Transcriptomics of Myosin XI Conditional Loss-of-function in Moss

A Major Qualifying Project
Submitted to the Faculty of
WORCESTER POLYTECHNIC INSTITUTE
In partial fulfillment of the requirements for the
Degree of Bachelor of Science

By: Carter Nakagawa
Bioinformatics and Computational Biology (BCB)
Report Submitted to: Prof. Luis Vidali

Date: April 23, 2024

Abstract

Plant biology is a key research topic due to the variety of functions plants perform for our society and in nature. Present dangers to plants including global warming and pollution may have ecological and economic consequences if understanding of plant biology is insufficient to solve such problems. Cellular mechanisms of growth are one such topic that is vital to understanding factors that affect plant health. With discoveries and advancements made in recent decades, the potential of this research has only increased. Model organisms provide a convenient framework for experimental plant research, while technologies such as next generation sequencing (NGS) have furthered our ability to identify the genomic properties underlying plant biology. This project leverages these developments and builds on previous research to investigate the role of Myosin XI in tip growth, with the moss *Physcomitrium patens* as a model. Myosin XI has already been implicated as a key factor in *P. patens* tip growth, with defects in Myosin XI resulting in lack of growth and cell death in developing moss. However, the exact mechanisms by which Myosin XI enables tip growth are not fully understood. We hypothesized that Myosin XI is involved in numerous pathways related to cell growth and survival. To confirm this, we analyzed RNAseq data from a temperature-sensitive Myosin XI moss line. Our results confirm widespread transcriptome remodeling as a result of Myosin XI loss-of-function and identify several potential key genes which may coordinate with functional Myosin XI to ensure proper tip growth in *P. patens*.

Table of Contents

Table of Figures

Table of Tables

4

1. Introduction

Plants are a vital resource for our society, playing a central role in agriculture and the production of natural resources. Currently, growing populations necessitate further consumption of plant material, motivating new approaches such as genetic modification to optimize production. Additionally, global climate change and environmental damage can negatively impact the health of plant populations and harm ecosystems that depend on them. As such, understanding the mechanisms of plant growth is a key topic of interest. Recent advances have allowed study of this field at a cellular and genetic level, enabling the exploration of many fundamental factors in plant growth. This project approaches the topic of plant growth by looking at the role of Myosin XI in moss tip growth, using the model moss *Physcomitrium patens*.

Another key factor in this project is the role of heat response in plant cellular biology. We utilize a temperature-sensitive (TS) Myosin XI mutant to conditionally induce Myosin XI loss-of-function. As such, another goal of this project will be to identify known plant heat responses in order to separately identify effects of Myosin XI loss-of-function. The mechanisms of heat shock response in plants, including *P. patens* have been researched and documented in previous literature (Elzanati et al., 2020). We aim to confirm the specific role of Myosin X in tip growth independent of heat response by comparing gene regulation in WT and TS strains. Understanding the genomic and transcriptomic changes that produce this heat sensitive phenotype will help to identify key genes involved tip growth in *P. patens*.

Investigation of Myosin XI loss-of-function will be done by analyzing the transcriptome and the changes that it undergoes in conditional loss-of-function. RNA-

seq data was obtained through past experiments comparing wild-type and mutant plant samples at 20°C and 32°C. The goal is to establish what genes are affected by the loss of Myosin XI and how this prevents growth and cell survival. Functional analysis of these genes will contextualize these results, which may be compared to current hypotheses surrounding the mechanisms of tip growth in plants.

2. Background/Literature Review

2.1 The Model Organism Physcomitrium patens

The moss *P. patens* is a well-studied model organism, useful in a wide range of topics. Research on the moss first began decades ago, with initial studies investigating plant biology using mutant *P. patens* strains (Engel, 1968). With advances in genomics from the past few decades, the potential of this organism for plant research has only increased. The reasons for this moss's popularity as a model organism are varied.

Firstly, as a non-vascular plant, *P. patens* fills an important niche in plant research. While there are numerous vascular model plants, including the highly popular *Arabidopsis thaliana*, there are far fewer commonly used non-vascular models. The divide between vascular and nonvascular plants stands as an important evolutionary and physiological difference. The lack of vasculature has many implications in plant biology, including the lack of true roots, stems, and leaves, as well as different methods of nutrient transport. Though not unique to non-vascular plants, the lack of seeds and use of spores instead is another important difference between *P. patens* and many model seed plants. The study of non-vascular plants is also important for understanding plant evolution, as the split between vascular and nonvascular plants occurred early on.

As such, comparison between vascular and nonvascular species can reveal potential characteristics of early plants. These differences necessitate study of nonvascular plants to understand the full scope of plant biology.

There are also many traits of *P. patens* that make it a convenient organism to study. One of its most important features is its ability to integrate exogenous DNA through homologous recombination, allowing researchers to achieve transformation efficiency on par with yeast, much higher than most other plant (Schaefer & Zryd, 1997). This allows simple creation of knockout mutants for targeted genetic studies. In addition to this, *P. patens* is a generally simple organism, consisting of few tissue types and spending most of its life cycle as a haploid gametophyte (Rensing et al., 2020). These various factors make *P. patens* a simple and flexible model organism.

