**Pathway Enrichment Analysis of Physcomitrium patens CESA Knockout DEGs**

The plant cell wall is a dynamic structure critical for growth, development, and environmental response. Cellulose synthase (CESA) genes encode enzymes responsible for synthesizing cellulose, the primary component of the cell wall. In the model moss *Physcomitrium patens*, knocking out CESA genes disrupts cellulose biosynthesis, making it possible to study compensatory pathways that maintain cell integrity. This project investigates how CESA knockout in *P. patens* affects functional pathways, protein domains, and biological processes among differentially expressed genes (DEGs). Using Pfam domain enrichment, Gene Ontology (GO) enrichment, and KEGG pathway analysis, the goal is to identify molecular functions and pathways activated in response to CESA loss. These analyses will provide insights into cell wall remodeling and stress adaptation mechanisms.

**Methods**

The list of DEGs used for enrichment was previously generated using the DESeq2 package. This list served as the input for the Pfam, GO, and KEGG enrichment analyses. All analyses and visualizations were performed in R (version 4.3.1). Data manipulation was conducted using the tidyverse package. Visualizations were generated using ggplot2, with patchwork used to assemble multi-panel figures and Viridis applied for color-blind-friendly palettes. Analysis scripts, input data, and output figures are available at the project’s GitHub repository:<https://github.com/kawhelan/cesa-ko-enrichment> for full reproducibility.

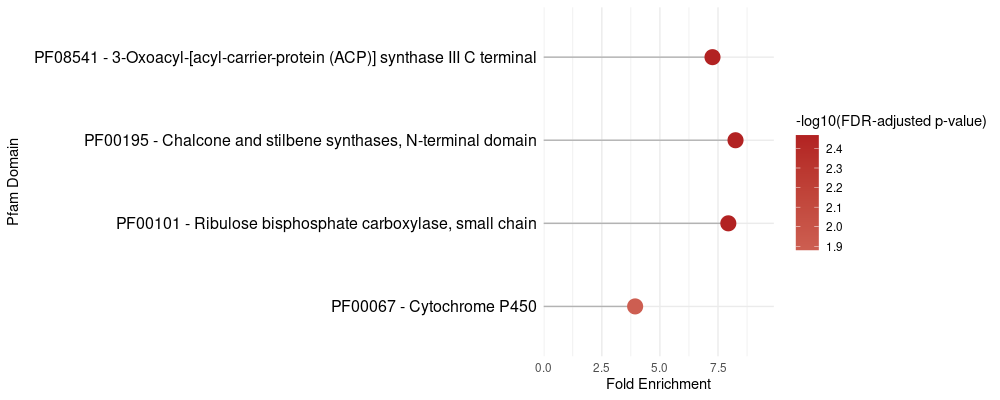
Differentially expressed genes identified from *P. patens* CESA knockout lines (n = 3 biological replicates) were filtered for FDR-adjusted p-values (padj) < 0.05 and used for enrichment analysis. Gene IDs from the DEG list were annotated with Pfam domains using the Phytozome annotation dataset. The frequency of Pfam domains in the DEG set was compared against the background frequency of domains in the entire genome. A Fisher’s exact test was performed for each domain to assess statistical enrichment. The resulting p-values were corrected for multiple testing using the Benjamini-Hochberg FDR method. Initial exploration of domain distributions was conducted using a bar plot of the ten most frequent Pfam domains and a heatmap comparing domain frequencies between upregulated and downregulated DEGs. However, the final enrichment results focused on statistically significant domains, which were visualized using a dot plot ranked by fold enrichment and FDR significance, and a lollipop plot that highlighted domains with FDR < 0.05. The lollipop plot was included in the results section as it concisely summarized the significant enrichment findings.

Gene Ontology enrichment analysis was conducted to identify biological processes (BP), molecular functions (MF), and cellular components (CC) overrepresented among CESA knockout DEGs. GO term assignments were obtained using the PANTHER classification system, and enrichment results were provided as separate output tables for each category. The DEG dataset used for this analysis was filtered for FDR-adjusted p-values (padj) < 0.05 and contained gene identifiers suitable for input into PANTHER. Enriched terms were filtered for FDR < 0.05 to identify significant results for each GO category. To visualize biological processes (BP), the top 15 terms were selected based on FDR significance and plotted as a bar graph, ranked by fold enrichment. For overall functional classification, pie charts were generated for BP, MF, and CC, showing the top seven terms in each category with the remaining grouped as “Others”. These exploratory visuals provided a broad overview of GO classifications among DEGs. To evaluate significant enrichment patterns, a GO dot plot was generated, highlighting the top 30 biological processes ranked by gene ratio (the proportion of DEGs annotated to a given term relative to the background). The gene ratio, FDR-adjusted p-values, and gene counts were incorporated into the plot to depict the magnitude and significance of enrichment. The bar plot with pie charts and GO dot plot were included in the results section to show the main functional categories and highlight statistically enriched processes.

