (TIP aquaporins), and sulfur transferase activity. The highest gene ratio (~ 0.03) is for the oleosin proteins, indicating a substantial induction of lipid droplet formation. Lipid droplets (LDs), known to accumulate under abiotic stresses like drought and salinity, act as reservoirs of neutral lipids and help stabilize cellular structures during stress (Huang, 2018) (Liang et al., 2020).

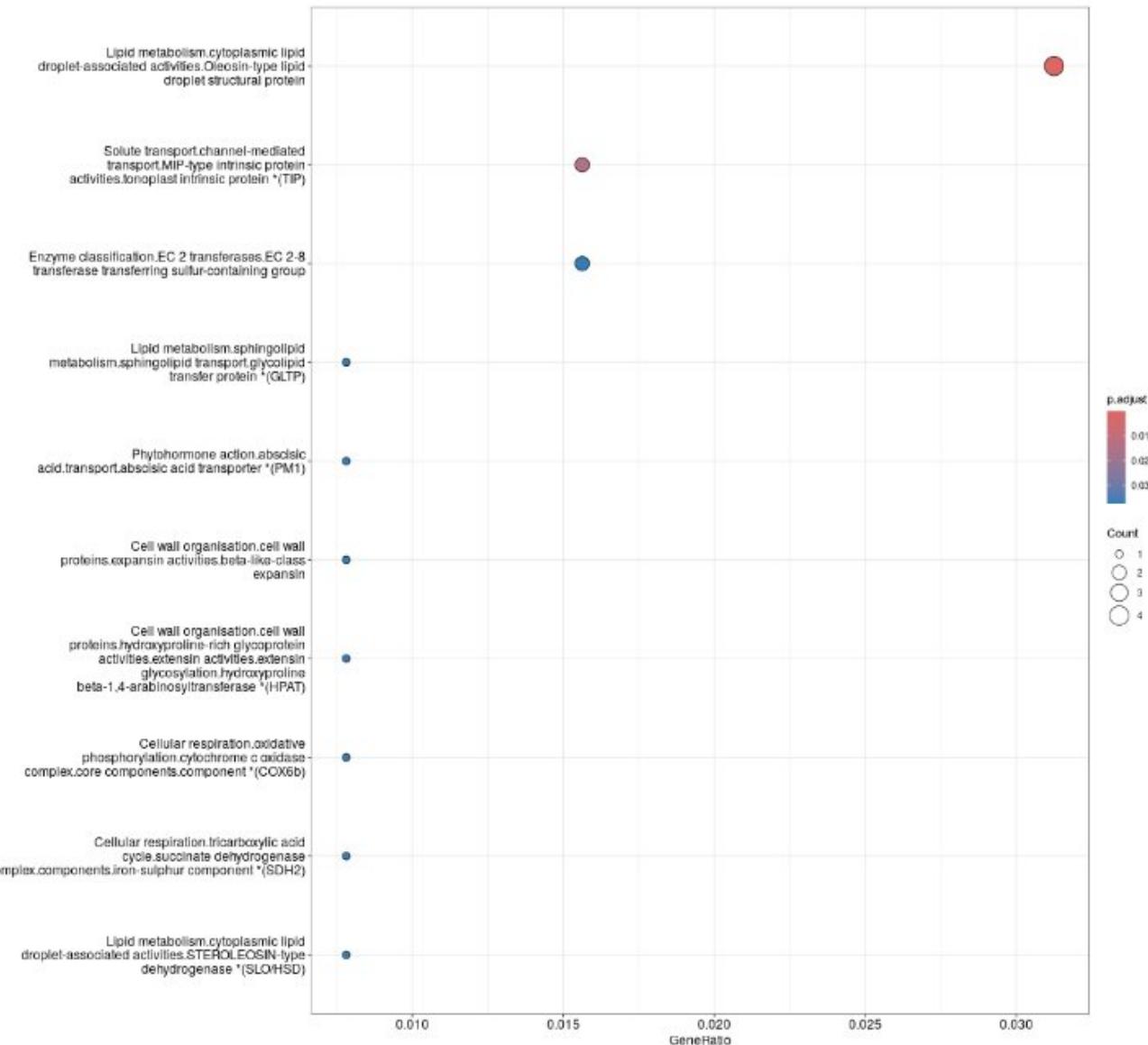


Figure 36 Functional enrichment of Bisque4 Module (WGCNA).

Figure 36 description: The Y axis represents the enriched terms in the module and X axis represents the gene ratio. Each dot size represents the number of genes contributing to that enriched term and color gradient is the p adjusted value.

The turquoise module is positively correlated with MI, a root nematode, which reveals strong transcriptional activation of genes related oxidative stress response particularly EC1 oxidoreductases, which are nematode induced reactive oxygen species (ROS) signaling (refer Figure 37). In tomato, the MI infection leads to a biphasic ROS burst that facilitates both defense and nematode feeding site formation. Solute transporters (DMT)which modulate ion and water flux during gall formation which is critical for the nematode induced vascular remodeling. TIP-type aquaporins and DMT transporters are upregulated during MI infection in tomatoes contributing to altered osmotic homeostasis (Shukla et al., 2017).

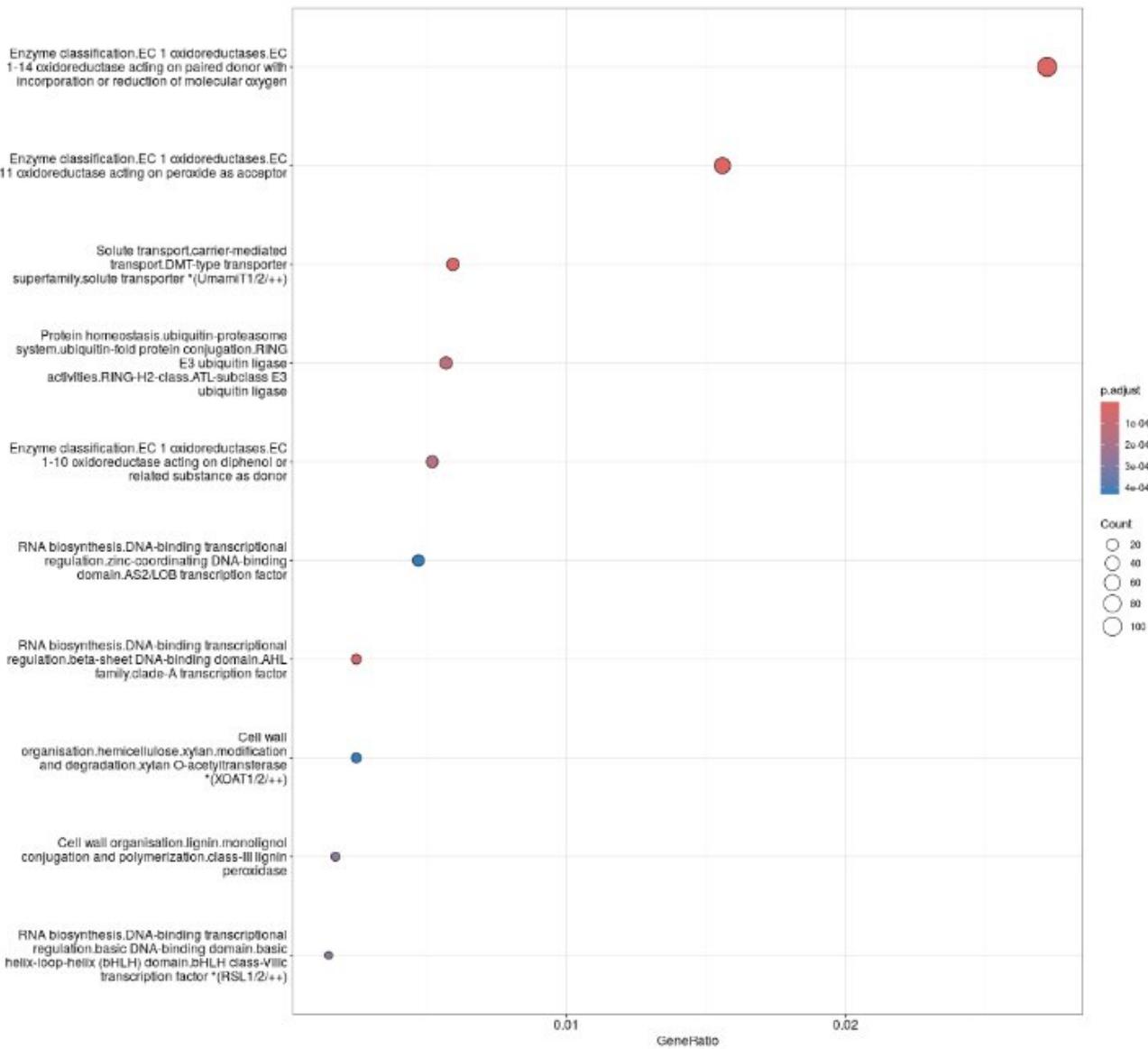


Figure 37 Functional enrichment of Turquoise Module (WGCNA).

Figure 37 description: The Y axis represents the enriched terms in the module and X axis represents the gene ratio. Each dot size represents the number of genes contributing to that enriched term and color gradient is the p adjusted value.

c. Regulatory modules

i. Overview of all the plots generated through GENIE3

GENIE3 was applied to construct a gene regulatory network using the variance stabilized expression data, identical to the input used for WGCNA. A curated list of transcription factors was supplied as regulators to infer directed regulatory links. The resulting output matrix was transformed into an edge list that captures potential regulator target gene interactions. Louvain clustering was then performed on the inferred network to identify modules of co-regulated genes. A resolution parameter of 2.1 was selected to obtain 43 distinct modules closely aligning with the 39 modules derived from WGCNA. To assess the similarity between GENIE 3 derived louvain modules and WGCNA modules, cosine similarity was calculated (refer Equation 2) between

binary gene by module membership matrices. This resulted in a similarity heatmap visualizing the degree of overlap between modules across the two methods (refer Figure 38). For downstream functional enrichment, gene sets were defined by the intersection of genes present in both WGCNA and GENIE 3 modules, while background genes were taken from the corresponding WGCNA modules.

ii. Cosine similarity

Louvain clustering was performed on GENIE3-inferred gene regulatory network to identify modules connected regions within a graph. Using a resolution parameter of 2.1, a total of 43 louvain modules were detected. To access the similarity or the overlap of genes between louvain and WGCNA modules, cosine similarity was calculated using binary gene-by module membership matrices. This produced matrix of similarity scores where each cell reflects the degree of gene membership overlap between a WGCNA module and a louvain cluster. To facilitate the visual interpretation, the cosine similarity heatmap was ordered by module size, with larger WGCNA modules positioned towards the top and the larger louvain cluster towards the left. The resulting heatmap (refer Figure 38) visualizes the degree of gene membership overlap between WGCNA modules (represented on the Y axis) and the louvain clusters (represented on the X axis), with deeper blue tones indicating higher similarity. This comparison revealed both overlapping and distinct module compositions, highlighting how co-expression and regulatory network approaches capture complimentary aspects of gene organization. This arrangement made it easier to identify patterns of overlap between dominant gene groupings. The highest observed cosine similarity was 0.12, between the WGCNA module “black” and “louvain_5”, suggesting a strong agreement in gene membership between these two groupings. A threshold of 0.03 was applied to identify louvain clusters moderately overlapped with each WGCNA modules, and gene IDs from these clusters were extracted for downstream functional enrichment. Overall, the comparison between Louvain and WGCNA modules using cosine similarity revealed both shared and distinct gene groupings, highlighting complementary insights into gene regulatory organization.

