

Asymmetric Transcriptional Responses: Shared Upregulation and Distinct Downregulation in *Arabidopsis thaliana* under Heat and Salt Stress

Brayden Ritter
brritter@ualberta.ca
University of Alberta
Edmonton, Alberta, Canada

Grace Padberg
gpadberg@ualberta.ca
University of Alberta
Edmonton, Alberta, Canada

Alyssa Wong
ajwong1@ualberta.ca
University of Alberta
Edmonton, Alberta, Canada



Figure 1: The model organism *Arabidopsis thaliana*. This study analyzes the transcriptomic response of this plant to heat and salt stress. (Image source: Dreamstime.com)

Abstract

Global climate change threatens agricultural stability by exposing crops to complex abiotic stressors, primarily heat and high salinity. Developing resilient crops requires understanding whether plants utilize a conserved "core" stress response or distinct, stress-specific regulatory programs. In this study, we analyzed publicly available RNA-seq data from *Arabidopsis thaliana* to compare transcriptomic responses under heat and salt stress using a standardized Galaxy-based workflow. Our results reveal a significant functional asymmetry: while salt stress induces broad, systemic genomic reprogramming affecting both upregulation and downregulation, heat stress triggers a targeted upregulation of nucleic acid metabolism with minimal downregulation. Despite these differences in scope, we identified a shared "core activation module" among upregulated genes. Protein-protein interaction (PPI) analysis revealed this module is anchored by high-degree network hubs, including heat shock proteins (HSP90-1, HSP70-4) and the transcriptional master regulator HSFA2, which converge on Reactive Oxygen Species (ROS) mitigation and protein folding. Conversely, downregulated pathways showed negligible overlap, indicating that mechanisms for metabolic suppression are highly stress-specific. These findings suggest that broad-spectrum resilience may be engineered by targeting shared upstream regulators like HSFA2, rather than manipulating the highly divergent downstream metabolic pathways.

Keywords

Arabidopsis thaliana, RNA-seq, Heat stress, Salt stress, Gene Ontology

ACM Reference Format:

Brayden Ritter, Grace Padberg, and Alyssa Wong. 2025. Asymmetric Transcriptional Responses: Shared Upregulation and Distinct Downregulation in *Arabidopsis thaliana* under Heat and Salt Stress. In *Proceedings of BME 415/615: Biomedical Engineering Project (BME 415/615)*. ACM, New York, NY, USA, 13 pages.

1 Introduction

Changing temperature and precipitation patterns caused by global climate change pose serious risks to plant life by inducing stress, impairing photosynthetic efficiency, and disrupting nutrient uptake [1]. Because these changes occur against a backdrop of steady global population growth that will reach nearly 10 billion people by 2050 [24], their impacts are felt most acutely in the agricultural sector. The cereal grains, maize, rice, and wheat supply 48% of the world's caloric intake [6], meaning that even small improvements in their stress tolerance can translate into substantial gains in yield and stability.

Among stressors related to climate change, salinity and heat stress are consistently identified as two of the most damaging to crop productivity. Soil salinity often rises when reduced precipitation and irrigation practices without adequate drainage drive salt accumulation in the root zone of plants, disrupting cellular homeostasis and osmotic balance, promoting ion toxicity, and ultimately suppressing growth and yield [15]. Elevated temperatures

on the other hand, reduce seed germination, destabilize protein and membrane structures, impair photosynthetic efficiency, and accelerate reproductive failure through widespread changes in gene expression and metabolism [11]. Collectively, abiotic stresses can reduce yields of major crops by up to 50% on a global average [3], and soil salinity alone is estimated to cause an estimated \$20-30 billion (USD) in crop-production losses annually [18]. These agricultural impacts represent a substantial, unpredictable, and recurring economic burden and highlight the need for tools that can rapidly identify molecular targets for improving stress resilience.

In practice, dissecting the genomic responses to these environmental changes directly in cereals is challenging. These crops have large, complex genomes, large biomass, and comparatively long life cycles [2], all of which make high-throughput genetic and transcriptomic studies more demanding in the laboratory compared to other plants. As a result, researchers often rely on model organisms such as *Arabidopsis thaliana* (*A. thaliana*; Thale Cress), which combines a small, fully sequenced genome, short life cycle, and ease of genetic manipulation with extensive community resources and functional annotations [22, 25]. Many hormone receptors, signalling pathways, and transcriptional regulators characterized in *A. thaliana* are conserved across flowering plants, allowing insights from this species to inform work in agronomically important cereals [25].

Plants mount both stress-specific and shared adaptive responses to abiotic stressors such as cold, salinity, drought, and heat. Well characterized examples include the ICE-CBF-COR pathways in cold stress and the Salt-Overly-Sensitive (SOS) pathway in ionic homeostasis [12]. Heat stress pathways are comparatively less well-defined, highlighting the novelty of our study in this context. At the same time, multiple stresses converge on common regulatory pathways such as Mitogen-Activated Protein Kinase (MAPK) cascades and Reactive Oxygen Species (ROS), which act as damaging by-products as well as secondary messengers that alter gene expression, metabolic function, and cellular redox status [10, 12, 19]. In animals, stress signals are often funneled through a limited number of “hub” pathways, such as the Hypothalamic-Pituitary-Adrenal (HPA) Axis [20]. By analogy, a key open question in plant biology is whether different abiotic stressors, such as heat and salt, are processed through a shared pathway of core-stress response genes and pathways, or whether they rely on distinct, stress specific transcriptional programs with limited, if any, overlap.

Addressing this question has practical implications as well. If we can develop a pipeline to identify genes or pathways that mediate tolerance to multiple stressors, then selective breeding, genome editing tools, or molecular priming strategies could be used to engineer broad-spectrum stress resilience [11]. Robust and accessible bioinformatics architectures are essential for systematically identifying stress-responsive genes and for mapping them onto higher-order pathways using resources such as Galaxy for data processing, and PANTHER and the Gene Ontology (GO) consortium for functional annotation.

In this project, we use *A. thaliana* as a model to compare the transcriptional responses to heat and salt stress. Publicly available RNA-seq datasets generated under well-defined stress conditions are processed through a standardized Galaxy-based workflow, including quality control, read alignment, and differential expression analysis, to identify genes that are up-regulated or down-regulated

compared to control groups. We then integrate these results with functional annotation databases to determine whether heat and salt stress primarily elicit overlapping or distinct gene expression profiles in *A. thaliana*, and to highlight candidate genes and pathways that may act as central integrators of abiotic stress. Ultimately, the objective of this study is to characterize the transcriptomic response of *Arabidopsis thaliana* to heat and salt stress. Specifically, we aim to compare the functional conservation of upregulated versus downregulated gene programs to determine if the plant’s strategies for activation and suppression are shared across stressors or distinct to each condition.

2 Methods

2.1 RNA-Seq Data Sources and Preprocessing

We used publicly available RNA-sequence data to compare how *A. thaliana* responds to heat and salt stress. The heat stress dataset (GSE132415), found on NCBI Gene Expression Omnibus (GEO), provided both processed expression data and access to the raw reads via the linked SRA project (SRP200947). The salt stress dataset (PRJNA950536) was obtained directly from NCBI Sequence Read Archive (SRA) and had expression matrices available. The differing data formats between the two datasets meant that steps would need to be taken to process the heat data into a format similar to that made available for the salt stress data. In order to process the heat data to match the format of the salt data (raw sequence data in FASTQ format) we first identified the individual SRA run files for both the control and heat stress treatments (SRR9257060 - SRR9257063) [7]. Each sample contained two trials, resulting in four overall files being processed. The following steps were conducted on usegalaxy.org. This platform was chosen over the usegalaxy.ca instance due to the higher cloud storage. We first ran “Faster Download and Extract Reads in FASTQ” on the SRA Accession list provided by the dataset (SRR9257060 - SRR9257063). This provided us with a “pair-end data” file correlating to each of the treatments.