The significant volume of research on *P. patens* has resulted in a wealth of data on the organism, further increasing its value as a model organism. Of particular relevance to this project are the genomic and transcriptomic data sources for the moss, which can be used to inform and support this analysis. One such collection is PEATmoss, an online gene expression atlas for *P. patens* specifically (Fernandez-Pozo, 2023). Such resources can be used to compare new findings to past data to check for consistency or novel results.

2.2 Heat Response in *P. patens*

The heat response in *P. patens* has been well-documented in previous research. Past work on the transcriptomic heat response has shown it to be complex and multifaceted (Elzanati et al., 2020). At high temperature, expression of hundreds of different genes is affected, with many of those genes relating to different cellular

processes. The degree of the response has also been shown to be proportionate to the severity of heat stress, as the same heat-related genes are even more differentially expressed in high stress at 37°C compared to moderate stress at 30°C (Elzanati et al., 2020). Duration of heat stress has also been identified as a key factor, with different pathways activating after a long period of heat stress despite not being involved in the initial response. Such findings show that there are a variety of factors yet to be fully explored in the heat response of *P. patens*.

2.3 P. patens Development and Tip Growth

The cellular processes behind *P. patens* tip growth are well-studied but remain an ongoing research topic. Of particular interest is the protonema stage of development, which follows germination. At this stage, the moss is still in a haploid state. As the moss develops, protonemata undergo apical extension, causing the cells themselves to elongate (Bibeau et al., 2021). It is this process specifically that is of interest to this project.

Previous work studying *P. patens* protonemata has identified some processes in tip growth (Bibeau et al., 2021). Vesicles are essential to this process, transporting cell wall components and wall-loosening enzymes to the growing tip of the cell. Cytoskeletal elements such as F-actin and Myosin XI have been implicated in the transport and concentration of these vesicles (Bibeau et al., 2018). This suggests a complex response involving cytoskeletal movement of vesicles and regulation of vesicle fusion events. However, the specifics of these processes require further investigation, with current understanding relying somewhat on theoretical models.

2.4 Myosin XIa Temperature Sensitive Line

The mutant strain used in this project possesses a temperature sensitive allele of Myosin XI. Myosin XI is an actin-based motor protein implicated in polarized cell growth in plants (Galotto et al., 2021). There are two highly similar Myosin XI genes, Myosin XIa and Myosin XIb. This mutant moss line has a mutated Myosin XIa gene with point mutations causing instability in high heat conditions. As Myosin XIb is functionally redundant, this line also has a Myosin XIb knockout (Vidali et al., 2010) (Galotto et al., 2021).

In a previous study, myoXIaTS cells in high temperature (32°C) did not grow properly, with stunted development and minimal tip growth (Galotto et al., 2021). A majority of myoXIaTS apical cells died, which is unlikely to have been caused by temperature stress, as most WT cells survived. Additionally, F-actin, which Myosin XI associates with, was found to be involved in this process, as depolymerization of actin using LatB in TS cells prevented cell death. This demonstrated the importance of Myosin XI in plant tip growth. This project aims to further study the transcriptome of this mutant line to understand the mechanisms of temperature sensitivity.

2.5 Modular Cell Biology

Modular cell biology is a framework that considers cellular processes on a higher level than individual molecules and genes, instead grouping them by function into “modules” (Hartwell et al., 1999). As ongoing research reveals the complexity of biological systems, it becomes necessary to consider cellular processes in this more abstract way, as individual components become too numerous to keep track of separately. The concept of modules can be applied to understanding gene expression,

where expression patterns of many genes may correlate, suggesting these genes all contribute to some higher-level function. In fact, correlation network analysis of gene expression can be used to identify these functional modules (Langfelder & Horvath, 2008).

For this project, applying a modular framework to understanding heat response may be helpful. As established in previous work, the *P. patens* heat response is complex and multifaceted, involving a multitude of different genes and affecting a variety of separate cell processes (Elzanati et al., 2020). Identifying the key modules involved in the transcriptomic heat response may be useful in characterizing the cell's heat response at a higher level.

2.6 RNA-seq Analysis

A key technique in studying gene expression and regulation is RNA-seq. This type of sequencing begins by taking RNA samples and converting them to cDNA, which can then be amplified and sequenced using next-generation sequencing (NGS) technologies (Wang et al., 2009). The resulting sequences can then be aligned to a species' transcriptome, which shows how many reads are associated with individual genes. The counts of these aligned reads can be used to quantify gene expression and can be compared across different experimental conditions. By comparing the per-gene counts under different conditions, it is possible to see how different factors affect gene expression. Genes with statistically significant differences in expression under different conditions are considered differentially expressed genes (DEGs) (Huang et al., 2015).

For this project, there are a number of comparisons to be made for which genes may be considered differentially expressed. Firstly, comparing gene expression at

different temperature conditions identifies key DEGs involved in heat response. This comparison may be conducted for both the wild-type and mutant strains, in order to establish the heat response in each. However, it is also possible to analyze differentially expressed genes between the wild-type and mutant strains. This is important, as it allows identification of baseline differences in gene expression as well as differences that only are seen in high temperature conditions.