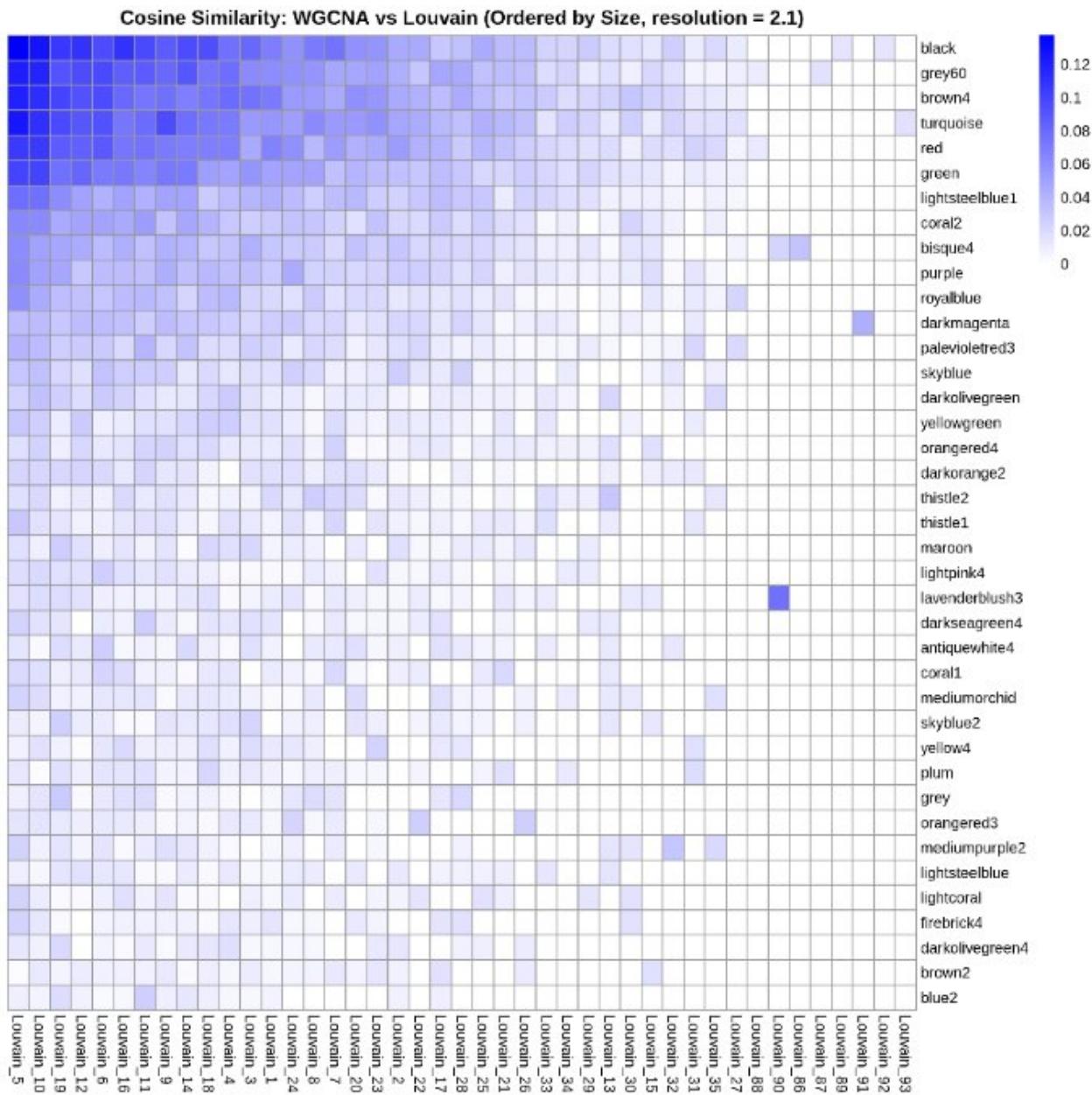


Figure 38 Cosine similarity computed between WGCNA modules and GENIE3-louvain clusters.

Figure 38 description: The X axis represents the louvain clusters and Y axis represents the WGCNA modules. The blue shade represents the cosine similarity intensity – darker the shade indicates higher similarity and lighter shade indicates low similarity.

iii. Functional enrichment of GENIE3

Functional enrichment analysis was conducted using the clusterProfiler package with gene annotations obtained from Mercator4. For each WGCNA module, genes that are overlapped with corresponding Louvain clusters based on cosine similarity were used as the genes of interest, while the full set of annotated genes within the same WGCNA module served as the background genes.

The brown4 module is strongly associated with AM interaction (refer Figure 27), functional enrichment of the intersecting gene set revealed pronounced involvement in photosynthetic processes. Enriched terms include photosystem I and II reaction center proteins, such as PsbA,

PsbD and CP43/PsbC, as well as chloroplast protein translocation and phototransferase activity (refer Figure 39). Many studies of tomato colonized with AM shows significant increase in chlorophyll content and photosynthetic activity and notably enhanced gas exchange parameters like CO₂ assimilation and stomatal conductance (Zouari et al., 2014)(Ullah et al., 2024). The elevated expression of these photosystems genes suggests that AM symbiosis may be reinforce chloroplast function and light-harvesting efficiency in tomato leaves. Additionally, several “not assigned/annotated” genes point to potentially novel components of the AM-mediated photosynthetic regulation, representing promising targets for future functional validation.

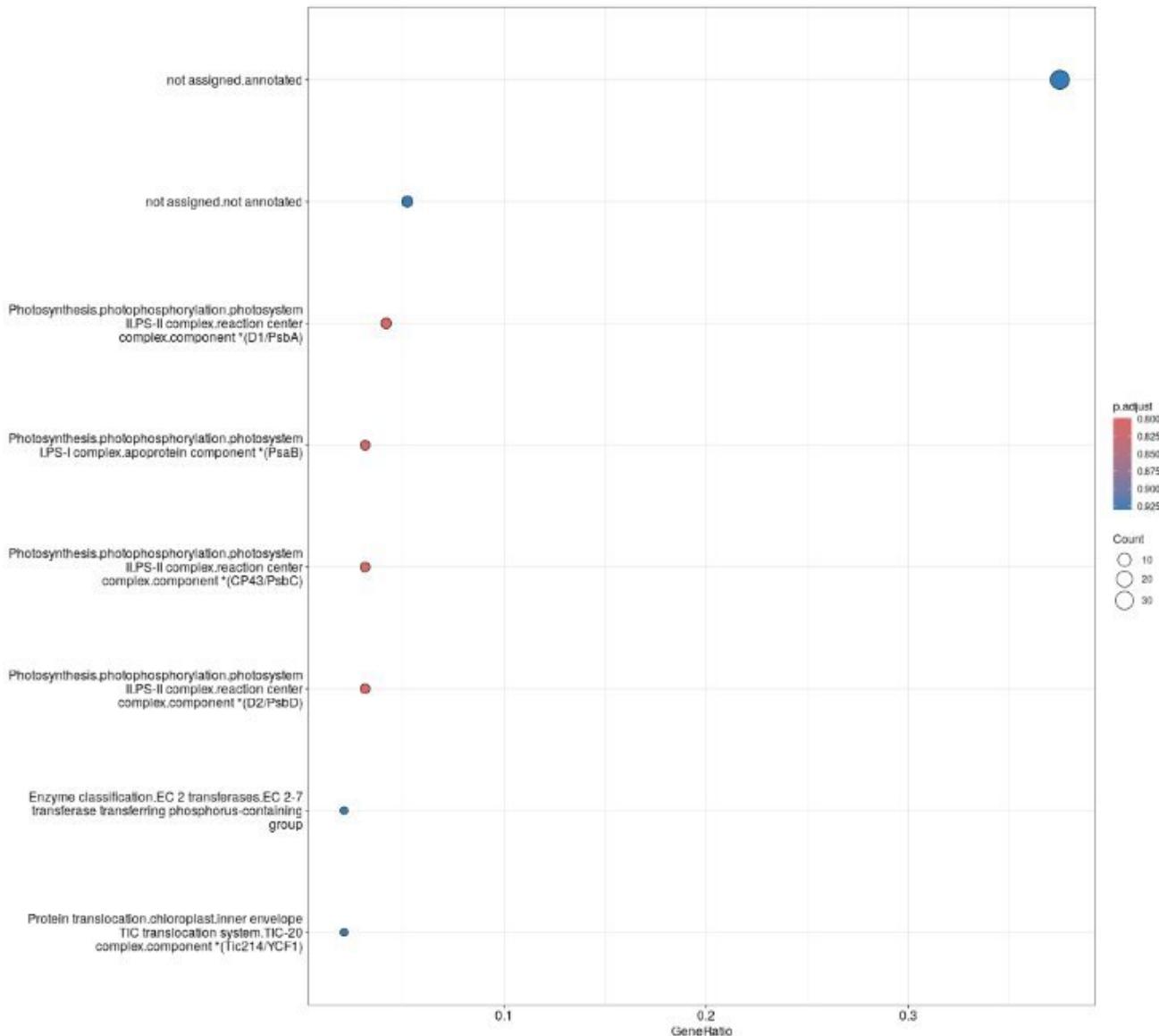


Figure 39 Functional enrichment of brown4 module (GENIE3).

Figure 39 description: The Y axis represents the enriched terms in the module and X axis represents the gene ratio. Each dot size represents the number of genes contributing to that enriched term and color gradient is the p adjusted value.

The orange/red module showed strong eigengene correlation with FM colonization in tomato, subjected to functional enrichment analysis. The analysis identified only one annotated category “not assigned/annotated”, with an adjusted p value of 0.74 suggesting a lack of significant

enrichment or the presence of poorly characterized genes in the subset (refer Figure 40). Therefore, the orangered3 module may represent a reservoir of novel or condition specific genes involved in early or late stages of FM colonization, underscoring the need for deeper annotation and targeted functional study.

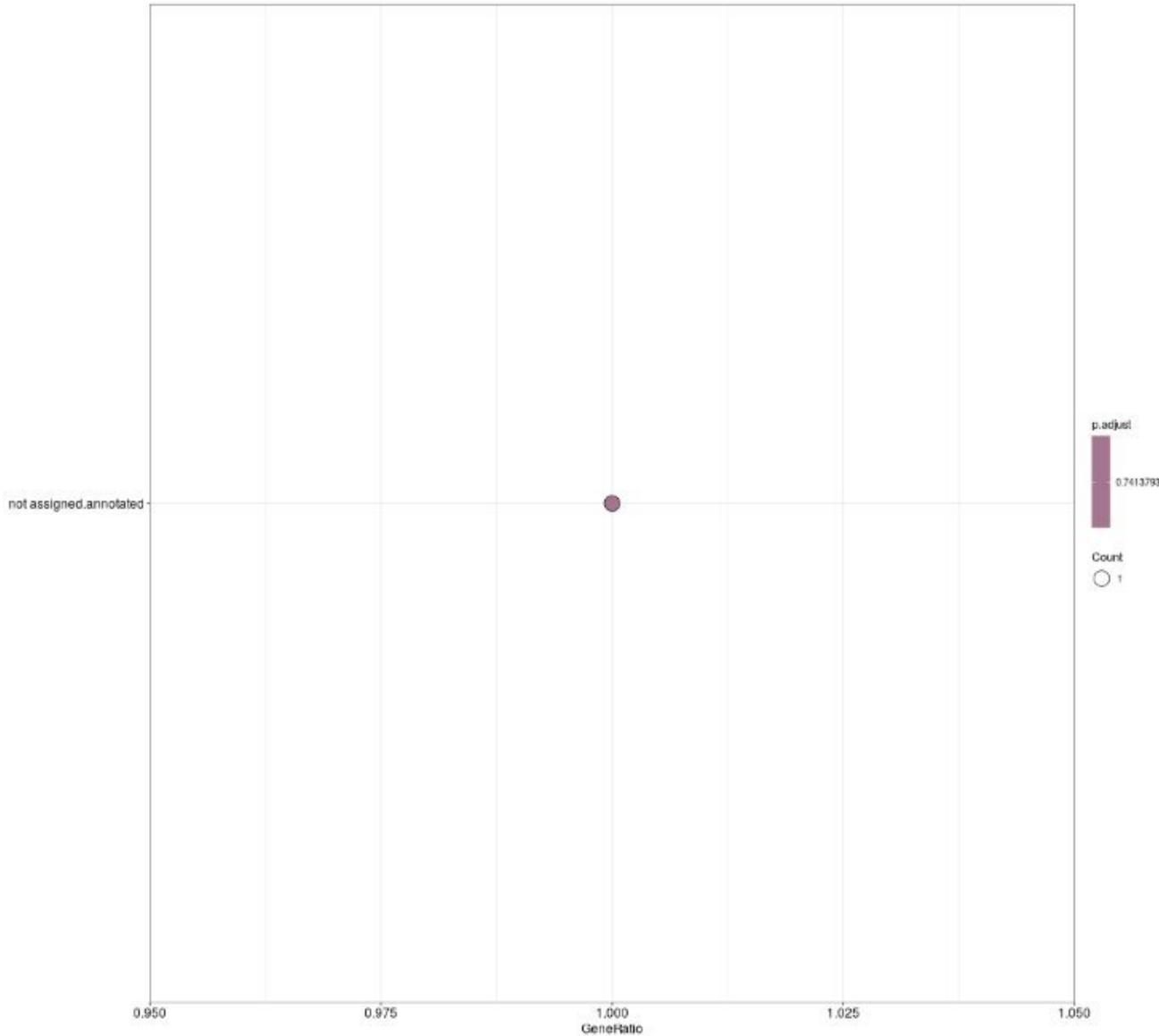


Figure 40 Functional enrichment of orangered3 module (GENIE3).