2.2 Quality Control with FastQC

Next, we ran FastQC [8] to assess the quality of both the heat control and heat stress treatment reads. We evaluated the following metrics, each of which we describe in the following:

- Per-base sequence quality
- Adapter content
- Sequence duplication levels
- Overrepresented sequences
- Per-sequence guanine-cytosine (GC) content

Per-base sequence quality is the most important plot in FastQC output. The primary metric is the Phred quality score, which represents the quality of the bases in the dataset. In order to be considered good quality, a Phred value greater than 30 is expected. Our samples produced Phred scores consistently greater than 30 and peaking at 40, confirming high-quality sequencing reads.

Adapter content was also assessed, with the ideal target being 0%. All four heat stress trials showed no detectable adapter contamination.

Sequence duplication represents over-sequencing, contamination, and Polymerase Chain Reaction (PCR) artifacts, and levels

ideally fall below 60-70%. The heat data we processed showed very low sequence duplication levels, with most below 2% and the highest approximately 31%, still much below the required threshold.

Overrepresented sequences were inspected to detect possible single-sequence contamination (e.g., from viral or bacterial sources). Fewer than 1% of reads were overrepresented across samples, indicating that the data was not significantly impacted by overrepresentation.

Finally, we examined per-sequence GC content. This is an important metric to ensure the data came from a single organism, that no PCR bias was present, and overall that the data was handled according to best practices. Typical *A. thaliana* GC content is approximately 43% [22]. Our samples had a GC content of about 45%, which is consistent with expectations given that stress-induced sequences are likely to carry a higher GC content (51–52%).

2.3 MultiQC Aggregated Quality Assessment

Although we explored the quality report for each trial independently, we also ran the UseGalaxy MultiQC tool [5] to ensure that the combined data fell within the expected thresholds. We examined the percentage of duplicate reads, average GC content, and total read counts for both forward and reverse directions.

Duplicate read percentages were 65.8% (forward) and 56.4% (reverse). Average GC content was 45.0% for both directions, consistent with FastQC results. Total read counts were 33.1 million (forward) and 26.0 million (reverse).

2.4 Alignment with STAR

We then used the STAR Aligner [4] via the Galaxy platform [9], a high-efficiency and high-accuracy aligner, to map reads to the *A. thaliana* reference genome. Running this tool provided data related to which genes came from each read, how many reads belong to each gene, and how cleanly the test and reference genome match. The TAIR10 genome and associated Gene Transfer Format (GTF) annotation files were obtained from the National Library of Medicine [17].

We selected “paired-end (as individual datasets)” mode and used the FASTA and GTF files obtained from NIH. We also selected “build index with gene model” since we were processing spliced mRNA and were using a GTF file to indicate the possible splice sites. This option also ensures compatibility with downstream tools such as featureCounts.

2.5 Conversion to BAM and Sorting

To prepare for the next analysis steps, we needed to ensure that our alignment files were in BAM format. BAM is a compressed binary format that allows for rapid searching and is generally considered the standard for storing read alignments. BAM files contain genomic positions, alignment quality, splice junctions, mismatches, paired-end formation, and strand orientation.

We used `samtools sort` to sort the aligned reads by genomic coordinate, an essential step for all downstream analyses.

2.6 FeatureCounts

We used the FeatureCounts tool from Galaxy [13] to convert the aligned BAM reads obtained from the previous step into a matrix

that can be used by DESeq2. For this step of processing we utilized the *A. thaliana* reference gene mentioned previously [17]. The alignment files were the direct output obtained from BAM. The strandedness option was set to unstranded (-s 0) to match the methods used to obtain the genetic material from the original researchers [7]. The paired setting was set to, “Yes, paired-end and count them as 1 single fragment”. The feature type selected was Exon, and gene identifier, gene_id. The outputs obtained from this step were then processed with DESeq2.

2.7 DESeq2

We ran DESeq2 analysis on Galaxy [14] in order to determine which genes showed significant differences or overlap in gene expression between conditions. “How” was set to select datasets per level. The factor name was set to condition with the factor levels being set to “salt” and “heat” and the count files came from the previous steps.

2.8 Python Processing After DESeq2

After the data processing was carried out, our data was in a consistent format: tabular DESeq2 files. We used Python scripting to carry out data analysis and convert our files into forms that were meaningful for biological interpretation. This process consisted of several steps, beginning by converting the tabular DESeq2 format to CSV. During this step, we also added filters on $\log_2(\text{FoldChange})$ and padj to retain only significant entries. We initially used a Log2 Fold Change (LFC) threshold of 1.0, but the resulting gene set was too large for meaningful downstream analysis, so we increased the threshold to an LFC of 2.0 to focus on genes with more substantial expression changes. Next, we merged both filtered datasets into a singular file to simplify the creation of intermediate datasets with biologically relevant groupings. We separated gene responses based on stress and regulation direction, obtaining a set of significant genes for each category: salt upregulated, salt downregulated, heat upregulated, and heat downregulated. Gene lists were also extracted that explored the intersection between the stresses: one dataset for upregulated genes, and one for downregulated genes.

2.9 PANTHER Functional Enrichment Analysis

We completed functional enrichment analysis of our significant gene sets using the PANTHER Overrepresentation Test [16] (PANTHER version 19.0, released 2024-06-20; <https://pantherdb.org/>). Gene lists for each stress and regulation direction (four total) were uploaded using *Arabidopsis thaliana* as the reference organism. We selected the Gene Ontology (GO) Biological Process Complete annotation set [23] and used the default *A. thaliana* reference list. Overrepresentation was evaluated using Fisher’s Exact Test with False Discovery Rate (FDR) correction for multiple testing. We exported the results table and retained only significantly enriched GO terms ($\text{FDR} \leq 0.05$) for downstream interpretation.

2.10 Functional Overlap Quantification and Normalization

Because the number of significant GO Biological Process terms differed substantially between heat and salt stress conditions, we quantified functional similarity using both raw overlap counts and normalized overlap metrics. For each stress and regulation direction

pair, we defined a set of significant GO terms using an FDR threshold of 0.05. We then evaluated functional overlap by computing the size of the intersection and union of the two GO-term sets.

To normalize the overlap, we computed three metrics (with sets $A = \text{heat}$ and $B = \text{salt}$):

- **Jaccard index**

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

- **Overlap coefficient**

$$O(A, B) = \max\left(\frac{|A \cap B|}{|A|}, \frac{|A \cap B|}{|B|}\right)$$

- **Directional overlap percentage relative to A (heat)**

$$\%A = \frac{|A \cap B|}{|A|} \times 100\%$$

These metrics were computed using a custom Python script. The Jaccard index is a measure of similarity between two datasets, ranging from 0 to 1, with 1 indicating complete similarity. The overlap coefficient quantifies the degree to which the smaller set is contained within the larger. The range for this metric is 0 to 1, with 1 indicating that the smaller set is completely contained in the larger set. The directional overlap percentage was taken relative to the heat dataset because it was smaller in both the upregulated and downregulated comparisons.

2.11 GO Term Plot Generation

After completing the PANTHER Overrepresentation Test on our significant gene sets (heat_up, heat_down, salt_up, salt_down), we created two Python scripts to process and visualize the resulting GO terms. The first script filtered each dataset to retain only statistically significant terms ($\text{FDR} \leq 0.05$). After filtering, we used the second script to generate seven visualizations:

- Four bar charts showing the top enriched GO terms for each dataset (limited to 15 terms for clarity).
- One bar chart showing GO-term overlap between heat and salt.
- One bar chart showing the number of significant terms per dataset.
- One heatmap summarizing all significant GO terms across conditions.