3. Methods

3.1 RNA-seq Experiment

Data used in this project was obtained from previous RNA-seq experiments. WT and TS protoplasts were regenerated for three days at 25°C, then transferred to growth media at 20°C for four days. Following this, protoplasts were transferred to media promoting caulonemata differentiation at 20°C for another four days. Lastly, cells were incubated at either 20°C or 32°C (restrictive temperature for TS line) for two hours before collection and flash freezing in liquid nitrogen. The experiment was performed in triplicate for statistical confidence. The RNeasy Plant Mini Kit (Quiagen) was used to extract RNA, which was then sequenced at the GENEWIZ facility. A HiSeq4000 sequencer was used with libraries prepared with the NEBNext Ultra RNA Library Prep Kit. 27-39 million reads were sequenced per sample.

3.2 Read Processing

Following the RNA-seq experiment, reads were first checked with the FastQC tool for sequence quality and length metrics (Andrews, 2010). Trimmomatic was then used to remove adapters and low-quality sequences using the ILLUMINACLIP and

SLIDINGWINDOW functions respectively with default options (Bolger et al., 2014). The tool HISAT2 was used to map reads to the *Physcomitrium patens* genome V3.3, with paired-end options --fr --no-discordant (Kim, 2019). SAM files obtained from this step were then used to generate count matrices using htseq-count (Anders et al., 2015). Genome and feature data for *Physcomitrium patens* V3.3 used in these steps was obtained from NCBI. All steps were performed via the Galaxy online platform (Galaxy, 2022).

3.3 Differentially Expressed Genes (DEGs)

Count data obtained from the previous workflow was then used to analyze differential expression across conditions. This analysis was performed in R using the DESeq2 package (Love et al., 2014). First, PCA analysis was done to assess large-scale sample differences. Following that, differential expression analysis was then performed with comparisons WT 20°C vs WT 32°C, TS 20°C vs TS 32°C, WT 20°C vs TS 20°C, and WT 32°C vs TS 32°C. As DESeq2 internally normalizes counts by library size, no additional normalization steps were needed, though genes with low counts are also filtered out with default settings. The main comparisons of interest were the WT vs TS comparisons at the two temperature conditions, as these comparisons would show expression differences between strains. However, the other two comparisons were included to assess the general temperature response for each strain individually. Plots illustrating gene expression across the transcriptome were made using DESeq2 functions.

Expression was also analyzed at the level of individual genes, in order to identify key genes of interest. DESeq2 provides multiple metrics for quantifying gene

expression, including $\log_2$ fold change (log2FC) and both raw and adjusted p-values.

The adjustment method implemented in DESeq2 is an adaptation of Benjamini-Hochberg false discovery rate correction. It is intended to limit the false discovery rate to a cutoff of 5%. Both log2FC and p-values are considered at a per-gene level, to assess the degree of differential expression and statistical confidence. Genes with log2FC > 1 or log2FC < -1 were considered differentially expressed.

3.4 Gene Annotation

Annotations were first obtained from UniProt, including gene ontology (GO) terms along with UniProt ID's (UniProt, 2023). More specific resources were also referenced for specific genes, including JGI's Phytozome for gene annotations (Goodstein et al., 2012). The online *P. patens* atlas PEATMoss was also referenced for expression data (Fernandez-Pozo, 2023). Annotations were used to identify genes of interest related to known pathways implicated in the TS line's loss of growth phenotype.

3.5 Gene Ontology (GO) Analysis

To identify key pathways involved in the differential response between strains, GO analysis was applied. The R package clusterprofiler was used for this step to identify enriched terms in DEGs for each comparison (Wu et al., 2021). As with the initial differential expression analysis, the main comparison of interest was WT vs TS at 32°C. The comparisons between the same strains at different temperatures were also important to investigate the overall heat response of each. Once significantly enriched terms ($p < 0.05$) were identified, enrichment was visualized using clusterprofiler

functions. Significantly enriched modules were investigated further to identify individual genes of interest.

4. Results

4.1 Differential Expression by Condition

Initial differential expression analysis revealed numerous DEGs across all conditions (Figure 1). Significantly over-expressed and under-expressed genes were seen in all comparisons. However, the magnitude of overall differential expression varied across conditions. Notably, though DEGs were still found in the “TS vs WT at 20°C” comparison, the total number of DEGs was less compared to all other conditions (Figure 1C). Compared to the 1037, 1154, and 428 DEGs found in the “WT 32°C vs WT 20°C”, “TS 32°C vs TS 20°C”, and “TS 32°C vs WT 32°C” comparisons respectively, only 112 DEGs were found in the “TS 20°C vs WT 20°C” comparison. This supports the assumption that while there is some baseline difference in gene expression between strains, the two strains are largely similar below the restrictive temperature.