Figure 40 description: The Y axis represents the enriched terms in the module and X axis represents the gene ratio. Each dot size represents the number of genes contributing to that enriched term and color gradient is the p adjusted value.

The Lightsteelblue module demonstrated a strong correlation with CMV which was explored through functional enrichment of genes. Surprisingly, the intersecting gene set only mapped to broad “not assigned” or “not annotated” categories with an adjusted p-value of ~0.96 which is statistically relevant revealing a complete absence of enriched functional terms (refer Figure 41).

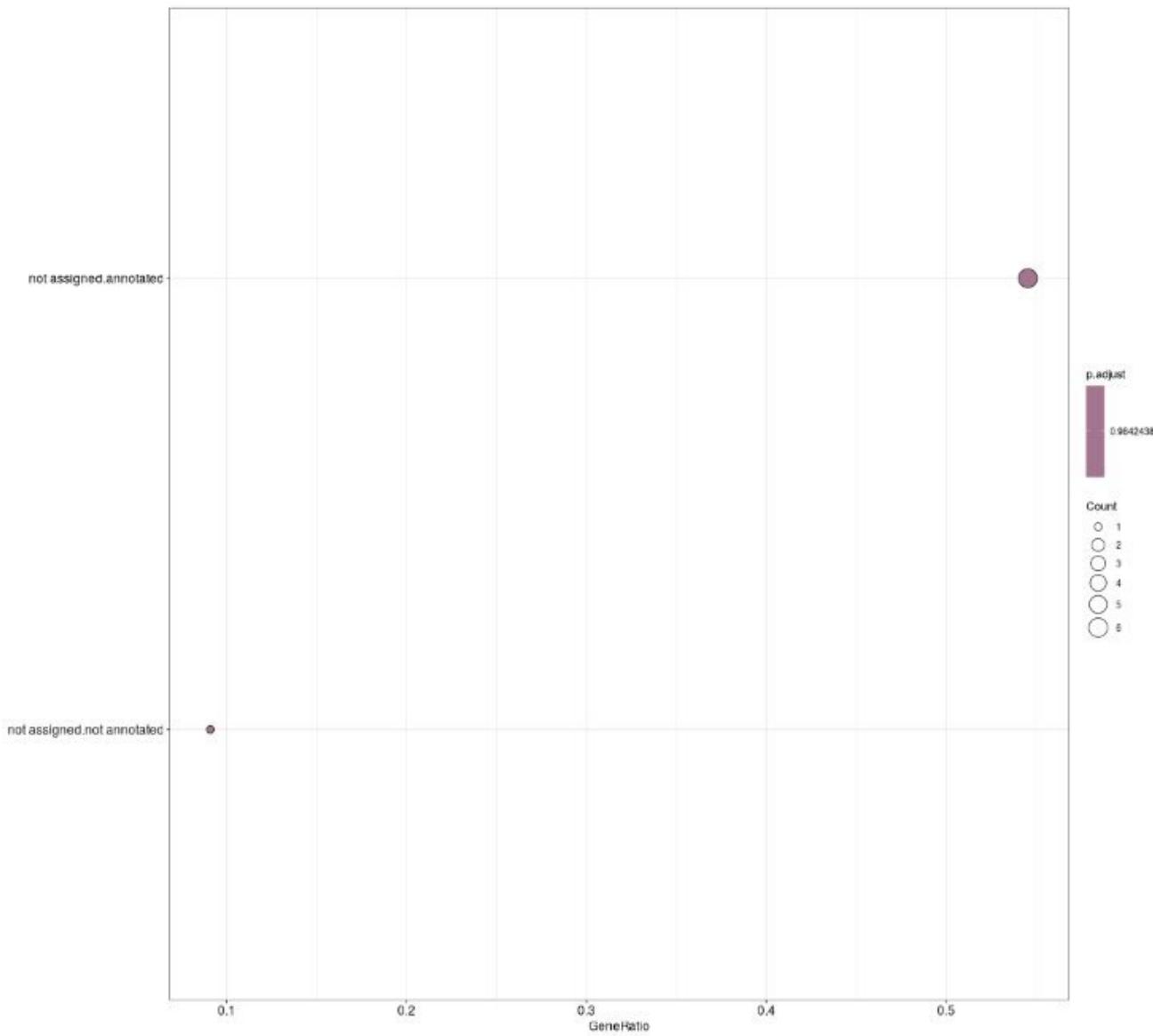


Figure 41 Functional enrichment of Lightsteelblue module (GENIE3).

Figure 41 description: The Y axis represents the enriched terms in the module and X axis represents the gene ratio. Each dot size represents the number of genes contributing to that enriched term and color gradient is the p adjusted value.

The turquoise module showed highest correlation with MI infection in tomato, which was analyzed for functional enrichment. The results revealed significant representation of genes involved in oxidoreductase activity, including EC 1.114 and EC 1.11 subclasses, as well as hydrolase activities acting on ester bond modification (refer Figure 42). Additional terms include ubiquitin-proteosome system components, transcription factors (e.g., AS2/LOB and NAC family), and pectin modification enzymes like pectin methylesterase. These enrichments likely reflect host attempts to counter nematode-induced damage and facilitate localized defense, while altered transcription factors and proteasome components may modulate downstream immune and developmental signaling (Shukla et al., 2018).

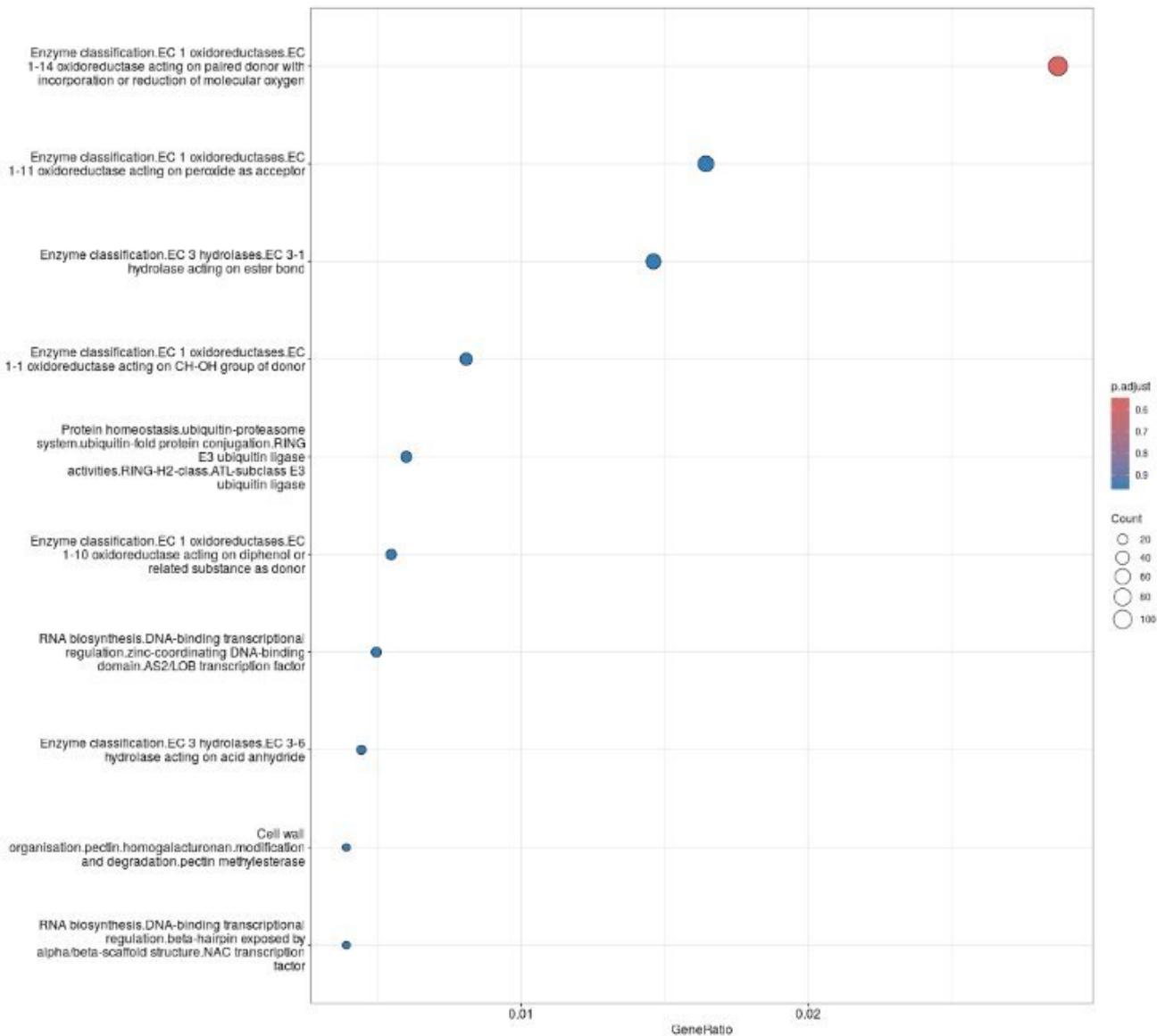


Figure 42 Functional enrichment of turquoise module (GENIE3).

Figure 42 description: The Y axis represents the enriched terms in the module and X axis represents the gene ratio. Each dot size represents the number of genes contributing to that enriched term and color gradient is the p adjusted value.

The red module demonstrated positive correlation with wide range of microbes such as AM, MI, SI, FM, CF and CMV, while showing negative correlation with BC, D, ON and all wild-type samples (refer table 1 for microbes). The functional enrichment of genes revealed terms linked to enzyme activity (oxidoreductases, transferases, hydrolases), cell cycle regulation (e.g CYCB) and kinase signaling, although adjusted p-value remained above 0.90 (refer Figure 43), indicating low statistical enrichment. Notably, a recent transcriptomic study in tomato showed that shared and unique transcriptional responses are triggered by both beneficial and pathogenic microbes, as well as abiotic stress, particularly cell wall compounds, signaling and hormones (Amoroso et al., 2023).