These visualizations allowed us to compare the breadth and specificity of enriched biological processes across stress conditions.

In addition, we generated a DEG heatmap summarizing log₂ fold change values for all significant genes across heat and salt, separated by up- and down-regulated groups, to visualize overlap and stress-specific expression patterns prior to GO-based pathway analysis.

2.12 STRING Network Analysis

After obtaining and comparing the GO-term lists for each stress and regulation direction, we sought additional quantitative evidence to validate our observations. To achieve this, we focused on the genes shared between both stresses and performed protein–protein interaction (PPI) network analysis.

We selected the STRING database for network visualization, as it is widely used in bioinformatics to display physical and functional protein interactions. STRING generates PPI network diagrams that reveal relationships between proteins, allowing us to assess whether the shared genes form functional complexes or pathways rather than representing a random collection.

PPI networks were constructed for the shared upregulated and downregulated gene sets using STRING [21]. For each set, we inputted the set of relevant genes, chose *Arabidopsis thaliana* as the organism, and applied Markov Cluster Algorithm (MCL) clustering with an inflation parameter of 3. The resulting network diagrams were saved for reference, and summary statistics were obtained from the “Network Stats” tab.

We also examined the functional enrichment tables under STRING’s “Analysis” tab, examining GO Biological Process, Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome pathways, and UniProt Annotated Keywords. Using the program’s filtering tools, we selected the most highly enriched and statistically significant terms ($\text{FDR} < 0.05$). These results were compiled into a summary table representing the core biological processes and pathways associated with the shared stress-responsive genes.

3 Results

Differential expression analysis revealed distinct transcriptional responses to heat and salt stress. To visualize these gene-level patterns prior to pathway-level interpretation, we generated a heatmap summarizing significant differential expression under heat and salt stress (Figure 2). Each column represents a stress condition and regulation direction (heat up/down, salt up/down), and cell values reflect log₂ fold change (genes not significant in a given condition/direction appear as 0). The heatmap illustrates that salt stress produces a substantially broader and stronger transcriptional response than heat stress, and that gene-level overlap between the two stresses is limited (approximately 15 shared DEGs under our thresholds). These gene-level patterns motivate the pathway analyses below, which test whether functional overlap is stronger at the level of enriched biological processes than at the level of individual genes.

We performed Gene Ontology (GO) enrichment analysis on the differentially expressed genes under heat and salt conditions. We identified significantly enriched GO terms ($\text{FDR} \leq 0.05$) across four categories: heat-upregulated, heat-downregulated, salt-upregulated, and salt-downregulated.

Figure 3 shows a high level of variation in the scale of the response between the two stresses. Salt stress caused a much broader response, with over 140 terms in the upregulated group and 110 in the downregulated group. Heat stress was much more limited, resulting in fewer than 30 upregulated terms and fewer than 10 downregulated terms.

For the heat-upregulated genes, the top terms were linked to nucleic acid metabolism, RNA processing, and gene expression (Figure 4).

The heat-downregulated group had significantly fewer terms, only seven total (Figure 5). These terms were related to broad categories such as “cellular process” and “metabolic process” rather than specific biological pathways.

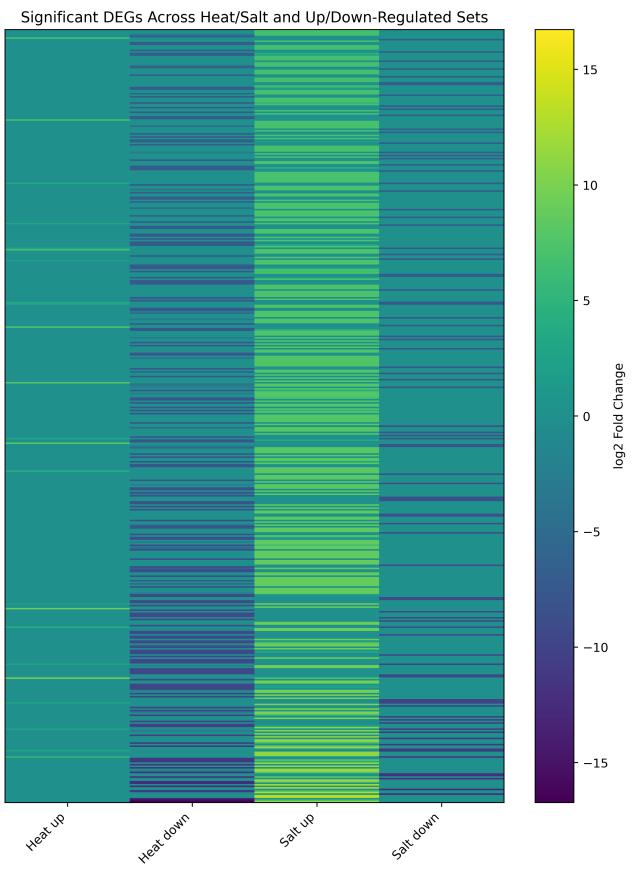


Figure 2: Heatmap of significant differentially expressed genes (DEGs) across heat and salt stress. Columns separate DEGs by condition and regulation direction. Cell color indicates \log_2 fold change; a value of 0 indicates the gene was not significant in that condition/direction under the chosen thresholds. This visualization highlights the comparatively broader salt response and the limited gene-level overlap between stresses.

In the salt stress analysis, the upregulated terms were focused heavily on stress response pathways. Figure 6 shows enrichment for “response to chemical,” “response to abiotic stimulus,” and specifically “response to oxygen levels/hypoxia”.

The salt-downregulated genes (Figure 7) showed enrichment for secondary metabolic processes, specifically phenylpropanoid and suberin biosynthesis.

Finally, we compared the enriched terms from the two conditions to obtain a set of shared processes. Figure 8 shows a moderate overlap in upregulated processes, with the 14 shared terms suggesting some common stress activation pathways. However, there was little to no overlap in the downregulated genes, with only one shared term.

Since the salt datasets produced substantially larger GO term sets than the heat datasets, we also calculated normalized overlap metrics (Table 1). Based on both the raw counts and the normalized

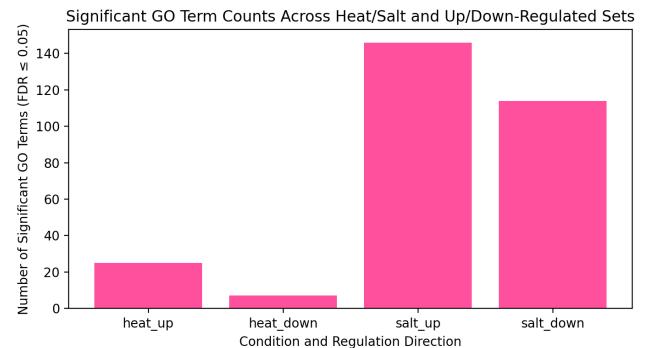


Figure 3: Count of significantly enriched Gene Ontology (GO) terms ($FDR \leq 0.05$) identified across heat and salt stress conditions.

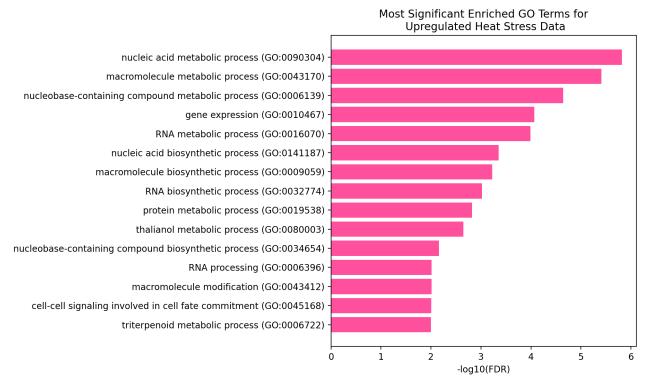


Figure 4: Top 15 significantly enriched GO terms for genes upregulated under heat stress.