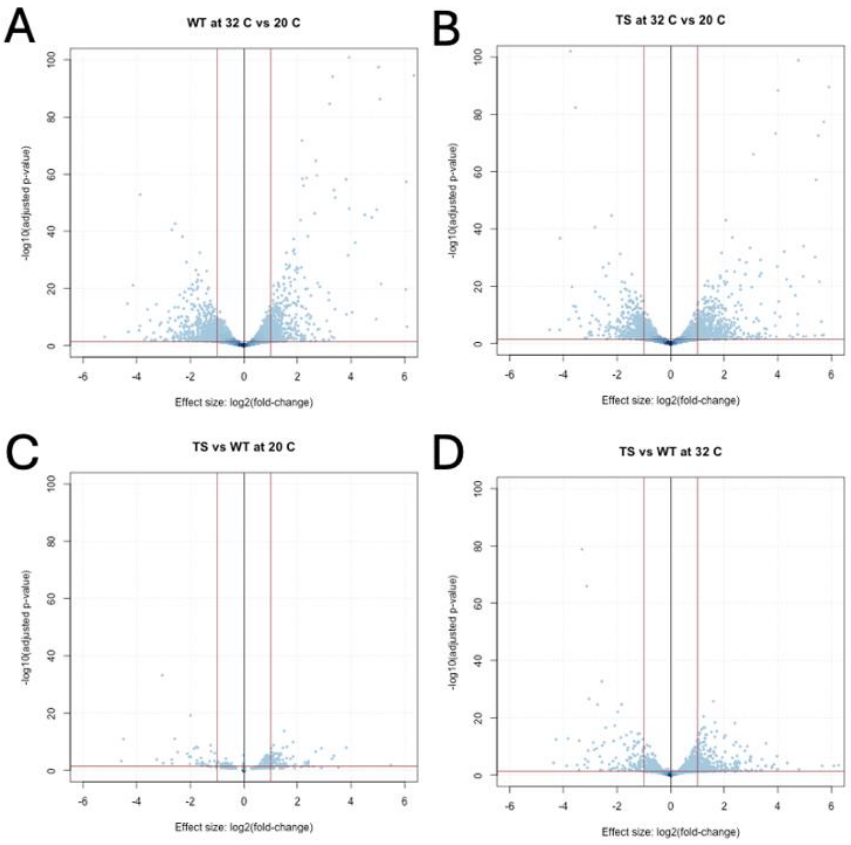

Figure 1: Volcano plots for conditions (a) WT at 32°C vs WT at 20°C (b) TS at 32°C vs TS 20°C (c) TS at 20°C vs WT at 20°C and (d) TS at 32°C vs WT at 32 °C. Cutoffs for DEG set at padj < 0.05 and |log2FC| > 1.

DEGs were found across a wide range of raw counts for all conditions (Figure 2). This confirmed that differential expression occurs both in highly expressed genes and those with minimal expression overall. Overall trends in DEGs across the distribution of normalized counts were somewhat similar as expected, with more highly differentially expressed genes at lower normalized counts, regardless of total DEGs in each comparison. However, comparisons between strains (Figure 2C,D) exhibited less highly differentially expressed genes with high counts compared to the temperature change conditions (Figure 2A,B).

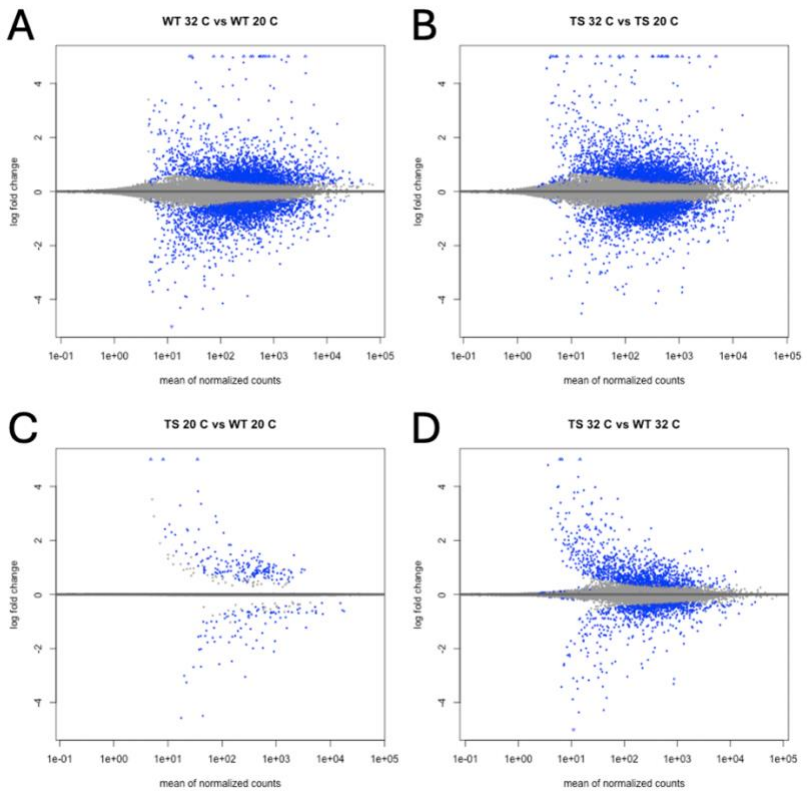

Figure 2: MA plots for conditions (a) WT at 32°C vs WT at 20°C (b) TS at 32°C vs TS 20°C (c) TS at 20°C vs WT at 20°C and (d) TS at 32°C vs WT at 32 °C. Blue points represent genes with padj < 0.05.