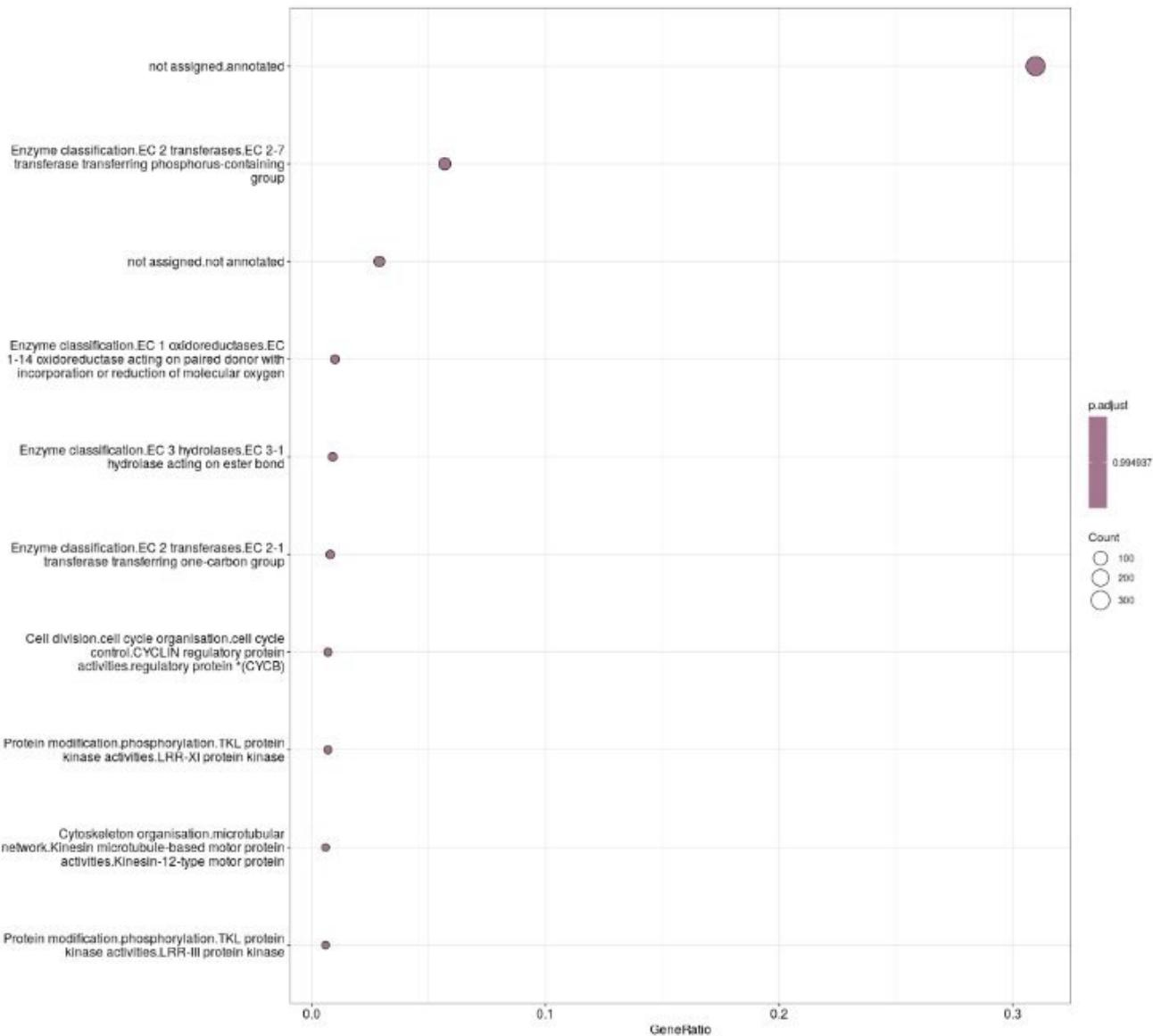


Figure 43 Functional enrichment of red module (GENIE3).

Figure 43 description: The Y axis represents the enriched terms in the module and X axis represents the gene ratio. Each dot size represents the number of genes contributing to that enriched term and color gradient is the p adjusted value.

The green module, which exhibited a complex pattern of correlations, positively correlated with AM, SI, CF, MI, ON and D and negatively correlated with FM, CMV, BC and wild type (refer Table 1 for microbes), was functionally enriched for multiple hydrolases and transferase activities (refer Figure 44). Notable categories included EC 3.1 and EC3.4 hydrolases, EC 2.4 glycosyltransferases, and ubiquitin-related ligase activity linked to hormone signaling, such as XERICO, known to modulate ABA and drought responses. These functions are consistent with known defense and stress associated transcriptional pathways (Dong et al., 2023). Additionally, several “not assigned not annotated” genes point to potentially novel components of the interaction with different abiotic and biotic stress, which needs to be studied further.

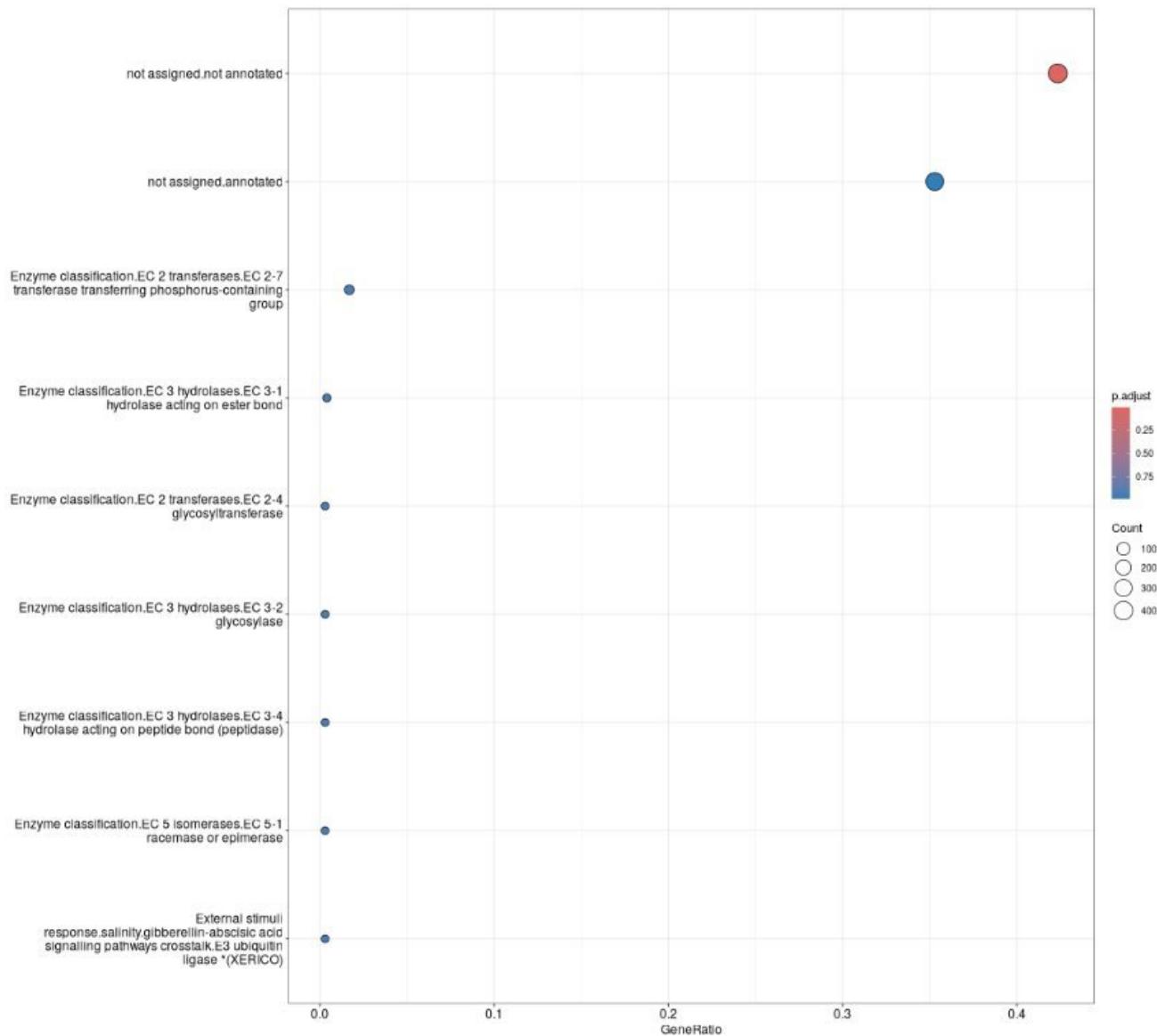


Figure 44 Functional enrichment of green module (GENIE3).

Figure 44: The Y axis represents the enriched terms in the module and X axis represents the gene ratio. Each dot represents the number of genes contributing to that enriched term and color gradient is the p adjusted value.

6. Discussion

This study applied computational approach to uncover gene co-expression and regulatory network in *solanum lycopersicum* during pathogenic and beneficial conditions. These modules that were identified through a data-driven approach involving 73 RNA-seq samples covering a range of microbial interactions, including fungal, viral, nematode pathogens as well as mutualistic fungi and abiotic stress condition such as drought. Using Weighted gene co-expression network analysis (WGCNA) and Gene network inference with ensemble of trees (GENIE3), distinct gene modules were identified that reflects biological responses which are unique and shared between pathogenic and beneficial interaction. Each module was assessed for its functional relevance using Mercator4 annotations and clusterProfiler enrichment, enabling the identification of enriched biological processes. Based on the aim of the study we were able to deduce and focus on the five most biologically informative modules – Brown4 for beneficial fungal interaction, Lightsteelblue for viral pathogenic interaction, Turquoise for nematode interaction, Red and Green module for shared interaction between pathogenic and mutualistic interactions.

a. Limitations and Future work

i. Data variability and experimental constraints

This study heavily relied on publicly available RNA-seq datasets, which, while valuable, introduce several inherent limitations. These datasets vary widely in terms of experimental design, tissue specificity, developmental stage, library preparation methods and environmental growth conditions. Although normalization and quality control procedures (such as DESeq2 normalization and variance stabilizing transformation) were applied to reduce batch effects and standardize gene expression estimates, complete harmonization of biological and technical variability was not possible. Consequently, some genes expression signals may reflect technical variation rather than true biological variations.

ii. Limitations of WGCNA

While WGCNA is a powerful tool for identifying co-expressed gene modules, it has intrinsic limitations. WGCNA assumes the input data has been normalized and properly pre-processed. The results of WGCNA can be biased or invalid when dealt with technical artifacts or tissue contaminations. Hence a balanced design is critical, for example similar number of replicates per condition, controlled growth conditions. Since this study relies on data from NCBI's SRA, there are mismatches in the number of replicates per sample and also varying growth conditions because of which biases could have been introduced (Ovens et al., 2021)(Langfelder & Horvath, 2008).

High quality RNA seq data with sufficient depth is needed for reliable co-expression networks. To achieve comparable co-expression performance, one needs to have > 20 samples with > 10 million reads each (Ballouz et al., 2015). Low read depth or noisy expression can also inflate spurious correlations. Outlier samples or genes can also distort the network, so WGCNA workflow often included filtering of outliers (e.g. using goodSamplesGenes function) to improve

data quality. For this study goodSamplesGenes function was computed and 2085 genes were removed from the total 34075 genes (Langfelder & Horvath, 2008).

This package is limited to undirected network essentially an adjacency matrix of gene-gene correlations (Langfelder & Horvath, 2008). This means it does not tell which gene might regulate another, or through what mechanism. In plant immune responses, for example, a TF and a defense gene might fall in the same module, but WGCNA alone won't reveal that the TF is the upstream of the defense gene. This lack of directionality is a built in limitation of co-expression network, making them useful for hypothesis generation but not sufficient for confirming regulatory pathways (Ovens et al., 2021)(Langfelder & Horvath, 2008).

iii. Limitations of GENIE3

The GENIE3 algorithm solves a regression problem for each gene (as a target) against all potential regulators, typically using random forest. Hence GENIE3's run time is far greater than any other methods such as LASSO or RidgeCO, where a study (Hillerton et al., 2022) stated that GENIE3 required 6 hours for 500 genes dataset and LSCON ran that in few seconds. Parallel computing can alleviate this to some extent by distributing the workload (Hu et al., 2020). Nonetheless, the runtime and memory demand are significant disadvantages when working with big gene expression matrices. GENIE3 also does not inherently account for the time-series or dynamic expression changes. As a result it can perform poorly on datasets with time course compared to other methods designed for handling time-series data (Hu et al., 2020). A dynamic extension (dynGENIE3) was later proposed to handle time-course data (Huynh-Thu et al., 2010).

During a tomato pathogenic attack, immune related genes could appear co-expressed simply because immune responsive cells increase in proportion, this could incorrectly infer regulatory links among those genes in GENIE3. A study (Kashima et al., 2021) cautioned that a GRN from bulk data might "merely reflect the changes in cell populations but not the interconnections of genes" and this must be considered when interpreting bulk RNA-inferred networks.

iv. Louvain clustering

After applying GENIE3, louvain clustering is also used to detect clusters (i.e., communities) of co-regulated genes. While this computationally efficient and widely used it also has limitations. Louvain may identify communities that are internally disconnected. It was demonstrated by (Traag et al., 2019) that Louvain often yields communities that are internally "badly connected" and in some cases even disconnected, meaning the algorithm grouped together genes that share few direct links. Up to 16% of the louvain communities were disconnected in certain benchmarks, a defect later addressed by the Leiden algorithm. In summary, Louvain's greedy merges can occasionally lump together disparate network regions, undermining the interpretability of modules.

v. Limitations of using cosine similarity

While comparing the gene modules obtained between WGCNA versus those derived from GENIE3's Louvain algorithm, the use of cosine similarity on binary module membership

matrices can introduce its own biases. Cosine similarity measures the angle between two module membership vectors, and while convenient, it has important limitations in assessing cluster overlap. One issue is its sensitivity to module size imbalance. A small WGCNA module completely nested inside a much larger GENIE3 module could yield a cosine similarity comparable to that of two moderately overlapping large modules, because the fractional overlap is similar. This indicates that the cosine similarity doesn't fully account for module size disparities, potentially overestimating the similarity of overlaps involving small modules and underestimating the significance of large absolute overlaps. While cosine similarity measure was used to compare between Louvain and WGCNA it cannot determine the significance of overlap unlike hypergeometric test (Fuxman Bass et al., 2013) (Liu et al., 2015).

vi. Advantage of implementing WGCNA and GENIE3

In this study both WGCNA and GENIE3 are independently applied for the same VST dataset to explore gene network structure in tomato during interaction with pathogenic and mutualistic microbes. WGCNA was used to identify co-regulated modules (i.e., group of genes with highly correlated expression profiles), which provides insights into coordinated functional responses to each microbe. In parallel, GENIE3 was used to infer directed gene regulatory interactions, which could capture potential causal relationships among genes based on decision tree. Although these methods were not used in an integrated pipeline, their parallel application revealed distinct layers of biological insight.