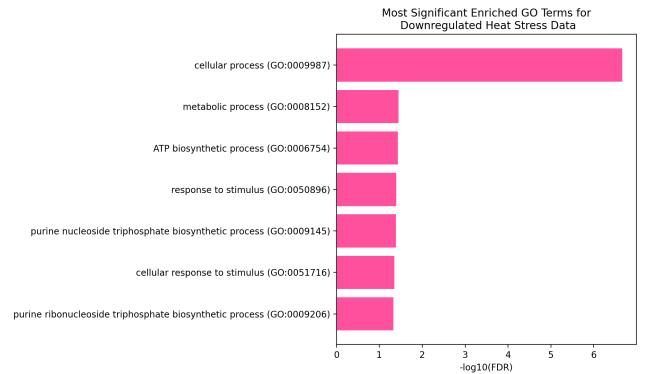


Figure 5: Significantly enriched GO terms for genes downregulated under heat stress.

results, the upregulated programs showed greater functional similarity than the downregulated programs. It can also be noted that the overlap coefficient indicated that 56% of the smaller upregulated term set (heat-up) was shared with salt-up, whereas only 14.3% of

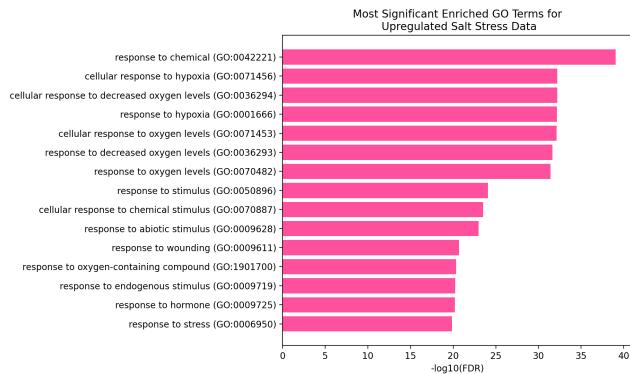


Figure 6: Top 15 significantly enriched GO terms for genes upregulated under salt stress.

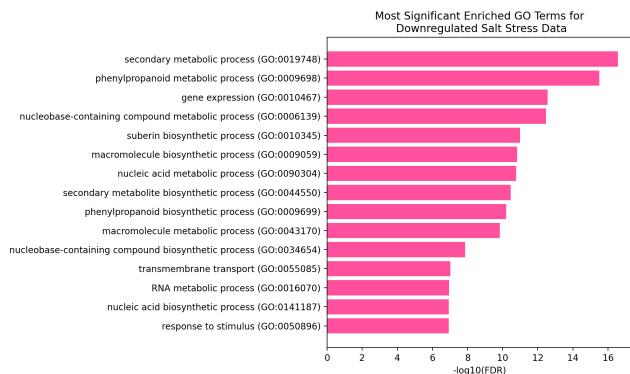


Figure 7: Top 15 significantly enriched GO terms for genes downregulated under salt stress.

Table 1: Normalized overlap metrics for enriched GO biological process terms (heat vs. salt).

Comparison	A	B	A ∩ B	Jaccard	Overlap coeff.	% of A
Up (heat vs salt)	25	146	14	0.0892	0.5600	56.00
Down (heat vs salt)	7	114	1	0.0083	0.1429	14.29

the smaller downregulated term set (heat-down) overlapped with salt-down. The heatmap in Figure 9 highlights these specific cases: hypoxia-related terms appear only in the salt-upregulated data, while cell-wall related terms (suberin and phenylpropanoid) are specific to the salt-downregulated data.

Protein-protein interaction (PPI) networks were constructed for the shared upregulated and downregulated gene sets using the STRING database.

Figure 10 shows the PPI network for the shared upregulated gene sets. The network contains 337 nodes and three regions of significant interaction. This network has a PPI enrichment p-value of $< 1.0 \times 10^{-16}$ and significantly more interactions than expected (751 edges versus 178 expected).

Visual inspection of the network reveals two distinct, high-density clusters (expanded views available in Appendix Figure 13

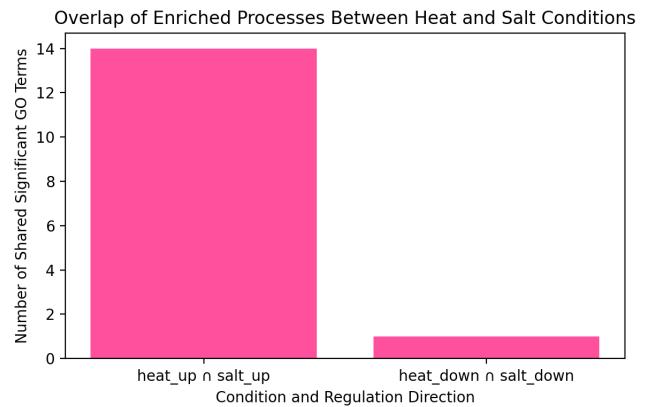


Figure 8: Intersection of all significant enriched GO terms between heat and salt stress conditions, showing shared upregulated and downregulated processes. This figure shows the raw overlap counts. For the normalized metrics, refer to (Table 1).

Table 2: List of shared Gene Ontology (GO) terms found in both Heat and Salt stress conditions (Intersection).

GO ID	Term Description
Shared Upregulated Processes (Core Activation)	
GO:0032774	RNA biosynthetic process
GO:0016070	RNA metabolic process
GO:0006396	RNA processing
GO:0010467	Gene expression
GO:00141187	Nucleic acid biosynthetic process
GO:0090304	Nucleic acid metabolic process
GO:0034654	Nucleobase-containing compound biosynthetic process
GO:0006139	Nucleobase-containing compound metabolic process
GO:0009059	Macromolecule biosynthetic process
GO:0043170	Macromolecule metabolic process
GO:0019538	Protein metabolic process
GO:0006629	Lipid metabolic process
GO:0006996	Organelle organization
GO:0042221	Response to chemical
Shared Downregulated Processes	
GO:0050896	Response to stimulus

and Figure 14). The primary, central cluster is anchored by highly connected “hub” genes, specifically the heat shock proteins HSP90-1 and HSP70-4. Their dense connectivity aligns with the enrichment strength observed for “Protein complex oligomerization” (Strength = 1.15). We also identified HSFA2, a known transcriptional master regulator, within this central cluster, suggesting it may drive the expression of these downstream chaperones. In addition, the presence of BIP3 (an ER chaperone) and APX2 (an antioxidant enzyme) within this same primary cluster provides direct molecular evidence for the “Protein processing in ER” pathway ($FDR 2.54 \times 10^{-13}$) and “Response to hydrogen peroxide” (Strength = 1.21), respectively. Additionally, the UniProt keyword “Chaperone” was significantly