4.2 Highly Differentially Expression between 32°C Conditions

To identify genes that may contribute significantly to the TS mutant's loss of growth phenotype, we selected genes with highly differential expression between TS and WT lines at the restrictive temperature of 32°C. Genes were first filtered by raw counts, such that genes without >50 counts in all replicates of at least one condition were not considered. While a cutoff of >10 counts is generally sufficient for filtering out insignificant genes, we wanted to focus on genes with higher expression in general. The adjusted p-value (p_{adj}) was also used to filter out genes without significantly differential expression, which we considered to be those with $p_{adj} \geq 0.05$. After filtering, genes were

ordered by $\log_2$ fold change (LFC) to identify the top 10 most differentially under-expressed and over-expressed genes.

With this approach, we identified multiple highly differential DEGs that may be relevant to phenotypic strain differences. Many of these genes are only differentially expressed in the “TS 32°C vs TS 20°C” and “TS 32°C vs WT 32°C” conditions, suggesting they are uniquely important to the TS strain and not the WT (Figure 3). Additionally, several of these highly differentially regulated DEGs were found to be uncharacterized or poorly characterized (Table 1,2).

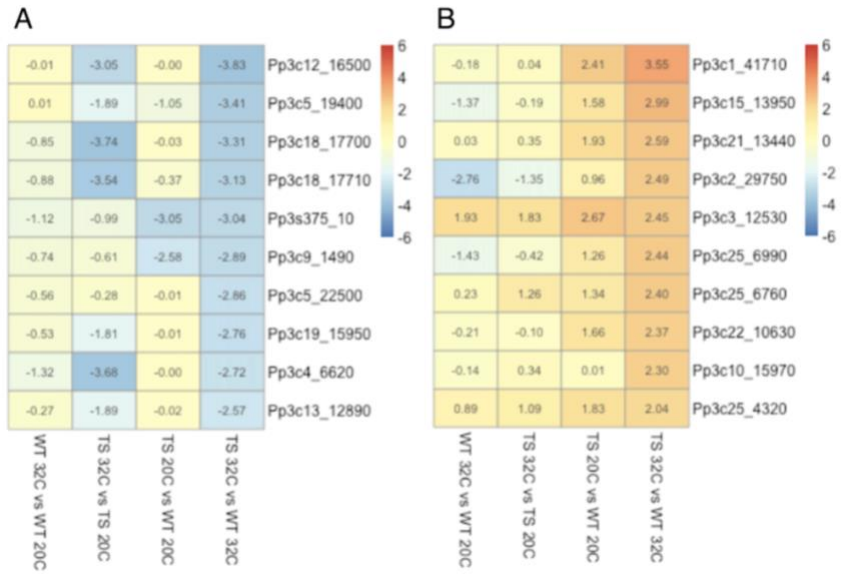

Figure 3: Log fold change heatmaps for the top 10 (a) downregulated and (b) upregulated genes in the TS 32°C vs WT 32°C condition.

Gene ID	Annotation
Pp3c4_6620	Uncharacterized protein
Pp3c5_19400	Uncharacterized protein
Pp3c5_22500	Uncharacterized protein
Pp3c9_1490	Pyruvate kinase (EC 2.7.1.40)

Pp3c12_16500	Peroxidase (EC 1.11.1.7)
Pp3c13_12890	Uncharacterized protein
Pp3c18_17700	Uncharacterized protein
Pp3c18_17710	Uncharacterized protein
Pp3c19_15950	Oxysterol-binding protein
Pp3s375_10	Pyruvate kinase (EC 2.7.1.40)

Table 1: Annotations for the top 10 under-expressed genes in the TS vs WT 32°C comparison, ordered to match Figure 3A.

Gene ID	**Annotation**
Pp3c1_41710	Hexosyltransferase (EC 2.4.1.-)
Pp3c2_29750	Uncharacterized protein
Pp3c3_12530	Ribulose bisphosphate carboxylase small subunit, chloroplastic (RuBisCO small subunit)
Pp3c10_15970	TIR domain-containing protein
Pp3c15_13950	Peroxidase (EC 1.11.1.7)
Pp3c21_13440	Uncharacterized protein
Pp3c22_10630	Protein kinase domain-containing protein
Pp3c25_4320	Steroid 5-alpha reductase C-terminal domain-containing protein
Pp3c25_6760	VQ domain-containing protein
Pp3c25_6990	Uncharacterized protein

Table 2: Annotations for the top 10 over-expressed genes in the TS vs WT 32°C comparison, ordered to match Figure 3B

4.4 Comparing WT and TS Heat Response

Firstly, though we are mainly interested in the effect of temperature on the TS mutant's Myosin XI, it was also necessary to characterize the heat response overall.

Understanding what aspects of the moss heat response is active at the restrictive temperature will help us better understand what changes are directly caused by loss of Myosin XI function. To do this, we applied GO enrichment analysis to identify key processes enriched with DEGs, specifically those from the “TS 32°C vs TS 20°C” and “WT 32°C vs WT 20°C” comparisons.

Gene ontology analysis showed similar enriched modules in both strains in response to temperature (Figure 4). This suggests that on a large scale, cells of both strains are behaving similarly in response to heat. Additionally, the lack of differences informs our assumptions about the expression-level cause of the TS loss-of-growth phenotype. Since we do not see significantly enriched processes unique to the TS strain’s heat response, it may be that the phenotype is caused by a small subset of key genes or that the relevant genes are spread across many processes.