The advantage of WGCNA was its ability to reduce dimensionality by grouping genes into modules with highly correlated expression profiles. This approach highlights coordinated gene groups, which often correspond to biological pathways or functional units such as defense, secondary metabolism, or symbiosis-related responses. These co-expression modules provide insights about each microbe-specific condition, facilitating downstream analysis such as enrichment tests or trait associations.

On the other hand, GENIE3 infers directed regulatory links based on the predictive influence of one gene over another using tree-based regression models. This yields a ranked list of potential regulators, which can suggest candidate TFs that may control specific sets of genes. However, GENIE3 alone doesn't organize genes into condition specific modules, it lacks the trait-association framework that WGCNA provides. As a result, while GENIE3 adds directionality and regulators structure to the expression data, it is WGCNA that facilitates the biological grouping of genes and their linkage to specific microbial treatments.

In summary, implementing both WGCNA and GENIE3 on the same dataset provided two independent views of the transcriptome: WGCNA provided correlation based modular structure while GENIE3 introduced directionality and regulatory inference. This layered approach strengthened the biological interpretation of network patterns and helped identify candidate regulators governing key microbial interaction responses in tomato.

b. Interpretation of key findings

i. Biological interpretation of enriched pathways

Among all the WGCNA derived modules, the MEyellowgreen module stood out as being specifically upregulated in response to biotrophic interactions, particularly *Oidium neolycopersici* (ON). It exhibited the strongest positive correlation with the biotrophic interaction type ($r = 0.4$, $p = 0.001$) (refer Figure 14) and negatively correlated with all others. Functional enrichment of this module revealed overrepresentation of the S-glutathionylation/glutaredoxin activity, suggesting that redox based post translational regulation is critical mechanism in the tomato response to biotrophic fungi.

In contrast, MEbrown4 and MEorangered3 modules were positively correlated with symbiotic interaction and fungal partners such as AM and FM respectively (refer Figure 14). These modules were enriched for photosynthesis related pathways, chloroplast import and nutrient exchange, consistent with metabolic reprogramming in mutualistic symbiosis. The MERed and MEGreen modules appeared to represent generalized stress-responsive modules, with diverse gene content and mixed associations to both biotic and abiotic treatments, including necrotrophic pathogen and drought.

Despite the strong correlation of yellowgreen module with biotrophic response in WGCNA, functional enrichment did not yield refined results, as the genes within this module lacked support from the GENIE3 regulatory network. This could indicate a lack of regulatory structure overlap and implies that yellowgreen's co-expression signature maybe governed by non-transcriptional mechanisms, or transcriptional factors not captured by the curated TF list used for GENIE3.

ii. Role of TF found in regulatory modules

The integration of WGCNA and GENIE3 revealed distinct regulatory landscapes within specific gene modules. While WGCNA grouped co-expressed genes under varying microbial conditions, GENIE3 allowed the inference of putative regulatory links, enabling the identification of TF potentially controlling the expression of module-specific genes. This brings us to the second part of aim for this study which is “How many genes in the co-regulated network of pathogen have a gene regulatory network?”

The brown4 module's GENIE3 analysis supported the WGCNA finding. Louvain clusters overlapping with this module contained TF linked to photosynthesis and nutrient transport, such as bZIPs and NAC domain proteins. This suggests that transcriptional regulation is an important part of the metabolic changes that happen during symbiosis, especially in response to AM fungi.

In the case of the Lightsteelblue module, which was specific to viral (CMV), few transcription factors were identified in the GENIE3 clusters. This could mean that the genes in this module are regulated post-transcriptionally or by non-canonical mechanism, possibly due to virus driven changes in translation and protein degradation pathways.

For the turquoise module, which was associated with nematode infection, GENIE3 identified TF such as WRKY and AP2/ERFs. These are known to regulate oxidative stress responses and cell wall modification, supporting the idea that this module is involved in defense and structural changes in root tissues during nematode invasion.

The green and red modules were associated with multiple conditions, including both pathogenic and beneficial interactions. GENIE3 identified general stress related TF in these modules, including MYBs, bZIPs, and TCP family proteins. These regulators are often involved in hormone signaling, cell division and growth defense balance.

7. Conclusion

This study presents an integrative network based transcriptomic analysis to unravel the unique and shared molecular responses of tomato during interaction with pathogenic and mutualistic microbes. Using RNA-seq datasets across eight distinct microbe-tomato interaction condition along with one sample for abiotic stress condition, a robust and replicable computational workflow that combined WGCNA and GENIE3 was implemented.

By analyzing modules across pathogenic, mutualistic, and abiotic conditions, biologically meaningful gene clusters associated with nutrient exchange, redox regulation, hormone signaling, and defense were identified. While several module enrichments returned terms labeled as “not assigned/annotated”, this may indicate the presence of novel or condition-specific genes whose functions are not yet characterized. These gene groups need to be further investigated through wet-lab validation to uncover their potential roles in microbial interaction and stress responses.

This network-based framework advances our understanding of how tomato responds to different microbes by turning external signals into specific biological responses. It shows the value of using both co-expression and regulatory network analysis to study complex plant-microbe interactions. This method is also flexible and can be used with other crops or conditions to find important genes and pathways that help plants adapt.

8. Acknowledgements

I would like to express my special gratitude to my supervisor Prof. Nadia Kamal, Dr. Amit Fenn and everyone at the chair of Computational Plant Biology for guiding me through this project and giving me the opportunity to work on such an interesting topic.

9. References

- 1) Ahmed, N., Li, J., Li, Y., Deng, L., Deng, L., Chachar, M., Chachar, Z., Chachar, S., Hayat, F., Raza, A., Umrani, J. H., Gong, L., & Tu, P. (2025). Symbiotic synergy: How Arbuscular Mycorrhizal Fungi enhance nutrient uptake, stress tolerance, and soil health through molecular mechanisms and hormonal regulation. *IMA Fungus*, 16, e144989. <https://doi.org/10.3897/imafungus.16.144989>
- 2) Amoroso, C. G., D'Esposito, D., Aiese Cigliano, R., & Ercolano, M. R. (2023). Comparison of Tomato Transcriptomic Profiles Reveals Overlapping Patterns in Abiotic and Biotic Stress Responses. *International Journal of Molecular Sciences*, 24(4), 4061. <https://doi.org/10.3390/ijms24044061>
- 3) Ayra, L., Reyero-Saavedra, M. del R., Isidra-Arellano, M. C., Lozano, L., Ramírez, M., Leija, A., Fuentes, S.-I., Girard, L., Valdés-López, O., & Hernández, G. (2021). Control of the Rhizobia Nitrogen-Fixing Symbiosis by Common Bean MADS-Domain/AGL Transcription Factors. *Frontiers in Plant Science*, 12. <https://doi.org/10.3389/fpls.2021.679463>
- 4) Badmi, R., Tengs, T., Brurberg, M. B., Elameen, A., Zhang, Y., Haugland, L. K., Fossdal, C. G., Hytönen, T., Krokene, P., & Thorstensen, T. (2022). Transcriptional profiling of defense responses to Botrytis cinerea infection in leaves of Fragaria vesca plants soil-drenched with β-aminobutyric acid. *Frontiers in Plant Science*, 13. <https://doi.org/10.3389/fpls.2022.1025422>
- 5) Bai, Y., & Lindhout, P. (2007). Domestication and Breeding of Tomatoes: What have We Gained and What Can We Gain in the Future? *Annals of Botany*, 100(5), 1085–1094. <https://doi.org/10.1093/aob/mcm150>
- 6) Ballouz, S., Verleyen, W., & Gillis, J. (2015). Guidance for RNA-seq co-expression network construction and analysis: Safety in numbers. *Bioinformatics*, 31(13), 2123–2130. <https://doi.org/10.1093/bioinformatics/btv118>
- 7) Bar-Joseph, Z., Gerber, G. K., Lee, T. I., Rinaldi, N. J., Yoo, J. Y., Robert, F., Gordon, D. B., Fraenkel, E., Jaakkola, T. S., Young, R. A., & Gifford, D. K. (2003). Computational discovery of gene modules and regulatory networks. *Nature Biotechnology*, 21(11), 1337–1342. <https://doi.org/10.1038/nbt890>
- 8) Blondel, V., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast Unfolding of Communities in Large Networks. *Journal of Statistical Mechanics Theory and Experiment*, 2008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- 9) Blüthgen, N., Brand, K., Cajavec Bernard, B., Swat, M., Herzel, H., & Beule, D. (2005). Biological profiling of Gene Groups utilizing Gene Ontology – A Statistical Framework. *Genome Informatics. International Conference on Genome Informatics*, 16, 106–115. <https://doi.org/10.11234/gi1990.16.106>
- 10) Bolger, M., Schwacke, R., & Usadel, B. (2021). MapMan Visualization of RNA-Seq Data Using Mercator4 Functional Annotations. In D. Dobnik, K. Gruden, Ž. Ramšak, & A. Coll (Eds.), *Solanum tuberosum* (Vol. 2354, pp. 195–212). Springer US. https://doi.org/10.1007/978-1-0716-1609-3_9
- 11) Borges, Á. V., Saraiva, R. M., & Maffia, L. A. (2014). Key factors to inoculate Botrytis cinerea in tomato plants. *Summa Phytopathologica*, 40, 221–225. <https://doi.org/10.1590/0100-5405/1929>
- 12) Broido, A. D., & Clauset, A. (2019). Scale-free networks are rare. *Nature Communications*, 10(1), 1017. <https://doi.org/10.1038/s41467-019-08746-5>