KEGG pathway analysis was used to investigate the metabolic and signaling pathways represented in CESA knockout DEGs. The list of DEGs, filtered for FDR-adjusted p-values (padj) < 0.05), was submitted to the DAVID functional annotation tool, which maps genes to KEGG pathways and calculates Benjamini-Hochberg adjusted p-values (FDR) to evaluate overrepresentation. Pathways with FDR < 0.05 were retained for downstream analysis. To visualize the distribution and significance of enriched pathways, the top 30 KEGG terms were plotted in a dot plot displaying the gene ratio (DEGs mapped to each pathway relative to the total number of background genes), the number of DEGs, and the adjusted significance of each pathway. This allowed enrichment strength and confidence to be assessed across pathways. The KEGG dot plot was included in the results to highlight key metabolic and regulatory pathways involved in the response to CESA disruption.

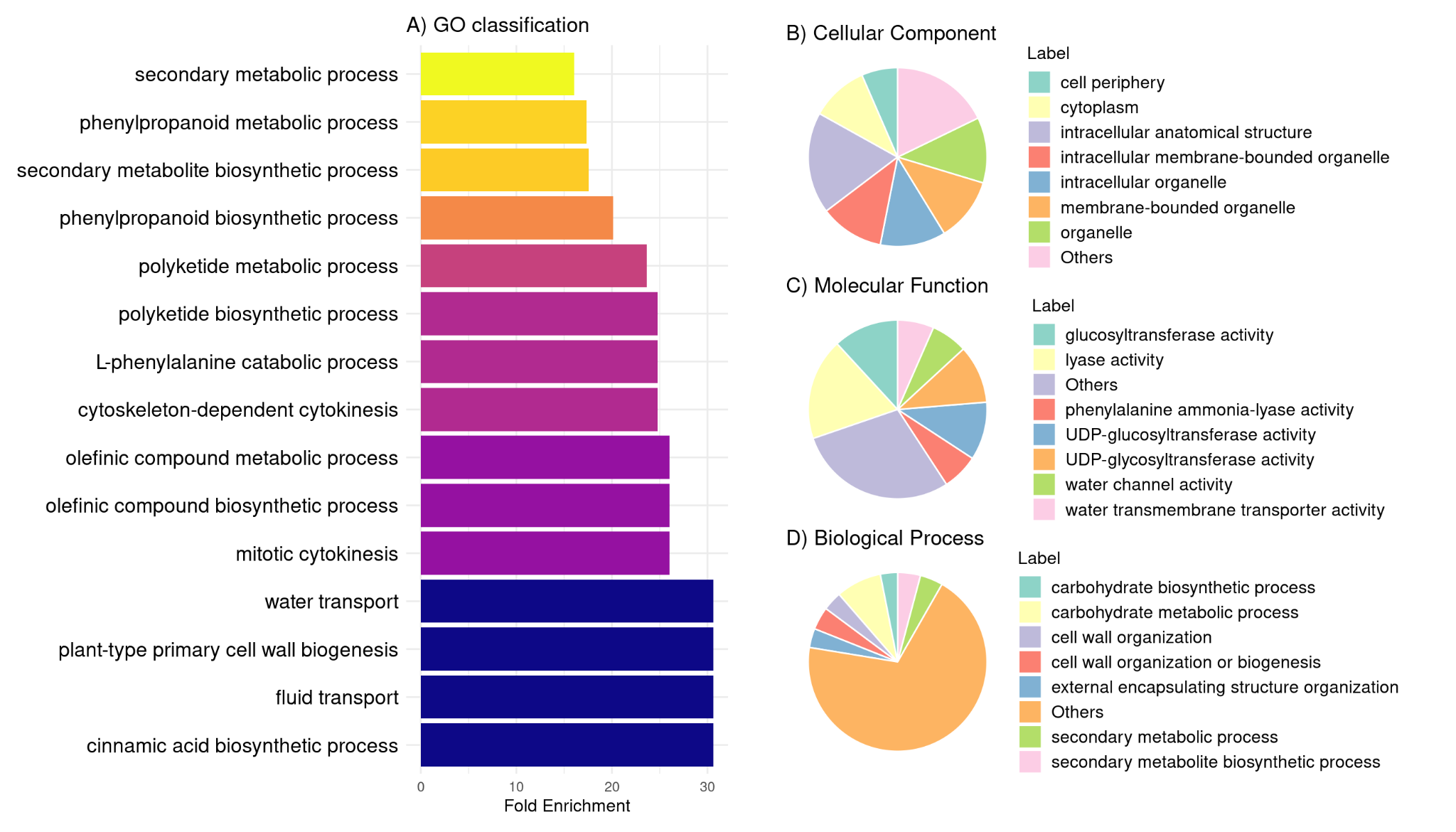