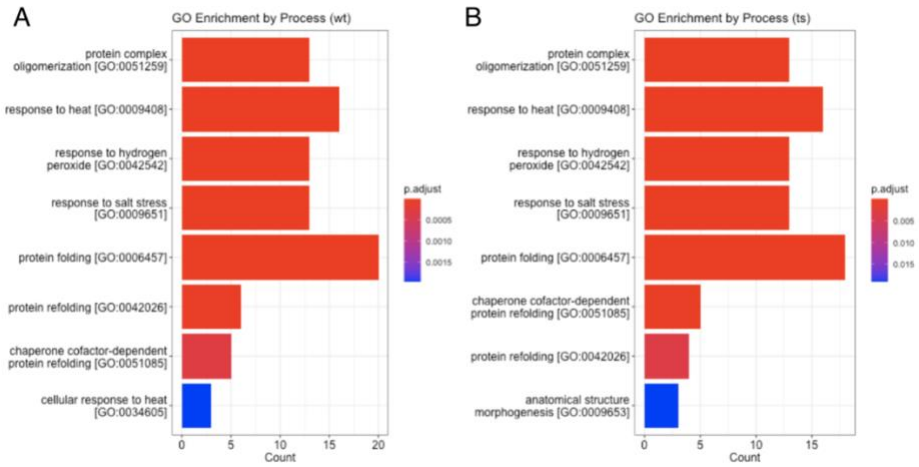

Figure 4: GO enrichment plots showing enriched processes as defined by GO terms in the (a) “WT 32°C vs WT 20°C” and (b) “TS 32°C vs TS 20°C” comparisons.

Alternatively, the enriched processes identified through GO terms may contain the genes responsible for the loss-of-growth phenotype. However, these genes may be expressed differently between strains, despite being differentially expressed in both. To

determine whether this is the case, we looked at expression of DEGs labeled with these enriched GO terms, starting with process categories.

The vast majority of DEGs annotated with enriched GO terms showed minimal expression differences between strains (Figure 5). However, four genes, all grouped under “anatomical structure morphogenesis” were found to be differentially expressed between WT and TS lines (Figure 5C). Notably, this module was only found to be enriched with DEGs in the TS line and not in the WT line when comparing temperature conditions. These genes were also all found to be expansins and beta-expansins, suggesting that expansins may play a key role in the TS loss-of-growth phenotype (Table 5).