- 13) Campos, M. D., Félix, M. do R., Patanita, M., Materatski, P., Albuquerque, A., Ribeiro, J. A., & Varanda, C. (2022). Defense Strategies: The Role of Transcription Factors in Tomato–Pathogen Interaction. *Biology*, 11(2), Article 2. <https://doi.org/10.3390/biology11020235>
- 14) Coelho, A., Horta Jung, M., Ebadzad, G., & Cravador, A. (2011). *Quercus suber – Phytophthora cinnamomi* interaction: A hypothetical molecular mechanism model. *N. Z. J. For. Sci.*, 41.
- 15) Colmán, A. A., Alves, J. L., da Silva, M., & Barreto, R. W. (2018). *Phoma destructiva* causing blight of tomato plants: A new fungal threat for tomato plantations in Brazil? *Tropical Plant Pathology*, 43(3), 257–262. <https://doi.org/10.1007/s40858-017-0200-2>
- 16) De Cal, A., Melgarejo, P., & Jimenez-Gasco, M. D. M. (2022). Editorial: Necrotrophic Fungal Plant Pathogens. *Frontiers in Plant Science*, 13. <https://doi.org/10.3389/fpls.2022.839674>
- 17) de Wit, P. J. G. M., van der Burgt, A., Ökmen, B., Stergiopoulos, I., Abd-Elsalam, K. A., Aerts, A. L., Bahkali, A. H., Beenen, H. G., Chettri, P., Cox, M. P., Datema, E., de Vries, R. P., Dhillon, B., Ganley, A. R., Griffiths, S. A., Guo, Y., Hamelin, R. C., Henrissat, B., Kabir, M. S., ... Bradshaw, R. E. (2012). The Genomes of the Fungal Plant Pathogens *Cladosporium fulvum* and *Dothistroma septosporum* Reveal Adaptation to Different Hosts and Lifestyles But Also Signatures of Common Ancestry. *PLoS Genetics*, 8(11), e1003088. <https://doi.org/10.1371/journal.pgen.1003088>
- 18) Dong, S., Ling, J., Song, L., Zhao, L., Wang, Y., & Zhao, T. (2023). Transcriptomic Profiling of Tomato Leaves Identifies Novel Transcription Factors Responding to Dehydration Stress. *International Journal of Molecular Sciences*, 24(11), Article 11. <https://doi.org/10.3390/ijms24119725>
- 19) Fuxman Bass, J. I., Diallo, A., Nelson, J., Soto, J. M., Myers, C. L., & Walhout, A. J. M. (2013). Using networks to measure similarity between genes: Association index selection. *Nature Methods*, 10(12), 1169–1176. <https://doi.org/10.1038/nmeth.2728>
- 20) Ghafouri-Fard, S., Safarzadeh, A., Taheri, M., & Jamali, E. (2023a). Identification of diagnostic biomarkers via weighted correlation network analysis in colorectal cancer using a system biology approach. *Scientific Reports*, 13(1), 13637. <https://doi.org/10.1038/s41598-023-40953-5>
- 21) Ghafouri-Fard, S., Safarzadeh, A., Taheri, M., & Jamali, E. (2023b). Identification of diagnostic biomarkers via weighted correlation network analysis in colorectal cancer using a system biology approach. *Scientific Reports*, 13(1), 13637. <https://doi.org/10.1038/s41598-023-40953-5>
- 22) Ghafouri-Fard, S., Safarzadeh, A., Taheri, M., & Jamali, E. (2023c). Identification of diagnostic biomarkers via weighted correlation network analysis in colorectal cancer using a system biology approach. *Scientific Reports*, 13(1), 13637. <https://doi.org/10.1038/s41598-023-40953-5>
- 23) González, M., Brito, N., Frías, M., & González, C. (2013). Botrytis cinerea Protein O-Mannosyltransferases Play Critical Roles in Morphogenesis, Growth, and Virulence. *PLOS ONE*, 8(6), e65924. <https://doi.org/10.1371/journal.pone.0065924>
- 24) *goodSamplesGenesMS function—RDocumentation*. (n.d.). Retrieved 18 June 2025, from <https://www.rdocumentation.org/packages/WGCNA/versions/1.72-5/topics/goodSamplesGenesMS>
- 25) Guttman, D. S., McHardy, A. C., & Schulze-Lefert, P. (2014). Microbial genome-enabled insights into plant–microorganism interactions. *Nature Reviews Genetics*, 15(12), 797–813. <https://doi.org/10.1038/nrg3748>

- 26) Hillerton, T., Seçilmiş, D., Nelander, S., & Sonnhammer, E. L. L. (2022). Fast and accurate gene regulatory network inference by normalized least squares regression. *Bioinformatics*, 38(8), 2263–2268. <https://doi.org/10.1093/bioinformatics/btac103>
- 27) Hu, X., Hu, Y., Wu, F., Leung, R. W. T., & Qin, J. (2020). Integration of single-cell multi-omics for gene regulatory network inference. *Computational and Structural Biotechnology Journal*, 18, 1925–1938. <https://doi.org/10.1016/j.csbj.2020.06.033>
- 28) Huang, A. H. C. (2018). Plant Lipid Droplets and Their Associated Proteins: Potential for Rapid Advances1[OPEN]. *Plant Physiology*, 176(3), 1894–1918. <https://doi.org/10.1104/pp.17.01677>
- 29) Huynh-Thu, V. A., & Geurts, P. (2018). dynGENIE3: Dynamical GENIE3 for the inference of gene networks from time series expression data. *Scientific Reports*, 8(1), 3384. <https://doi.org/10.1038/s41598-018-21715-0>
- 30) Huynh-Thu, V. A., Irrthum, A., Wehenkel, L., & Geurts, P. (2010). Inferring Regulatory Networks from Expression Data Using Tree-Based Methods. *PLOS ONE*, 5(9), e12776. <https://doi.org/10.1371/journal.pone.0012776>
- 31) Jiang, X., Li, Y., Li, R., Gao, Y., Liu, Z., Yang, H., Li, J., Jiang, J., Zhao, T., & Xu, X. (2022). Transcriptome Analysis of the Cf-13-Mediated Hypersensitive Response of Tomato to Cladosporium fulvum Infection. *International Journal of Molecular Sciences*, 23(9), 4844. <https://doi.org/10.3390/ijms23094844>
- 32) Jones, H., Whipps, J. M., & Gurr, S. J. (2001). The tomato powdery mildew fungus Oidium neolympersici. *Molecular Plant Pathology*, 2(6), 303–309. <https://doi.org/10.1046/j.1464-6722.2001.00084.x>
- 33) Kalinina, E., & Novichkova, M. (2021). Glutathione in Protein Redox Modulation through S-Glutathionylation and S-Nitrosylation. *Molecules*, 26(2), Article 2. <https://doi.org/10.3390/molecules26020435>
- 34) Kashima, M., Shida, Y., Yamashiro, T., Hirata, H., & Kurosaka, H. (2021). Intracellular and Intercellular Gene Regulatory Network Inference From Time-Course Individual RNA-Seq. *Frontiers in Bioinformatics*, 1. <https://doi.org/10.3389/fbinf.2021.777299>
- 35) Khaliq, A., Perveen, S., Alamer, K. H., Zia Ul Haq, M., Rafique, Z., Alsudays, I. M., Althobaiti, A. T., Saleh, M. A., Hussain, S., & Attia, H. (2022). Arbuscular Mycorrhizal Fungi Symbiosis to Enhance Plant–Soil Interaction. *Sustainability*, 14(13), Article 13. <https://doi.org/10.3390/su14137840>
- 36) Ko, D. K., & Brandizzi, F. (2023). Coexpression Network Construction and Visualization from Transcriptomes Underlying ER Stress Responses. *Methods in Molecular Biology (Clifton, N.J.)*, 2581, 385–401. https://doi.org/10.1007/978-1-0716-2784-6_27
- 37) Langfelder, P. (2018, November 25). Signed and signed hybrid: What's the difference? *Insights from a Billion Correlations*. https://peterlangfelder.com/2018/11/25/_trashed/
- 38) Langfelder, P., & Horvath, S. (2007). Eigengene networks for studying the relationships between co-expression modules. *BMC Systems Biology*, 1(1), 54. <https://doi.org/10.1186/1752-0509-1-54>
- 39) Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, 9(1), 559. <https://doi.org/10.1186/1471-2105-9-559>
- 40) Langfelder, P., & Horvath, S. (2012). Fast R Functions For Robust Correlations And Hierarchical Clustering. *Journal of Statistical Software*, 46. <https://doi.org/10.18637/jss.v046.i11>

- 41) Liang, G., Liu, J., Zhang, J., & Guo, J. (2020). Effects of Drought Stress on Photosynthetic and Physiological Parameters of Tomato. *Journal of the American Society for Horticultural Science*, 145(1), 12–17. <https://doi.org/10.21273/JASHS04725-19>
- 42) Liu, H., Liu, W., Lin, Y., Liu, T., Ma, Z., Li, M., Zhang, H.-M., Kenneth Wang, Q., & Guo, A.-Y. (2015). Scoring the correlation of genes by their shared properties using OSCal, an improved overlap quantification model. *Scientific Reports*, 5(1), 10583. <https://doi.org/10.1038/srep10583>
- 43) Lohse, M., Nagel, A., Herter, T., May, P., Schroda, M., Zrenner, R., Tohge, T., Fernie, A. R., Stitt, M., & Usadel, B. (2014). Mercator: A fast and simple web server for genome scale functional annotation of plant sequence data. *Plant, Cell & Environment*, 37(5), 1250–1258. <https://doi.org/10.1111/pce.12231>
- 44) Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 550. <https://doi.org/10.1186/s13059-014-0550-8>
- 45) Ma, Y., MacMillan, C. P., de Vries, L., Mansfield, S. D., Hao, P., Ratcliffe, J., Bacic, A., & Johnson, K. L. (2022). FLA11 and FLA12 glycoproteins fine-tune stem secondary wall properties in response to mechanical stresses. *The New Phytologist*, 233(4), 1750–1767. <https://doi.org/10.1111/nph.17898>
- 46) Mapuranga, J., Zhang, N., Zhang, L., Chang, J., & Yang, W. (2022). Infection Strategies and Pathogenicity of Biotrophic Plant Fungal Pathogens. *Frontiers in Microbiology*, 13. <https://doi.org/10.3389/fmicb.2022.799396>
- 47) Matsui, R., Ferran, B., Oh, A., Croteau, D., Shao, D., Han, J., Pimentel, D. R., & Bachschmid, M. M. (2020). Redox Regulation via Glutaredoxin-1 and Protein S-Glutathionylation. *Antioxidants & Redox Signaling*, 32(10), 677–700. <https://doi.org/10.1089/ars.2019.7963>
- 48) Meidani, C., Telioglanidis, K., Giannoutsou, E., Ntalli, N., & Adamakis, I. D. S. (2025). Meloidogyne incognita-Induced Giant Cells in Tomato and the Impact of Acetic Acid. *Plants*, 14(7), Article 7. <https://doi.org/10.3390/plants14071015>
- 49) Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., & Thomas, P. D. (2017). PANTHER version 11: Expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Research*, 45(D1), D183–D189. <https://doi.org/10.1093/nar/gkw1138>
- 50) Miozzi, L., Vaira, A. M., Brilli, F., Casarin, V., Berti, M., Ferrandino, A., Nerva, L., Accotto, G. P., & Lanfranco, L. (2020). Arbuscular Mycorrhizal Symbiosis Primes Tolerance to Cucumber Mosaic Virus in Tomato. *Viruses*, 12(6), Article 6. <https://doi.org/10.3390/v12060675>
- 51) Moustakas, M., Bayçu, G., Sperdouli, I., Eroğlu, H., & Eleftheriou, E. P. (2020). Arbuscular Mycorrhizal Symbiosis Enhances Photosynthesis in the Medicinal Herb Salvia fruticosa by Improving Photosystem II Photochemistry. *Plants*, 9(8), Article 8. <https://doi.org/10.3390/plants9080962>
- 52) Muñoz, C. Y., de Jong, A., & Kuipers, O. P. (2021). Draft Genome Sequences of a *Bacillus subtilis* Strain, a *Bacillus velezensis* Strain, a *Paenibacillus* Strain, and an *Acinetobacter baumannii* Strain, All Isolated from the Phyllosphere of *Lactuca sativa* or *Solanum lycopersicum*. *Microbiology Resource Announcements*, 10(4), 10.1128/mra.01092-20. <https://doi.org/10.1128/mra.01092-20>
- 53) Ntana, F., Johnson, S. R., Hamberger, B., Jensen, B., Jørgensen, H. J. L., & Collinge, D. B. (2022). Regulation of Tomato Specialised Metabolism after Establishment of Symbiosis with the