**Results**

Pfam domain enrichment analysis revealed that several protein domains were overrepresented among CESA knockout DEGs compared to the genomic background. Although a bar plot and heatmap were initially used to explore the frequency and regulation patterns of domains, enrichment testing identified four domains that were significantly enriched (FDR < 0.05) (Figure 1). These domains were associated with secondary metabolism, photosynthesis, and stress responses. This suggests that CESA disruption triggers activation of compensatory metabolic and defense mechanisms at the protein domain level.



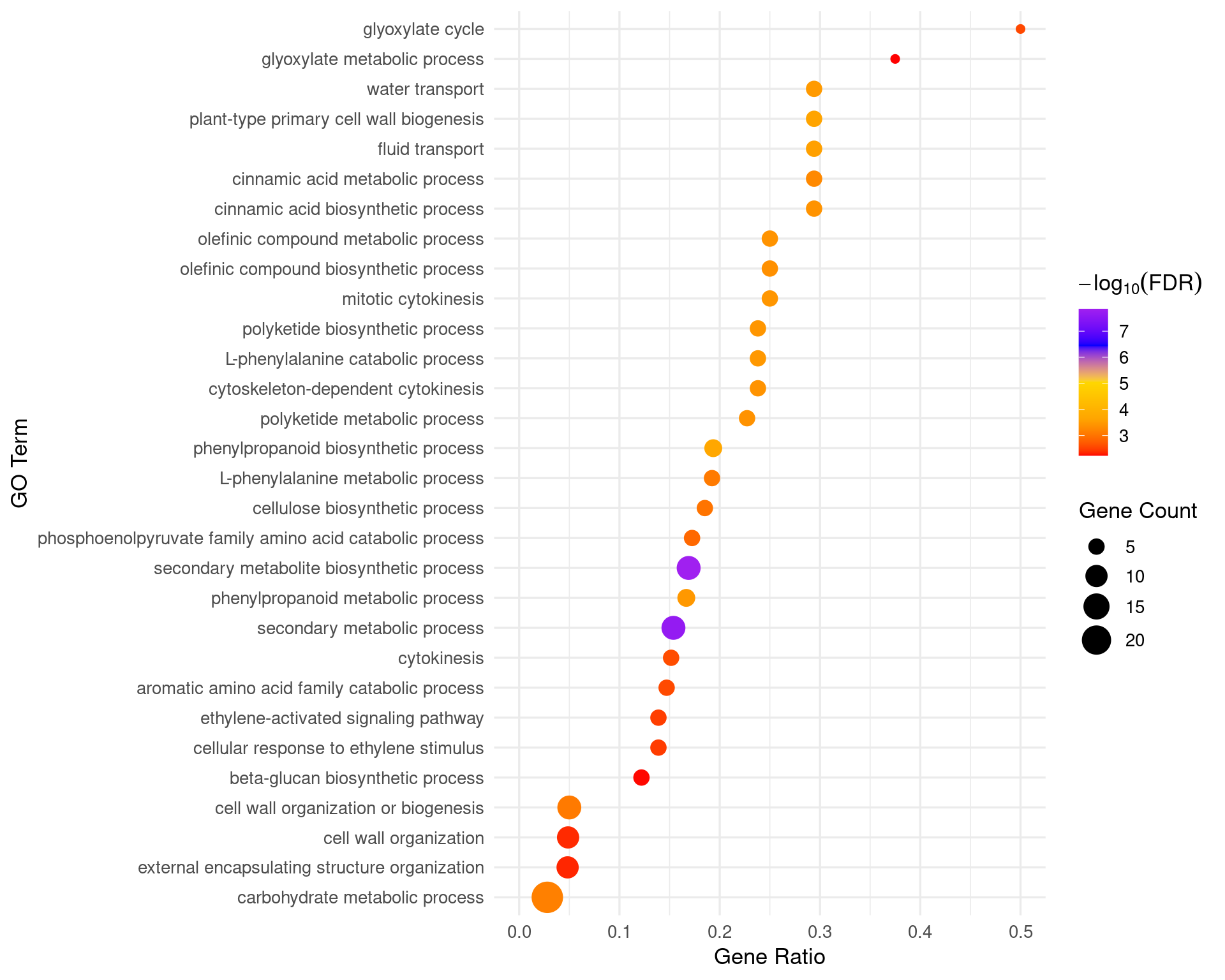
**Figure 1.** Secondary metabolism, photosynthesis, and stress-response domains are significantly enriched in CESA knockout DEGs. Pfam domains were assessed for enrichment in DEGs from *P. patens* CESA knockout lines (n = 3 biological replicates) relative to the genome background using Fisher’s exact test with false discovery rate (FDR) correction. Only domains with FDR < 0.05 are shown. Four Pfam domains were significantly enriched, including domains associated with secondary metabolism (chalcone/stilbene synthase), photosynthesis (ribulose bisphosphate carboxylase), and stress responses (cytochrome P450).