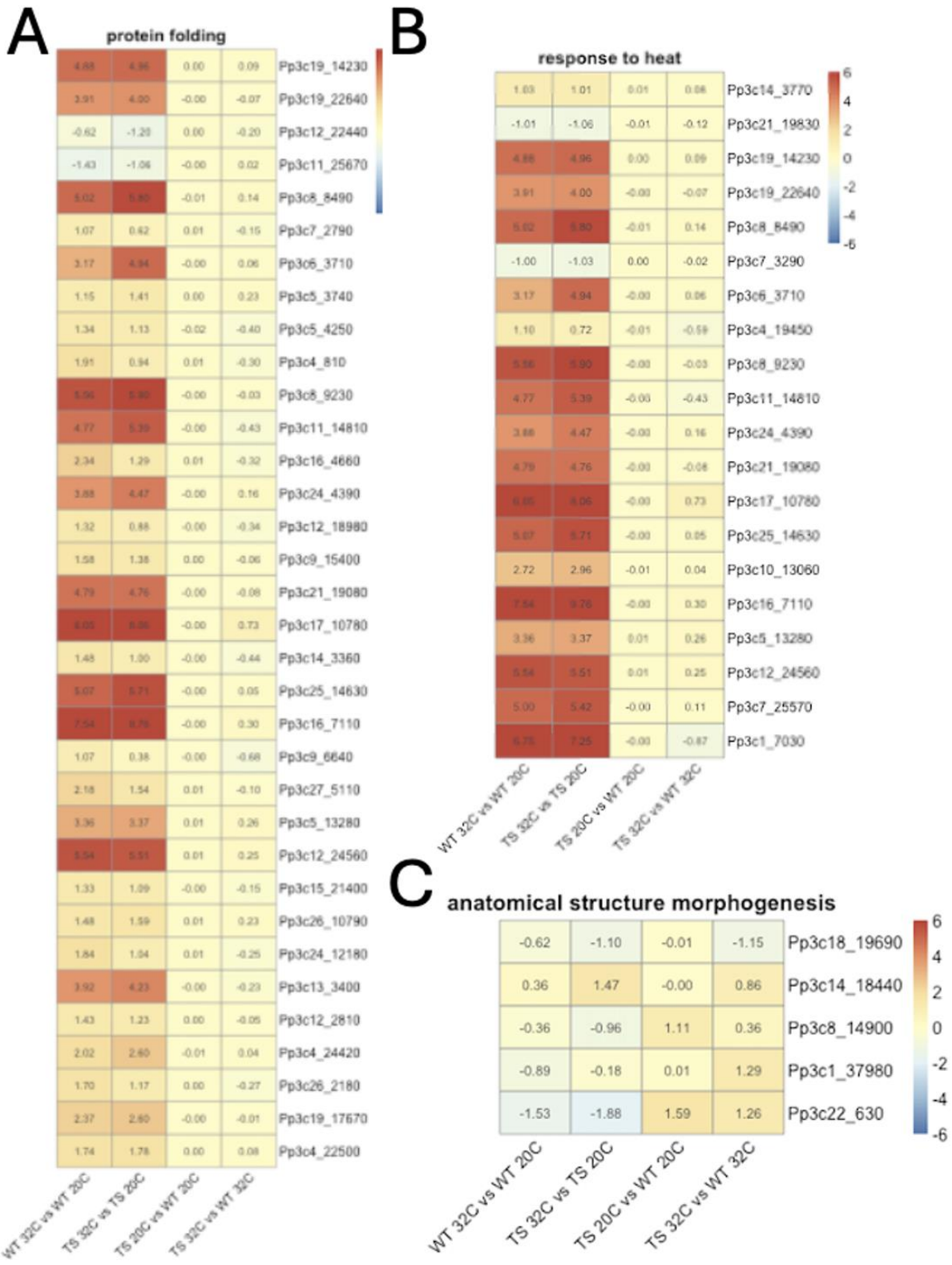

Figure 5: Log fold change heatmaps for key enriched heat response GO terms (a) protein folding (b) response to heat and (c) anatomical structure morphogenesis

Gene ID	Annotation
Pp3c19_14230	SHSP domain-containing protein
Pp3c19_22640	SHSP domain-containing protein
Pp3c12_22440	Histidine kinase/HSP90-like ATPase domain-containing protein
Pp3c11_25670	Peptidyl-prolyl cis-trans isomerase (PPIase) (EC 5.2.1.8)
Pp3c8_8490	SHSP domain-containing protein
Pp3c7_2790	Peptidyl-prolyl cis-trans isomerase (PPIase) (EC 5.2.1.8)
Pp3c6_3710	SHSP domain-containing protein
Pp3c5_3740	DNL-type domain-containing protein
Pp3c5_4250	Activator of Hsp90 ATPase AHSA1-like N-terminal domain-containing protein
Pp3c4_810	Histidine kinase/HSP90-like ATPase domain-containing protein
Pp3c8_9230	SHSP domain-containing protein
Pp3c11_14810	SHSP domain-containing protein
Pp3c16_4660	RuBisCO large subunit-binding protein subunit beta, chloroplastic
Pp3c24_4390	SHSP domain-containing protein
Pp3c12_18980	PPIase cyclophilin-type domain-containing protein
Pp3c9_15400	Activator of Hsp90 ATPase AHSA1-like N-terminal domain-containing protein
Pp3c21_19080	SHSP domain-containing protein
Pp3c17_10780	SHSP domain-containing protein
Pp3c14_3360	Histidine kinase/HSP90-like ATPase domain-containing protein
Pp3c25_14630	SHSP domain-containing protein
Pp3c16_7110	SHSP domain-containing protein
Pp3c9_6640	Histidine kinase/HSP90-like ATPase domain-containing protein
Pp3c27_5110	RuBisCO large subunit-binding protein subunit beta, chloroplastic

Pp3c5_13280	SHSP domain-containing protein
Pp3c12_24560	SHSP domain-containing protein
Pp3c15_21400	Activator of Hsp90 ATPase AHSA1-like N-terminal domain-containing protein
Pp3c26_10790	BAG domain-containing protein
Pp3c24_12180	RuBisCO large subunit-binding protein subunit alpha, chloroplastic
Pp3c13_3400	BAG domain-containing protein
Pp3c12_2810	J domain-containing protein
Pp3c4_24420	BAG domain-containing protein
Pp3c26_2180	Uncharacterized protein
Pp3c19_17670	BAG domain-containing protein
Pp3c4_22500	Uncharacterized protein

Table 3: Annotations for genes associated with ‘protein folding’ GO term.

Gene ID	Annotation
Pp3c14_3770	Ornithine aminotransferase (EC 2.6.1.13)
Pp3c21_19830	DnaJ-like protein
Pp3c19_14230	SHSP domain-containing protein
Pp3c19_22640	SHSP domain-containing protein
Pp3c8_8490	SHSP domain-containing protein
Pp3c7_3290	Uncharacterized protein
Pp3c6_3710	SHSP domain-containing protein
Pp3c4_19450	Saposin B-type domain-containing protein
Pp3c8_9230	SHSP domain-containing protein
Pp3c11_14810	SHSP domain-containing protein

Pp3c24_4390	SHSP domain-containing protein
Pp3c21_19080	SHSP domain-containing protein
Pp3c17_10780	SHSP domain-containing protein
Pp3c25_14630	SHSP domain-containing protein
Pp3c10_13060	HTH cro/C1-type domain-containing protein
Pp3c16_7110	SHSP domain-containing protein
Pp3c5_13280	SHSP domain-containing protein
Pp3c12_24560	SHSP domain-containing protein
Pp3c7_25570	SHSP domain-containing protein
Pp3c1_7030	SHSP domain-containing protein

Table 4: Annotations for genes associated with ‘heat response’ GO term.

Gene ID	Annotation
Pp3c18_19690	Expansin
Pp3c14_18440	Expansin
Pp3c8_14900	Expansin
Pp3c1_37980	Beta-expansin 2
Pp3c22_630	Beta-expansin 1

Table 5: Annotations for genes associated with ‘anatomical structure morphogenesis’ GO term.