- Endophytic Fungus Serendipita indica. *Microorganisms*, 10(1), 194. <https://doi.org/10.3390/microorganisms10010194>
- 54) Ovens, K., Eames, B. F., & McQuillan, I. (2021). Comparative Analyses of Gene Co-expression Networks: Implementations and Applications in the Study of Evolution. *Frontiers in Genetics*, 12. <https://doi.org/10.3389/fgene.2021.695399>
- 55) Perrin, D., & Zuccon, G. (2018). Recursive module extraction using Louvain and PageRank. *F1000Research*, 7, 1286. <https://doi.org/10.12688/f1000research.15845.1>
- 56) Petralia, F., Wang, P., Yang, J., & Tu, Z. (2015). Integrative random forest for gene regulatory network inference. *Bioinformatics*, 31(12), i197–i205. <https://doi.org/10.1093/bioinformatics/btv268>
- 57) *pickSoftThreshold: Analysis of scale free topology for soft-thresholding in WGCNA: Weighted Correlation Network Analysis.* (n.d.). Retrieved 18 June 2025, from <https://rdrr.io/cran/WGCNA/man/pickSoftThreshold.html>
- 58) Plouznikoff, K., Asins, M. J., de Boulois, H. D., Carbonell, E. A., & Declerck, S. (2019). Genetic analysis of tomato root colonization by arbuscular mycorrhizal fungi. *Annals of Botany*, 124(6), 933–946. <https://doi.org/10.1093/aob/mcy240>
- 59) Qin, S., Veloso, J., Baak, M., Boogmans, B., Bosman, T., Puccetti, G., Shi-Kunne, X., Smit, S., Grant-Downton, R., Leisen, T., Hahn, M., & van Kan, J. A. L. (2023). Molecular characterization reveals no functional evidence for naturally occurring cross-kingdom RNA interference in the early stages of *Botrytis cinerea*–tomato interaction. *Molecular Plant Pathology*, 24(1), 3–15. <https://doi.org/10.1111/mpp.13269>
- 60) Real, N., Garcia-Molina, A., Stolze, S. C., Harzen, A., Nakagami, H., & Martín-Hernández, A. M. (2025). Comprehensive proteomic profiling of Cucumber mosaic virus infection: Identifying key proteins and pathways involved in resistance and susceptibility in melon. *BMC Plant Biology*, 25(1), 434. <https://doi.org/10.1186/s12870-025-06464-3>
- 61) Rick, C. M. (1980). Tomato. In *Hybridization of Crop Plants* (pp. 669–680). John Wiley & Sons, Ltd. <https://doi.org/10.2135/1980.hybridizationofcrops.c48>
- 62) Risso, D., Schwartz, K., Sherlock, G., & Dudoit, S. (2011). GC-Content Normalization for RNA-Seq Data. *BMC Bioinformatics*, 12(1), 480. <https://doi.org/10.1186/1471-2105-12-480>
- 63) Rochaix, J. (2022). Chloroplast protein import machinery and quality control. *The Febs Journal*, 289(22), 6908–6918. <https://doi.org/10.1111/febs.16464>
- 64) Rodrigues, K. M., & Rodrigues, B. F. (2015). Endomycorrhizal association of Funneliformis mosseae with transformed roots of *Linum usitatissimum*: Germination, colonization, and sporulation studies. *Mycology*, 6(1), 42–49. <https://doi.org/10.1080/21501203.2015.1024777>
- 65) Sanchez-Bel, P., Troncho, P., Gamir, J., Pozo, M. J., Camañes, G., Cerezo, M., & Flors, V. (2016). The Nitrogen Availability Interferes with Mycorrhiza-Induced Resistance against *Botrytis cinerea* in Tomato. *Frontiers in Microbiology*, 7. <https://doi.org/10.3389/fmicb.2016.01598>
- 66) Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D., & Friedman, N. (2003). Module networks: Identifying regulatory modules and their condition-specific regulators from gene expression data. *Nature Genetics*, 34(2), 166–176. <https://doi.org/10.1038/ng1165>
- 67) Shiu, S.-H., & Bleecker, A. B. (2003). Expansion of the Receptor-Like Kinase/Pelle Gene Family and Receptor-Like Proteins in *Arabidopsis*. *Plant Physiology*, 132(2), 530–543. <https://doi.org/10.1104/pp.103.021964>

- 68) Shu, K., & Yang, W. (2017). E3 Ubiquitin Ligases: Ubiquitous Actors in Plant Development and Abiotic Stress Responses. *Plant and Cell Physiology*, 58(9), 1461–1476. <https://doi.org/10.1093/pcp/pcx071>
- 69) Shukla, N., Yadav, R., Kaur, P., Rasmussen, S., Goel, S., Agarwal, M., Jagannath, A., Gupta, R., & Kumar, A. (2017). Transcriptome analysis of root-knot nematode (*Meloidogyne incognita*)-infected tomato (*Solanum lycopersicum*) roots reveals complex gene expression profiles and metabolic networks of both host and nematode during susceptible and resistance responses. *Molecular Plant Pathology*, 19(3), 615–633. <https://doi.org/10.1111/mpp.12547>
- 70) Shukla, N., Yadav, R., Kaur, P., Rasmussen, S., Goel, S., Agarwal, M., Jagannath, A., Gupta, R., & Kumar, A. (2018). Transcriptome analysis of root-knot nematode (*Meloidogyne incognita*)-infected tomato (*Solanum lycopersicum*) roots reveals complex gene expression profiles and metabolic networks of both host and nematode during susceptible and resistance responses. *Molecular Plant Pathology*, 19(3), 615–633. <https://doi.org/10.1111/mpp.12547>
- 71) Singh, M., Chauhan, A., Srivastava, D. K., & Singh, P. K. (2024). Arbuscular mycorrhizal fungi promote growth and enhance the accumulation of bioactive compounds in Tomato (*Solanum lycopersicum* L.). *Biologia Futura*, 75(2), 251–257. <https://doi.org/10.1007/s42977-024-00214-6>
- 72) Sol Genomics Network. (n.d.). Retrieved 18 June 2025, from <https://solgenomics.sgn.cornell.edu/>
- 73) Šubr, Z., Predajňa, L., Soltys, K., Bokor, B., Budis, J., & Glasa, M. (2020). Comparative Transcriptome Analysis of Two Cucumber Cultivars with Different Sensitivity to Cucumber Mosaic Virus Infection. *Pathogens*, 9, 145. <https://doi.org/10.3390/pathogens9020145>
- 74) Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43), 15545–15550. <https://doi.org/10.1073/pnas.0506580102>
- 75) Suzuki, R., Ueda, T., Wada, T., Ito, M., Ishida, T., & Sawa, S. (2021). Identification of genes involved in *Meloidogyne incognita*-induced gall formation processes in *Arabidopsis thaliana*. *Plant Biotechnology*, 38(1), 1–8. <https://doi.org/10.5511/plantbiotechnology.20.0716a>
- 76) Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to Leiden: Guaranteeing well-connected communities. *Scientific Reports*, 9, 5233. <https://doi.org/10.1038/s41598-019-41695-z>
- 77) Ullah, F., Ullah, H., Ishfaq, M., Khan, R., Gul, S. L., Gulfraz, A., Wang, C., & Zhifang, L. (2024). Genotypic variation of tomato to AMF inoculation in improving growth, nutrient uptake, yield, and photosynthetic activity. *Symbiosis*, 92(1), 111–124. <https://doi.org/10.1007/s13199-023-00961-5>
- 78) Varma, A., Savita Verma, Sudha, Sahay, N., Bütehorn, B., & Franken, P. (1999). Piriformospora indica, a Cultivable Plant-Growth-Promoting Root Endophyte. *Applied and Environmental Microbiology*, 65(6), 2741–2744. <https://doi.org/10.1128/AEM.65.6.2741-2744.1999>
- 79) Wu, P., Wu, Y., Liu, C.-C., Liu, L.-W., Ma, F.-F., Wu, X.-Y., Wu, M., Hang, Y.-Y., Chen, J.-Q., Shao, Z.-Q., & Wang, B. (2016). Identification of Arbuscular Mycorrhiza (AM)-Responsive microRNAs in Tomato. *Frontiers in Plant Science*, 7, 429. <https://doi.org/10.3389/fpls.2016.00429>
- 80) Wu, Y., Ma, X., Pan, Z., Kale, S. D., Song, Y., King, H., Zhang, Q., Presley, C., Deng, X., Wei, C.-I., & Xiao, S. (2018). Comparative genome analyses reveal sequence features reflecting distinct modes of host-adaptation between dicot and monocot powdery mildew. *BMC Genomics*, 19(1), 705. <https://doi.org/10.1186/s12864-018-5069-z>

- 81) Wu, Y., & Zhou, J.-M. (2013). Receptor-Like Kinases in Plant Innate Immunity. *Journal of Integrative Plant Biology*, 55(12), 1271–1286. <https://doi.org/10.1111/jipb.12123>
- 82) Xu, X., Wang, L., & Ding, D. (2004). Learning module networks from genome-wide location and expression data. *FEBS Letters*, 578(3), 297–304. <https://doi.org/10.1016/j.febslet.2004.11.019>
- 83) Xue, D.-Q., Chen, X.-L., Zhang, H., Chai, X.-F., Jiang, J.-B., Xu, X.-Y., & Li, J.-F. (2017). Transcriptome Analysis of the Cf-12-Mediated Resistance Response to Cladosporium fulvum in Tomato. *Frontiers in Plant Science*, 7. <https://doi.org/10.3389/fpls.2016.02012>
- 84) Xue, Y., Li, W., Li, M., Ru, N., Chen, S., Jiu, M., Feng, H., Wei, L., Daly, P., & Zhou, D. (2024). Biological Control of a Root-Knot Nematode Meloidogyne incognita Infection of Tomato (*Solanum lycopersicum* L.) by the Oomycete Biocontrol Agent *Pythium oligandrum*. *Journal of Fungi*, 10(4), Article 4. <https://doi.org/10.3390/jof10040265>
- 85) Yao, K., Yao, Y., Ding, Z., Pan, X., Zheng, Y., Huang, Y., Zhang, Z., Li, A., Wang, C., Li, C., & Liao, W. (2023). Characterization of the FLA Gene Family in Tomato (*Solanum lycopersicum* L.) and the Expression Analysis of SFLAs in Response to Hormone and Abiotic Stresses. *International Journal of Molecular Sciences*, 24(22), Article 22. <https://doi.org/10.3390/ijms242216063>
- 86) Yao, X., Guo, H., Zhang, K., Zhao, M., Ruan, J., & Chen, J. (2023). Trichoderma and its role in biological control of plant fungal and nematode disease. *Frontiers in Microbiology*, 14. <https://doi.org/10.3389/fmicb.2023.1160551>
- 87) Zeng, D.-E., Hou, P., Xiao, F., & Liu, Y. (2015). Overexpression of Arabidopsis XERICO gene confers enhanced drought and salt stress tolerance in rice (*Oryza Sativa* L.). *Journal of Plant Biochemistry and Biotechnology*, 24(1), 56–64. <https://doi.org/10.1007/s13562-013-0236-4>
- 88) Zeng, Z., Liu, Y., Feng, X.-Y., Li, S.-X., Jiang, X.-M., Chen, J.-Q., & Shao, Z.-Q. (2023). The RNAome landscape of tomato during arbuscular mycorrhizal symbiosis reveals an evolving RNA layer symbiotic regulatory network. *Plant Communications*, 4(1). <https://doi.org/10.1016/j.xplc.2022.100429>
- 89) Zhang, B., & Horvath, S. (2005). A General Framework for Weighted Gene Co-Expression Network Analysis. *Statistical Applications in Genetics and Molecular Biology*, 4(1). <https://doi.org/10.2202/1544-6115.1128>
- 90) Zhao, T., Liu, W., Zhao, Z., Yang, H., Bao, Y., Zhang, D., Wang, Z., Jiang, J., Xu, Y., Zhang, H., Li, J., Chen, Q., & Xu, X. (2019). Transcriptome profiling reveals the response process of tomato carrying Cf-19 and Cladosporium fulvum interaction. *BMC Plant Biology*, 19(1), 572. <https://doi.org/10.1186/s12870-019-2150-y>
- 91) Zhu, Y., Zhang, X., Zhang, Q., Chai, S., Yin, W., Gao, M., Li, Z., & Wang, X. (2022). The transcription factors VaERF16 and VaMYB306 interact to enhance resistance of grapevine to *Botrytis cinerea* infection. *Molecular Plant Pathology*, 23(10), 1415–1432. <https://doi.org/10.1111/mpp.13223>
- 92) Zouari, I., Salvioli, A., Chialva, M., Novero, M., Miozzi, L., Tenore, G. C., Bagnaresi, P., & Bonfante, P. (2014). From root to fruit: RNA-Seq analysis shows that arbuscular mycorrhizal symbiosis may affect tomato fruit metabolism. *BMC Genomics*, 15(1), 221. <https://doi.org/10.1186/1471-2164-15-221>

Copyright license

The datasets used in this thesis were obtained from the NCBI Sequence Read Archive (SRA) and are publicly available. Figure 1 was created using a licensed BioRender template, in accordance with Bio Render's academic publication guidelines. Figure 2 is a workflow that I designed in draw.io website. Figure 3 to 44 were generated by the author using data derived from the SRA datasets (refer Table 3). All sources have been appropriately cited.

Table 3 Figure licenses

Figure number	License
Figure 1	Biorender template
Figure 2	Designed in draw.io
Figure 3 - 44	Data generated plots

Supplementary

i. Metadata containing the accession numbers.

Table 4 represents the metadata of this study. This table contains the SRR ids, timepoint of sample collection mentioned in hours, whether the sequencing is paired-end or single-end, instrument used for sequencing, treatment, microbe infected, interaction type, tissue sample, and study from which the SRR ids were obtained for this study to be carried out.