To gain a broader understanding of functional classifications, Gene Ontology (GO) enrichment analysis was performed across biological process (BP), molecular function (MF), and cellular component (CC) categories. The top 15 enriched biological processes were visualized by fold enrichment, and pie charts summarized the distribution of the most frequent GO terms within each category (Figure 2). Processes related to secondary metabolism, cytokinesis, water transport, and cell wall organization were particularly enriched among DEGs. Molecular functions were dominated by glycosyltransferase activities and water channel transporters, and cellular components were enriched for intracellular organelles and membrane structures.



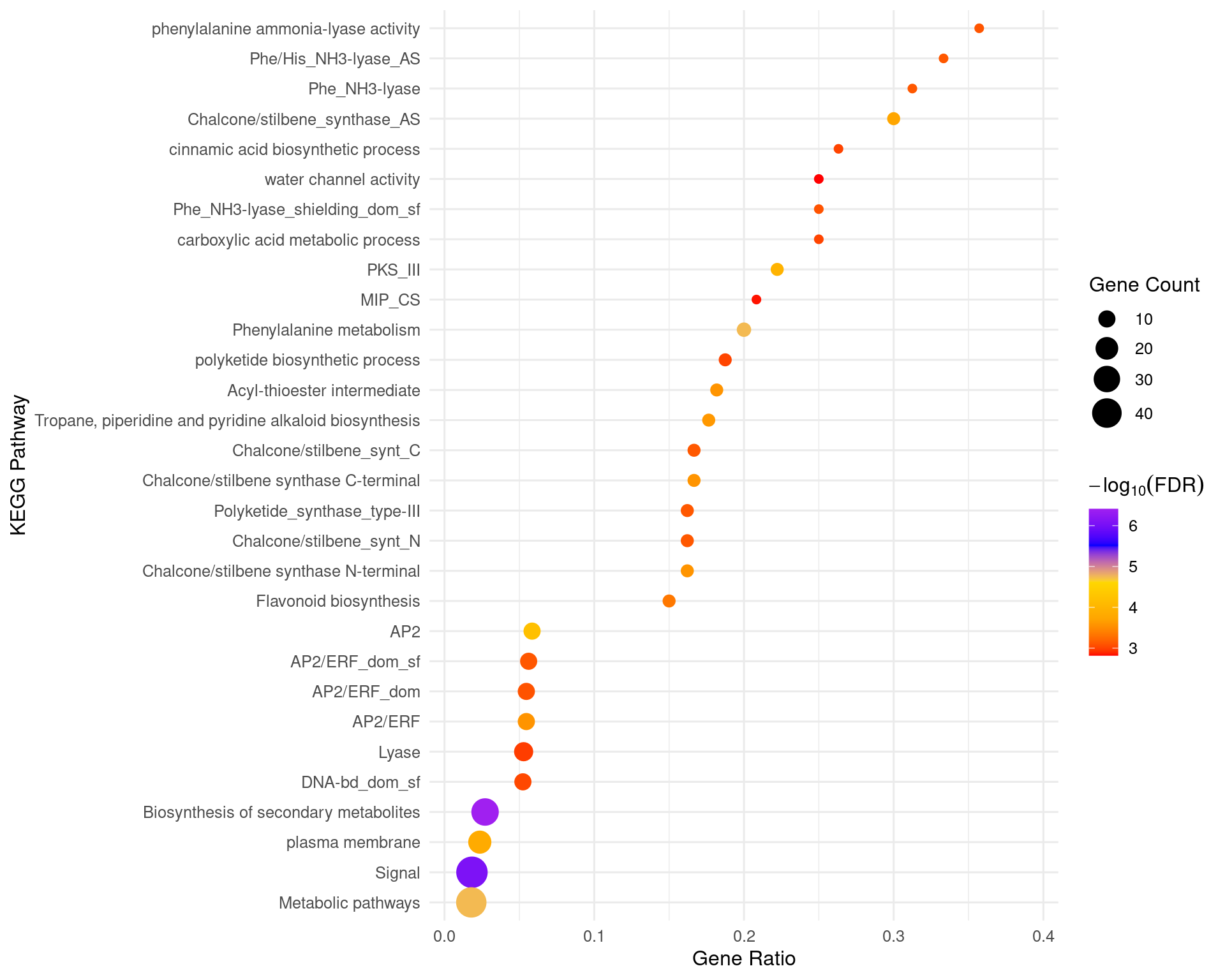
**Figure 2.** CESA knockout DEGs are enriched in secondary metabolism, transport, and cell wall biogenesis processes and localize to intracellular organelles.GO enrichment analysis was performed on DEGs from *P. patens* CESA knockout lines (n = 3 biological replicates) using the PANTHER classification system. A) The bar plot displays the top 15 enriched biological processes (BP) based on fold enrichment, filtered for FDR < 0.05. B–D) Pie charts summarize the distribution of enriched cellular components (CC), molecular functions (MF), and biological processes (BP). A) Biological processes were enriched for secondary metabolism, cytoskeleton-dependent cytokinesis, water transport, and cell wall biogenesis. B) Cellular components were predominantly intracellular organelles and membrane-bound structures. C) Molecular functions included glycosyltransferase activity and water transport-related activities. D) Biological process distribution confirmed the dominance of secondary metabolic and cell wall-related processes, highlighting the metabolic and structural reprogramming triggered by CESA loss.