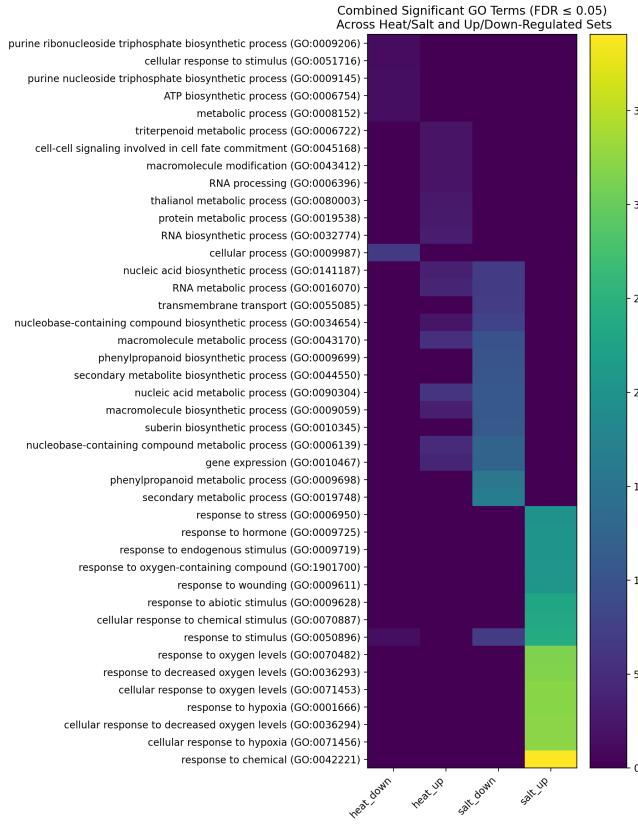


Figure 9: Heatmap showing the enrichment significance ($-\log_{10}$ FDR) of top GO terms across all four experimental conditions.

Table 3: Selected functional enrichment terms for the shared upregulated gene network (Source: STRING-db).

Category	Term Description	Strength	FDR
Biological Process	Response to hydrogen peroxide	1.21	3.05e-09
	Protein complex oligomerization	1.15	6.62e-07
	Response to heat	0.97	4.67e-15
Molecular Function	Unfolded protein binding	1.09	1.33e-09
KEGG Pathways	Protein processing in ER	0.96	2.54e-13
	Oxidative phosphorylation	0.70	0.008
Keywords	Chaperone	0.62	0.024

enriched (FDR = 0.024), covering a large proportion of the nodes in this central region.

In the secondary cluster (Figure 14), the identification of mitochondrial components such as COX1 and COX2 confirms that the shared response extends beyond the nucleus to metabolic energy regulation.

The PPI network for the shared downregulated gene sets is shown in Figure 11, consisting of 92 nodes. This network has significantly more interactions than expected, with 35 edges versus the

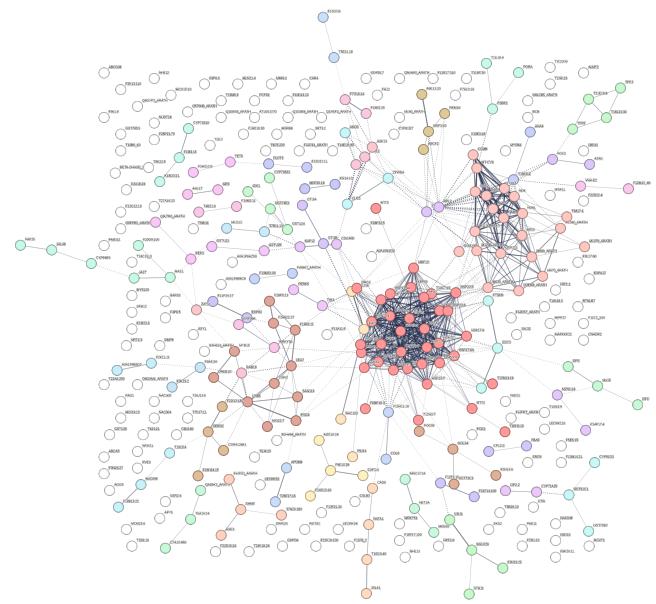


Figure 10: PPI network for shared upregulated gene sets containing 337 nodes and three regions of significant interaction. This network has a PPI enrichment p-value of $< 1.0 \times 10^{-16}$ and 751 edges versus 178 expected. An enlarged image of this network (Figure 12) and supplemental figures of the two major groups (Figure 13, Figure 14) are available in the Appendix.

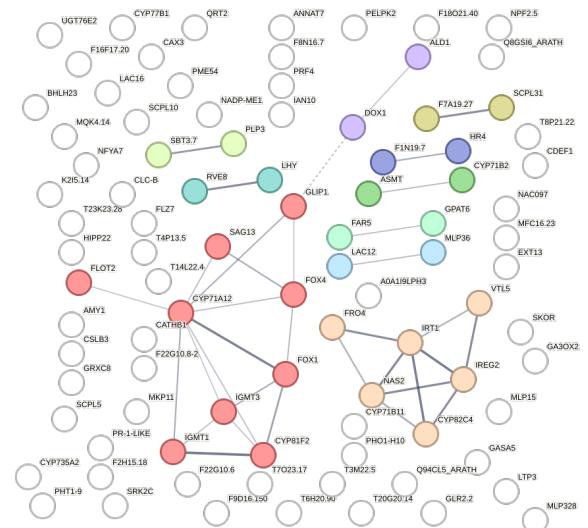


Figure 11: PPI network for shared downregulated gene sets consisting of 92 nodes and 35 edges versus 5 expected. This network has a PPI enrichment p-value of $< 1.0 \times 10^{-16}$.

expected count of 5. This network has a PPI enrichment p-value of $< 1.0 \times 10^{-16}$.

Table 4: Selected functional enrichment terms for the shared downregulated gene network (Source: STRING-db).

Category	Term Description	Strength	FDR
Biological Process	Secondary metabolic process	0.85	0.037
	Inorganic ion transmembrane transport	0.83	0.017
	Transition metal ion homeostasis	1.13	0.050
Molecular Function	Oxidoreductase activity	0.59	2.2e-04
	Iron ion transmembrane transporter activity	1.61	0.049
	Heme binding	0.79	0.041
Reactome Pathways	Cytochrome P450	0.84	0.019
Keywords	Iron	0.70	4.8e-04
	Heme	0.84	0.001

Functional enrichment analysis (Table 4) identified significant enrichment for “Secondary metabolic process” (Strength = 0.85) and “Inorganic ion transmembrane transport” (Strength = 0.83). Molecular function analysis highlighted “Oxidoreductase activity” (Strength = 0.59) and “Iron ion transmembrane transporter activity” (Strength = 1.61) as key repressed functions. Additionally, Reactome pathway analysis identified “Cytochrome P450” as a significantly enriched term within the downregulated set.

In sharp contrast to the dense interconnectivity of the upregulated genes, the shared downregulated network (Figure 11) appears highly fragmented. It consists of small, isolated clusters rather than a cohesive regulatory module.

4 Discussion

The disparity in the number of enriched GO terms reveals significant differences in how *A. thaliana* responds to heat versus salt stress. The high number of significant GO terms in salt (more than 250) suggests that it induces a broad, systemic genomic reprogramming. Conversely, heat stress triggers a much more targeted, narrow response in the plant, with fewer than 40 enriched pathways. This implies that salt stress requires adjustments to metabolism, signaling, and growth at a system-wide level. In contrast, heat stress was characterized mainly by the upregulation of RNA and nucleic-acid-associated processes, representing a narrower and more specialized response.

The heat-upregulated genes were enriched primarily for terms associated with nucleic acid metabolism, RNA processing, and gene expression. This likely reflects an increased transcriptional and translational capacity to synthesize heat shock proteins and chaperones required to prevent protein misfolding during thermal stress. Heat-downregulated pathways were few and general, suggesting that the plant relies mainly on activating protective mechanisms rather than suppressing specific biological processes under elevated temperatures.

In salt-stressed *A. thaliana*, there was overlap among enriched terms related to “hypoxia” and “oxygen levels,” suggesting crosstalk between osmotic stress signaling and low-oxygen sensing pathways. Suberin and phenylpropanoid are key components of cell walls, and downregulation of their biosynthetic processes implies suppressed growth. This likely reflects a strategy to reduce growth and costly secondary metabolism to conserve resources for survival.