Table 4 Metadata containing the accession numbers.

SRR Ids	Timepoint (Hours)	Sequencing	Instrument	Treatment	Microbe	Interaction type	Tissue type	Study
SRR16643549	1008	PAIRED	Illumina HiSeq	lncRNA-seq of AMF infected roots	arbuscular mycorrhizal	Symbiotic	Root	(Z. Zeng et al., 2023)
SRR16643548	1008	PAIRED	Illumina HiSeq	lncRNA-seq of AMF infected roots	arbuscular mycorrhizal	Symbiotic	Root	
SRR16643547	1008	PAIRED	Illumina HiSeq	lncRNA-seq of AMF infected roots	arbuscular mycorrhizal	Symbiotic	Root	
SRR16643551	1008	PAIRED	Illumina HiSeq	lncRNA-seq of AMF not infected roots	arbuscular mycorrhizal	Wild type	Root	
SRR16643550	1008	PAIRED	Illumina HiSeq	lncRNA-seq of AMF not infected roots	arbuscular mycorrhizal	Wild type	Root	
SRR16643552	1008	PAIRED	Illumina HiSeq	lncRNA-seq of AMF not infected roots	arbuscular mycorrhizal	Wild type	Root	
SRR12775028	NA	PAIRED	NextSeq 550	treated	Bacillus subtilis	Symbiotic	Leaf	(Muñoz et al., 2021)
SRR29790750	NA	PAIRED	Illumina NovaSeq 6000	Tomato_AMS_rep1	arbuscular mycorrhizal	Symbiotic	Root	Publication not available
SRR29790749	NA	PAIRED	Illumina NovaSeq 6000	Tomato_AMS_rep2	arbuscular mycorrhizal	Symbiotic	Root	

SRR29790748	NA	PAIRED	Illumina NovaSeq 6000	Tomato_AMS_rep3	arbuscular mycorrhizal	Symbiotic	Root	
SRR29790753	NA	PAIRED	Illumina NovaSeq 6000	Tomato_CK_rep1	arbuscular mycorrhizal	Wild type	Root	
SRR29790752	NA	PAIRED	Illumina NovaSeq 6000	Tomato_CK_rep2	arbuscular mycorrhizal	Wild type	Root	
SRR29790751	NA	PAIRED	Illumina NovaSeq 6000	Tomato_CK_rep3	arbuscular mycorrhizal	Wild type	Root	
SRR3037371	NA	SINGLE	Illumina HiSeq 2000	with microbe infection	Rhizophagus irregularis	Symbiotic	Root	(P. Wu et al., 2016)
SRR3037372	NA	SINGLE	Illumina HiSeq 2000	without microbe infection	Rhizophagus irregularis	Wild type	Root	
SRR17235452	264	PAIRED	HiSeq X Ten	Roots from plants treated with fungal spore suspension	Serendipita indica	Symbiotic	Root	(Ntana et al., 2022)
SRR17235453	264	PAIRED	HiSeq X Ten	Roots from plants treated with water	Serendipita indica	Wild type	Root	
SRR17235454	264	PAIRED	HiSeq X Ten	Leaves from plants treated with fungal spore suspension	Serendipita indica	Symbiotic	Leaf	
SRR17235455	264	PAIRED	HiSeq X Ten	Leaves from plants treated with water	Serendipita indica	Wild type	Leaf	
SRR5679560	1461	SINGLE	Illumina HiSeq 1500	tomato+funnellifomis Mosseae	Funneliformis mosseae	Symbiotic	Leaf	(Miozzi et al., 2020)
SRR5679561	1461	SINGLE	Illumina HiSeq 1500	tomato+funnellifomis Mosseae	Funneliformis mosseae	Symbiotic	Leaf	
SRR5679565	1461	SINGLE	Illumina HiSeq 1500	tomato+funnellifomis mosseae	Funneliformis mosseae	Symbiotic	Leaf	
SRR5679557	1461	SINGLE	Illumina HiSeq 1500	tomato control	Funneliformis mosseae + Cucumber Mosaic Virus	Wild type	Leaf	
SRR5679558	1461	SINGLE	Illumina HiSeq 1500	tomato control	Funneliformis mosseae + Cucumber Mosaic Virus	Wild type	Leaf	
SRR5679559	1461	SINGLE	Illumina HiSeq 1500	tomato control	Funneliformis mosseae + Cucumber Mosaic Virus	Wild type	Leaf	
SRR5679562	672	SINGLE	Illumina HiSeq 1500	tomato+ cucumber mosaic virus	Cucumber Mosaic Virus	Biotrophic	Leaf	
SRR5679563	672	SINGLE	Illumina HiSeq 1500	tomato+ cucumber mosaic virus	Cucumber Mosaic Virus	Biotrophic	Leaf	
SRR5679556	672	SINGLE	Illumina HiSeq 1500	tomato+ cucumber mosaic virus	Cucumber Mosaic Virus	Biotrophic	Leaf	
SRR18426570	0	PAIRED	Illumina HiSeq 2500	0 dpi susceptible biological replicate 1	Cladosporium Fulvum	Wild type	Leaf	(Jiang et al., 2022)
SRR18426585	0	PAIRED	Illumina HiSeq 2500	0 dpi susceptible biological replicate 2	Cladosporium Fulvum	Wild type	Leaf	
SRR18426584	0	PAIRED	Illumina HiSeq 2500	0 dpi susceptible biological replicate 3	Cladosporium Fulvum	Wild type	Leaf	
SRR18426583	216	PAIRED	Illumina HiSeq 2500	9 dpi susceptible biological replicate 1	Cladosporium Fulvum	Biotrophic	Leaf	

SRR18426582	216	PAIRED	Illumina HiSeq 2500	9 dpi susceptible biological replicate 2	Cladosporium Fulvum	Biotrophic	Leaf	
SRR18426581	216	PAIRED	Illumina HiSeq 2500	9 dpi susceptible biological replicate 3	Cladosporium fulvum	Biotrophic	Leaf	
SRR18426580	360	PAIRED	Illumina HiSeq 2500	15 dpi susceptible biological replicate 1	Cladosporium Fulvum	Biotrophic	Leaf	
SRR18426579	360	PAIRED	Illumina HiSeq 2500	15 dpi susceptible biological replicate 2	Cladosporium fulvum	Biotrophic	Leaf	
SRR18426578	360	PAIRED	Illumina HiSeq 2500	15 dpi susceptible biological replicate 3	Cladosporium fulvum	Biotrophic	Leaf	
SRR28479822	24	PAIRED	Illumina NovaSeq 6000	Inoculated _ rep1	Meloidogyne incognita	Biotrophic	Root	(Y. Xue et al., 2024)
SRR28479823	24	PAIRED	Illumina NovaSeq 6000	Inoculated _ rep2	Meloidogyne incognita	Biotrophic	Root	
SRR28479824	24	PAIRED	Illumina NovaSeq 6000	Inoculated _ rep3	Meloidogyne incognita	Biotrophic	Root	
SRR28479825	24	PAIRED	Illumina NovaSeq 6000	Control	Meloidogyne incognita	Wild type	Root	
SRR28479826	24	PAIRED	Illumina NovaSeq 6000	Control	Meloidogyne incognita	Wild type	Root	
SRR28479827	24	PAIRED	Illumina NovaSeq 6000	control	Meloidogyne incognita	Wild type	Root	
SRR4048988	144	SINGLE	Illumina HiSeq 1500	tomato (for OnM2)	Oidium neolyopersici-powdery mildew	Biotrophic	Leaf	(Y. Wu et al., 2018)
SRR4048989	144	SINGLE	Illumina HiSeq 1500	tomato (for OnM2)	Oidium neolyopersici-powdery mildew	Biotrophic	Leaf	
SRR4048990	144	SINGLE	Illumina HiSeq 1500	tomato (for OnM2)	Oidium neolyopersici-powdery mildew	Biotrophic	Leaf	
SRR19164834	12	SINGLE	NextSeq 500	Mock tomato 12 h	Botrytis cinerea	Control-wild type	Leaf	(Qin et al., 2023)
SRR8074946	12	SINGLE	NextSeq 500	Mock tomato 12 h	Botrytis cinerea	Control-wild type	Leaf	
SRR8074951	12	SINGLE	NextSeq 500	Mock tomato 12 h	Botrytis cinerea	Control-wild type	Leaf	
SRR19164833	16	SINGLE	NextSeq 500	Mock tomato 16 h	Botrytis cinerea	Control-wild type	Leaf	
SRR8074944	16	SINGLE	NextSeq 500	Mock tomato 16 h	Botrytis cinerea	Control-wild type	Leaf	
SRR8074947	16	SINGLE	NextSeq 500	Mock tomato 16 h	Botrytis cinerea	Control-wild type	Leaf	
SRR23048700	24	SINGLE	NextSeq 500	Mock tomato 24 h	Botrytis cinerea	Control-wild type	Leaf	
SRR8074945	24	SINGLE	NextSeq 500	Mock tomato 24 h	Botrytis cinerea	Control-wild type	Leaf	
SRR8074942	24	SINGLE	NextSeq 500	Mock tomato 24 h	Botrytis cinerea	Control-wild type	Leaf	

SRR8074956	12	SINGLE	NextSeq 500	B.cinerea-tomato 12 h	Botrytis cinerea	Control	Leaf	
SRR8074958	12	SINGLE	NextSeq 500	B.cinerea-tomato 12 h	Botrytis cinerea	Control	Leaf	
SRR8074959	12	SINGLE	NextSeq 500	B.cinerea-tomato 12 h	Botrytis cinerea	Control	Leaf	
SRR8074957	16	SINGLE	NextSeq 500	B.cinerea-tomato 16 h	Botrytis cinerea	Control	Leaf	
SRR8074954	16	SINGLE	NextSeq 500	B.cinerea-tomato 16 h	Botrytis cinerea	Control	Leaf	
SRR8074955	16	SINGLE	NextSeq 500	B.cinerea-tomato 16 h	Botrytis cinerea	Control	Leaf	
SRR8074950	24	SINGLE	NextSeq 500	B.cinerea-tomato 24 h	Botrytis cinerea	Control	Leaf	
SRR8074952	24	SINGLE	NextSeq 500	B.cinerea-tomato 24 h	Botrytis cinerea	Control	Leaf	
SRR8074953	24	SINGLE	NextSeq 500	B.cinerea-tomato 24 h	Botrytis cinerea	Control	Leaf	
SRR23959357	0	PAIRED	Illumina NovaSeq 6000	biological replicate 1 of 0h control	drought stress	Control-wild type	Leaf	(Dong et al., 2023)
SRR23959356	0	PAIRED	Illumina NovaSeq 6000	biological replicate 2 of 0h control	drought stress	Control-wild type	Leaf	
SRR23959355	0	PAIRED	Illumina NovaSeq 6000	biological replicate 3 of 0h control	drought stress	Control-wild type	Leaf	
SRR23959354	2	PAIRED	Illumina NovaSeq 6000	biological replicate 1 of 2h dehydration	drought stress	Control	Leaf	
SRR23959353	2	PAIRED	Illumina NovaSeq 6000	biological replicate 2 of 2h dehydration	drought stress	Control	Leaf	
SRR23959352	2	PAIRED	Illumina NovaSeq 6000	biological replicate 3 of 2h dehydration	drought stress	Control	Leaf	
SRR23959351	4	PAIRED	Illumina NovaSeq 6000	biological replicate 1 of 4h dehydration	drought stress	Control	Leaf	
SRR23959350	4	PAIRED	Illumina NovaSeq 6000	biological replicate 2 of 4h dehydration	drought stress	Control	Leaf	
SRR23959349	4	PAIRED	Illumina NovaSeq 6000	biological replicate 3 of 4h dehydration	drought stress	Control	Leaf	
SRR25551615	48	SINGLE	Illumina	2 days after infection with TYLCV by agroinoculation	tomato yellow leaf curl virus	Biotrophic	Leaf	Publication not available
SRR25551616	48	SINGLE	Illumina	2 days after infection with TYLCV by agroinoculation	tomato yellow leaf curl virus	Biotrophic	Leaf	
SRR25521634	48	SINGLE	Illumina	2 days after control treatment (non-treated)	tomato yellow leaf curl virus	Wild type	Leaf	
SRR25521635	48	SINGLE	Illumina	2 days after control treatment (non-treated)	tomato yellow leaf curl virus	Wild type	Leaf	