Statistical enrichment of biological processes was further examined using a GO dot plot (Figure 3). This analysis confirmed the enrichment of secondary metabolism, cell wall biogenesis, ethylene signaling, and cytoskeletal organization processes among CESA knockout DEGs. The visualization also showed that the most significant biological processes had both high fold enrichment and strong statistical support, reinforcing the trend toward metabolic and structural remodeling.



**Figure 3.** Secondary metabolism, cytokinesis, and cell wall organization processes are enriched in CesA knockout DEGs. GO enrichment analysis of biological processes (BP) was performed on DEGs from *P. patens* CESA knockout lines (n = 3 biological replicates) using the PANTHER classification system. The top 30 enriched GO terms were plotted based on gene ratio (DEG count/reference count). Statistical significance was assessed via FDR-adjusted p-values, and -log10(FDR) was used to color the dots. Enriched processes include secondary metabolic pathways, cytokinesis, ethylene signaling, and cell wall biogenesis, reflecting metabolic reprogramming and structural adaptation to CESA loss.

Finally, KEGG pathway enrichment analysis revealed key metabolic pathways impacted by CESA disruption (Figure 4). The top enriched pathways included phenylpropanoid biosynthesis, polyketide biosynthesis, and secondary metabolite biosynthesis. These pathways are frequently associated with stress adaptation and structural reinforcement in plants, supporting the conclusion that CESA loss leads to broad transcriptional reprogramming to compensate for disruptions in cellulose biosynthesis.



**Figure 4.** CESA knockout DEGs are enriched in secondary metabolite biosynthesis, phenylalanine metabolism, and cell wall-related pathways. KEGG pathway enrichment analysis was performed on differentially expressed genes (DEGs) from CESA knockout lines (n = 3 biological replicates) using DAVID functional annotation. Pathways with Benjamini-Hochberg adjusted p-values (FDR) < 0.05 were retained. The top 30 enriched pathways were plotted based on gene ratio (DEG count/reference population count). Enriched pathways included secondary metabolite biosynthesis, phenylpropanoid biosynthesis, polyketide biosynthesis, and cell wall biogenesis, alongside regulatory functions such as ethylene signaling and water transport.

**Conclusion**

Pfam domain enrichment, GO term analysis, and KEGG pathway analysis of CESA knockout DEGs in *Physcomitrium patens* revealed major shifts in metabolic processes, stress response pathways, and cell wall-related activities. These results suggest that loss of CESA function triggers transcriptional adaptation to reinforce cellular structures and activate compensatory metabolic programs. These findings highlight how plants respond at the molecular level to disruptions in cell wall biosynthesis and provide a foundation for future functional studies.