The moderate overlap in upregulated processes suggests the existence of a “core environmental stress response” in *A. thaliana*,

in which general defense mechanisms are activated regardless of the specific stressor. In contrast, the minimal overlap in downregulated processes indicates that there is no comparable “core” program for pathway suppression. Instead, downregulation appears highly stress-specific; for example, reduced cell-wall-related biosynthesis is prominent under salt stress but not under heat stress.

We then focused on the overlapping upregulated and downregulated genes shared between salt and heat stress, as we expected gene-level overlap to provide stronger biological evidence for shared responses than GO-term overlap alone. The initial results suggested strong coherence among the shared gene sets. To evaluate this more formally, we performed a STRING protein–protein interaction (PPI) analysis on the shared upregulated and downregulated genes.

For the shared upregulated genes, STRING identified a dense, highly interconnected network containing 337 nodes and 751 edges, compared to an expected 178 edges (PPI enrichment $p < 1.0 \times 10^{-16}$). This demonstrates that the shared upregulated genes form a coordinated, non-random “core stress response” module, despite representing a small fraction of total DEGs. Enriched terms included “response to hydrogen peroxide,” “response to heat,” “protein complex oligomerization,” “unfolded protein binding,” and “protein processing in the endoplasmic reticulum.” Together, these results suggest that both heat and salt converge on shared proteostasis and ROS-centered stress response mechanisms.

STRING analysis of the shared downregulated genes revealed a smaller network (92 nodes, 35 edges) that was still substantially enriched compared to the expected 5 edges (PPI enrichment $p < 1.0 \times 10^{-16}$). The enriched functional categories included “secondary metabolic processes,” “inorganic ion transport,” “iron homeostasis,” “cytochrome P450 pathways,” and “iron binding.” This indicates that when both stresses suppress genes, they do so in biologically meaningful clusters, largely related to metabolism and redox-active enzymatic systems. It should also be mentioned that when considering gene expression and subsequent GO pathways, it is important to remember that individual genes may be activated in multiple pathways and thus this may skew our results. This means that the up-regulation of any given gene may indicate altered expression in multiple GO pathways and this could contribute to the high number of pathways we observed activated. Future works may benefit from quantifying the overlap of individual genes in these overlapping pathways as targeting these genes specifically may also provide resistance to abiotic stresses in plants.

The findings indicate that while heat and salt share some activation mechanisms, the plant’s strategy for conserving energy via downregulation is more stress-specific. However, transcriptomic data alone only provides correlative rather than causative evidence. To definitively determine what regulatory mechanisms are behind this conserved activation, investigation of functional validation would be necessary. Specifically, future work on functional validation should focus on the high-degree nodes identified in this paper, such as HSFA2 and MBF1C. The broad downstream regulation of chaperones regulated by these transcription factors represent an ideal target for future genetic engineering. As mentioned in the introduction, discovering methods to induce broad-spectrum stress-resilience in crop plants would provide direct benefits to the agriculture industry. Master regulators such as HSFA2 may allow simultaneous resistance to both salt and heat stress and offer a

solution to the problem posed in this paper. Future works should prioritize CRISPR-Cas9 knockout or overexpression of MBF1C and HSFA2 to quantify their specific impacts on plant biology. Experiments that turn off important "hub" genes would help us show whether the shared ROS and chaperone pathways we identified are actually necessary for the plant cells to survive under both conditions.

Overall, the results obtained proved to be very strong, and show a clear overlap in gene expression as a response to both salt and heat stress independently. Despite the strength of the results, the study still has limitations to be discussed. The first is time; the processing of the data alone took nearly 50 hours, though most of this was idle waiting for the steps to process. The analysis of this study also proved to be very time consuming, with each gene showing up-regulation when exposed to salt or heat stress needing to be researched individually and in-depth to understand the full function of the gene and how this relates to observed physiological states. This, however, also opens the opportunity for future directions of research. The genes outlined in this study provide a strong baseline for a genetic modification study looking at how gene biology may be altered in order to negate the impacts of stress responses in the given crop. Other future works could see the pipeline outlined in our work applied to cereal crops by teams that have more human hours and compute to dedicate. As indicated in the introduction, the main benefit of results from a study of this nature are felt by agricultural companies, who can employ these findings to produce crops that have an innate ability to resist these stressors, although this comes with its own list of ethical concerns.

5 Conclusion

In conclusion, these analyses support a two-part model of stress response in *A. thaliana*. First, heat and salt share a tightly connected activation module dominated by ROS mitigation, chaperone activity, and protein-folding machinery, representing a conserved core stress response. Second, the two stresses differ markedly in the pathways they suppress: salt strongly suppresses secondary metabolism and ion-related processes, whereas heat exhibits minimal targeted downregulation. Thus, pathway suppression appears to be highly stress-specific rather than part of a broadly conserved program.

Overall, the STRING results strengthen the GO-based conclusions by confirming that the observed patterns reflect coordinated biological strategies rather than statistical artifacts. To build on this, future investigations should prioritize the functional characterization of high-degree nodes in our network analysis. By eliminating specific genes in the shared ROS mitigation and protein-folding pathways, we can test whether they are key in stress tolerance. Although it would be valuable to expand the study toward testing the universality of this core module, narrowing in on the specific genetic regulation of the hubs we identified offers the most direct path to engineering improved crop resilience.

6 Data and Code Availability

The full list of differentially expressed genes, custom Python scripts for DESeq2 conversion and GO-term plotting, and additional high-resolution plots are available in the project repository at: <https://github.com/gpadberg/bme-415-615-project>.

References

- [1] Preetha Bhadra, Sagar Maitra, Tanmoy Shankar, Akbar Hossain, Subhashisa Praharaj, and Tariq Aftab. 2022. Climate change impact on plants: Plant responses and adaptations. In *Plant perspectives to global climate changes*. Elsevier, 1–24.
- [2] Philipp Borrill. 2020. Blurring the boundaries between cereal crops and model plants. *New Phytologist* 228, 6 (2020), 1721–1727.
- [3] John S Boyer. 1982. Plant productivity and environment. *Science* 218, 4571 (1982), 443–448.
- [4] Alexander Dobin, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 1 (2013), 15–21. doi:10.1093/bioinformatics/bts635
- [5] Philip Ewels et al. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 19 (2016), 3047–3048.
- [6] Food and Agriculture Organization of the United Nations. 2021. *The impact of disasters and crises on agriculture and food security: 2021*. FAO, Rome. doi:10.4060/cb3673en Accessed: 2025-11-26.
- [7] National Center for Biotechnology Information. 2019. Gene Expression Omnibus: GSE132415. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE132415>. Accessed: 2023-11-26.
- [8] Galaxy Training Network. 2025. Quality Control: Galaxy Training Network Tutorial. <https://training.galaxyproject.org/training-material/topics/sequence-analysis/tutorials/quality-control/tutorial.html> Accessed: 2025-11-26.
- [9] Galaxy Training Network. 2025. RNA-seq reads to counts using STAR: Galaxy Training Network Tutorial. <https://training.galaxyproject.org/training-material/topics/transcriptomics/tutorials/rna-seq-bash-star-align/tutorial.html> Accessed: 2025-11-26.
- [10] Sarvajeet Singh Gill and Narendra Tuteja. 2010. Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. *Plant physiology and biochemistry* 48, 12 (2010), 909–930.
- [11] Miguel Gonzalez Guzman, Francesco Cellini, Vasileios Fotopoulos, Raffaella Balestrini, and Vicent Arbona. 2022. New approaches to improve crop tolerance to biotic and abiotic stresses. *Physiologia plantarum* 174, 1 (2022), e13547.
- [12] Guo-Tao Huang, Shi-Liang Ma, Li-Ping Bai, Li Zhang, Hui Ma, Ping Jia, Jun Liu, Ming Zhong, and Zhi-Fu Guo. 2012. Signal transduction during cold, salt, and drought stresses in plants. *Molecular biology reports* 39, 2 (2012), 969–987.
- [13] Yang Liao, Gordon K. Smyth, and Wei Shi. 2013. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 7 (2013), 923–930. doi:10.1093/bioinformatics/btt656
- [14] Michael I. Love, Wolfgang Huber, and Simon Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15, 12 (2014). doi:10.1186/s13059-014-0550-8
- [15] Ying Ma, Maria Celeste Dias, and Helena Freitas. 2020. Drought and salinity stress responses and microbe-induced tolerance in plants. *Frontiers in plant science* 11 (2020), 591911.
- [16] Huaiyu Mi, Anushya Muruganujan, Dinath Ebert, Xiaosong Huang, and Paul D Thomas. 2019. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Research* 47, D1 (2019), D419–D426. doi:10.1093/nar/gky1038
- [17] National Center for Biotechnology Information. 2020. Arabidopsis thaliana Genome Assembly TAIR10.1 (GCF_000001735.3). https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001735.3/. Accessed: 2025-11-26.
- [18] Manzoor Qadir, Emmanuelle Quillérou, Vinay Nangia, Ghulam Murtaza, Minakshi Singh, Richard J. Thomas, Pay Drechsel, and Andrew D. Noble. 2014. Economics of salt-induced land degradation and restoration. *Natural Resources Forum* 38, 4 (2014), 282–295. doi:10.1111/1477-8947.12054
- [19] Hong-bo Shao, Li-ye Chu, Ming-an Shao, Cheruth Abdul Jaleel, and Mi Hong-mei. 2008. Higher plant antioxidants and redox signaling under environmental stresses. *Comptes rendus Biologies* 331, 6 (2008), 433–441.
- [20] Julietta A Sheng, Natalie J Bales, Sage A Myers, Anna I Bautista, Mina Roueinfar, Taben M Hale, and Robert J Handa. 2021. The hypothalamic-pituitary-adrenal axis: development, programming actions of hormones, and maternal-fetal interactions. *Frontiers in behavioral neuroscience* 14 (2021), 601939.
- [21] STRING Consortium. 2025. STRING: functional protein association networks. <https://string-db.org/> Accessed: 2025-11-26.
- [22] David Swarbreck et al. 2010. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research* 38, suppl_1 (2010), D1020–D1027.

- [23] The Gene Ontology Consortium. 2025. Gene Ontology Database. <https://doi.org/10.5281/zenodo.16423886>. doi:10.5281/zenodo.16423886 Released 2025-07-22.
- [24] United Nations Department of Economic and Social Affairs, Population Division. 2022. *World Population Prospects 2022: Summary of Results*. United Nations, New York. <https://www.un.org/development/desa/pd/content/World-Population-Prospects-2022> Accessed: 2025-11-26.
- [25] Andrew W Woodward and Bonnie Bartel. 2018. Biology in bloom: a primer on the *Arabidopsis thaliana* model system. *Genetics* 208, 4 (2018), 1337–1349.

A Supplementary Network Analysis

This appendix provides high-resolution views of the Protein-Protein Interaction (PPI) networks discussed in the Results section.

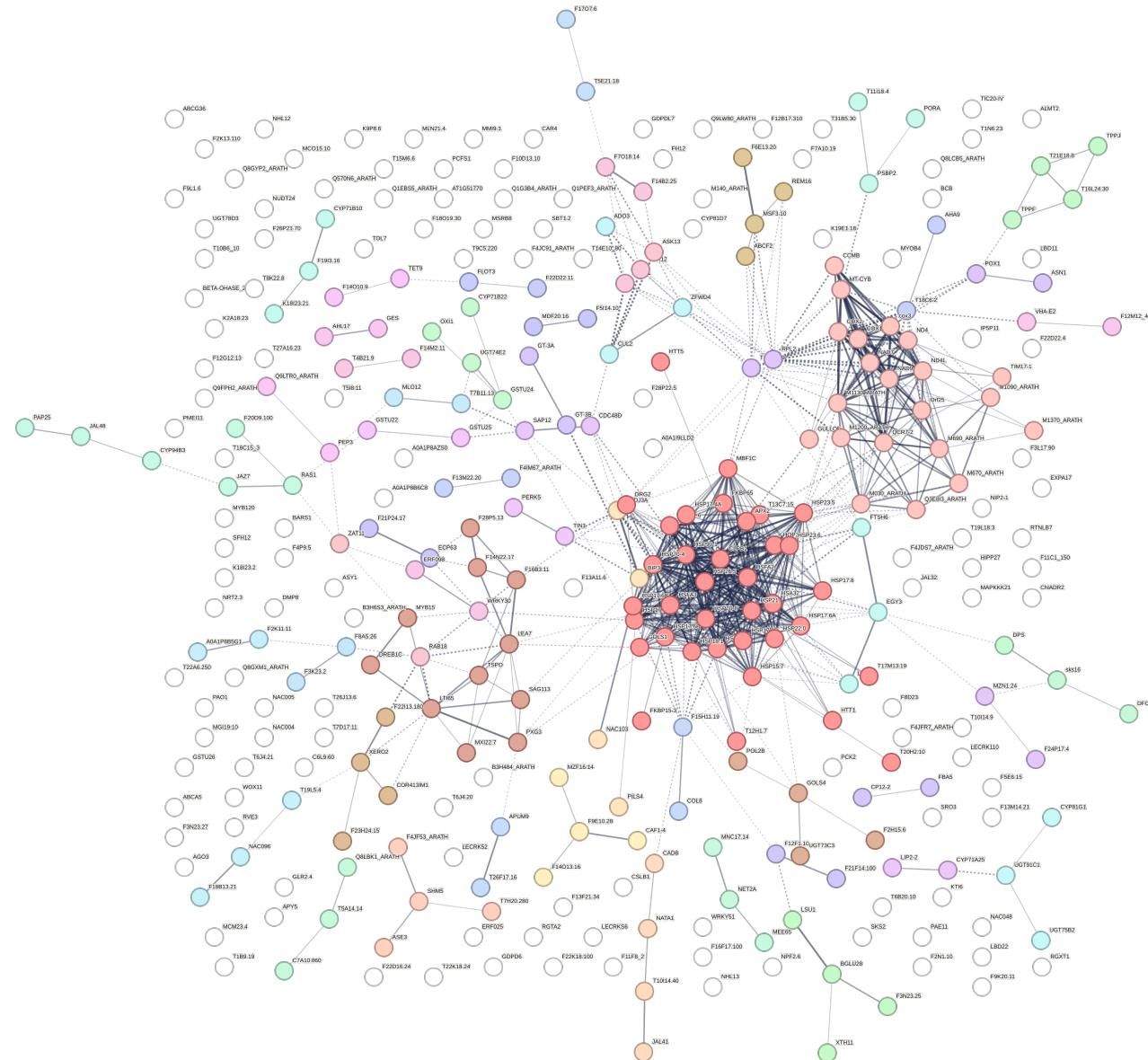


Figure 12: Full-size version of the shared upregulated gene PPI network (same as Figure 10). The network contains 337 nodes and 751 edges (178 expected), with PPI enrichment $p < 1.0 \times 10^{-16}$.

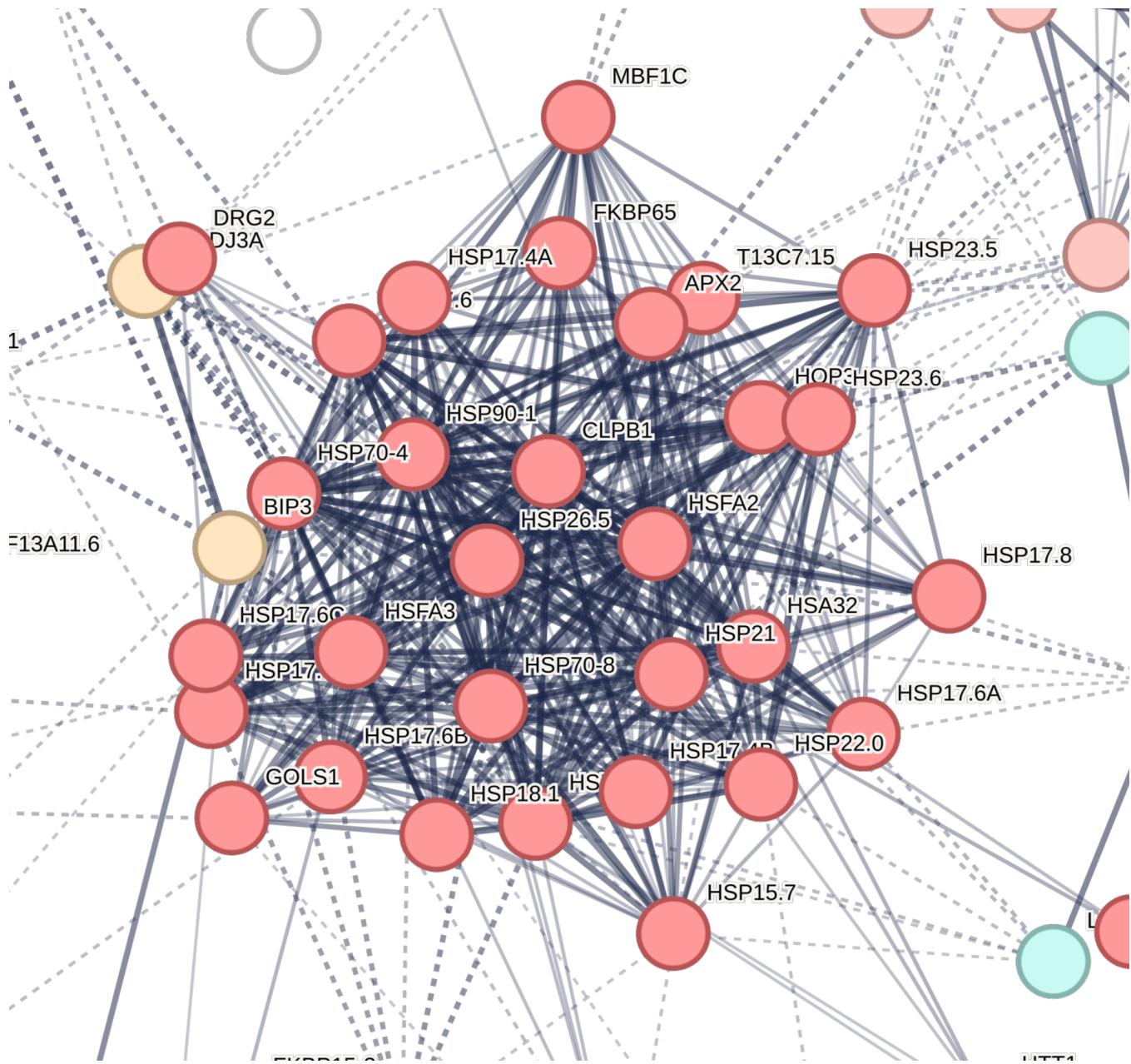


Figure 13: Enlarged image of the primary central cluster (Protein Folding/Chaperones) from the shared upregulated PPI network.

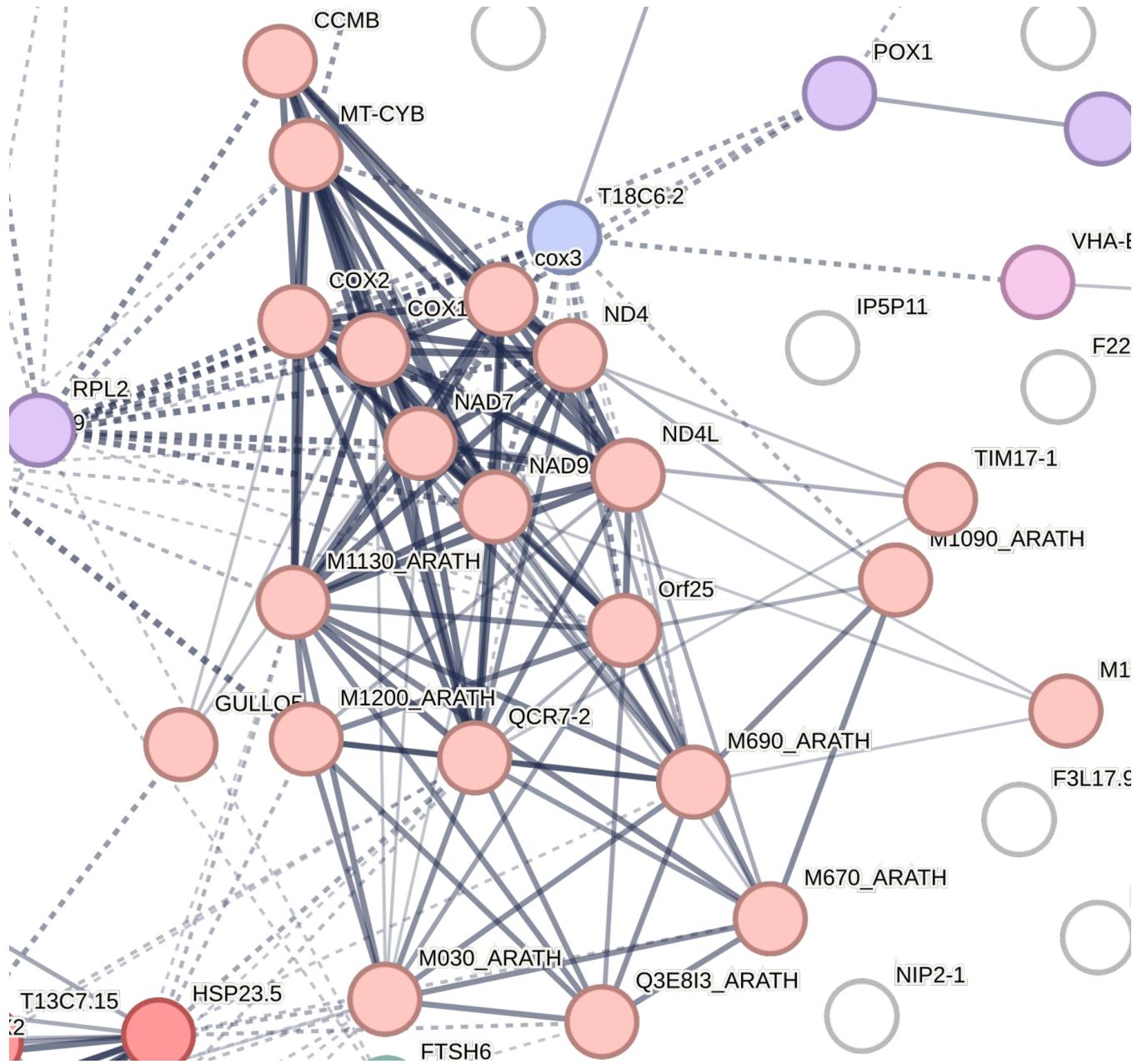


Figure 14: Enlarged image of the secondary central cluster (ROS/Oxidative Stress) from the shared upregulated PPI network.