We also performed the same GO enrichment analysis using functional Gene Ontology categories instead of process categories. This was to ensure we didn’t miss any key genes, as functional and process categories are generally similar but do not overlap completely.

Analysis based on functional categories identified five more genes with differential expression in the TS compared to the WT (Figure 6). However, as compared to the genes found through GO process enrichment analysis, which all were from the same category, these five genes are more varied, being spread across four categories (Table 3). We will consider these genes as genes of interest, particularly those with differential expression between WT and TS lines only at the restrictive temperature.

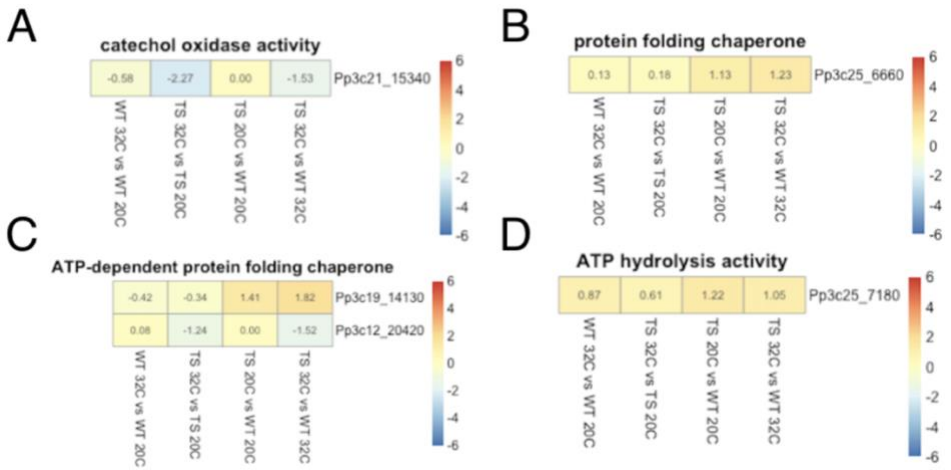

Figure 6: Log fold change heatmaps for DEGs associated with enriched functional GO terms (a) catechol oxidase activity (b) protein folding chaperone (c) ATP-dependent protein folding chaperone and (d) ATP hydrolysis activity found by comparing 32°C and 20°C conditions.

Gene ID	Annotation	GO Functional Term
Pp3c21_15340	Tyrosinase copper-binding domain-containing protein	Catechol oxidase activity
Pp3c25_6660	Chaperonin-like RBCX protein 1, chloroplastic	Protein folding chaperone
Pp3c19_14130	ALPHA KINASE/ELONGATION FACTOR 2 KINASE	ATP-dependent protein folding chaperone
Pp3c12_20420	ALPHA KINASE/ELONGATION FACTOR 2 KINASE	ATP-dependent protein folding chaperone
Pp3c25_7180	Clp R domain-containing protein	ATP hydrolysis activity

Table 6: Genes from Figure 6 listed with annotations and functional categories.

In combination with the expansin genes previously identified, nine total genes were found from this GO enrichment analysis. Five of these genes also appear to be differentially expressed between lines only at the restrictive temperature, which may suggest they are related to loss of Myosin XI function and the TS line's stunted growth.

4.5 Direct TS and WT Comparison

In addition to comparing the respective heat responses of both lines, we also compared gene expression between lines directly. Firstly, we compared gene expression at the 20°C condition to investigate any baseline differences between strains in the absence of impacted Myosin XI function. While both lines should be similar at non-restrictive temperatures, observations of the mutant TS line at 20°C did suggest some phenotypic difference. We performed the same GO enrichment analysis as before, looking at both functional and process categories.

Analysis using process categories did not identify significant genes. Functional analysis did identify multiple enriched categories with significant, differentially expressed genes, though only two distinct genes were identified (Figure 7). Both genes are enzymes expressed significantly less in the TS strain compared to the WT (Table 7).

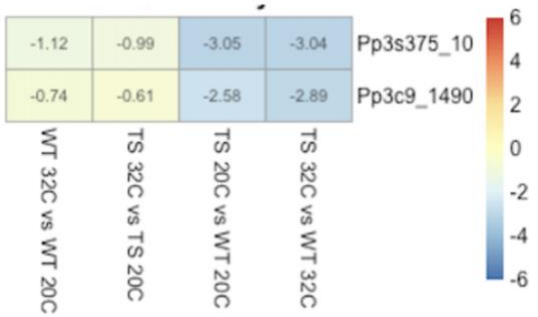

Figure 7: Log fold change heatmap for genes identified through GO enrichment analysis comparing TS and WT lines at 20°C.

Gene ID	Annotation
Pp3s375_10	Pyruvate kinase, cytosolic isozyme
Pp3c9_1490	Starch synthase

Table 7: Annotations for the two genes from DEG enriched modules at 20°C conditions.

Following GO enrichment analysis for the 20°C comparison, we performed the same analysis for the 32°C comparison. Using process categories, we identified eight more DEGs including three RuBisCo genes, two peroxidase genes, and three defense response genes (Figure 8, Table 8). Four of these genes (Pp3c10_15970, Pp3c15_13950, Pp3c12_16500, Pp3c3_12530) were also found in the top over-expressed and under-expressed genes for the “TS 32°C vs WT 32°C” comparison, further supporting their importance in this condition.

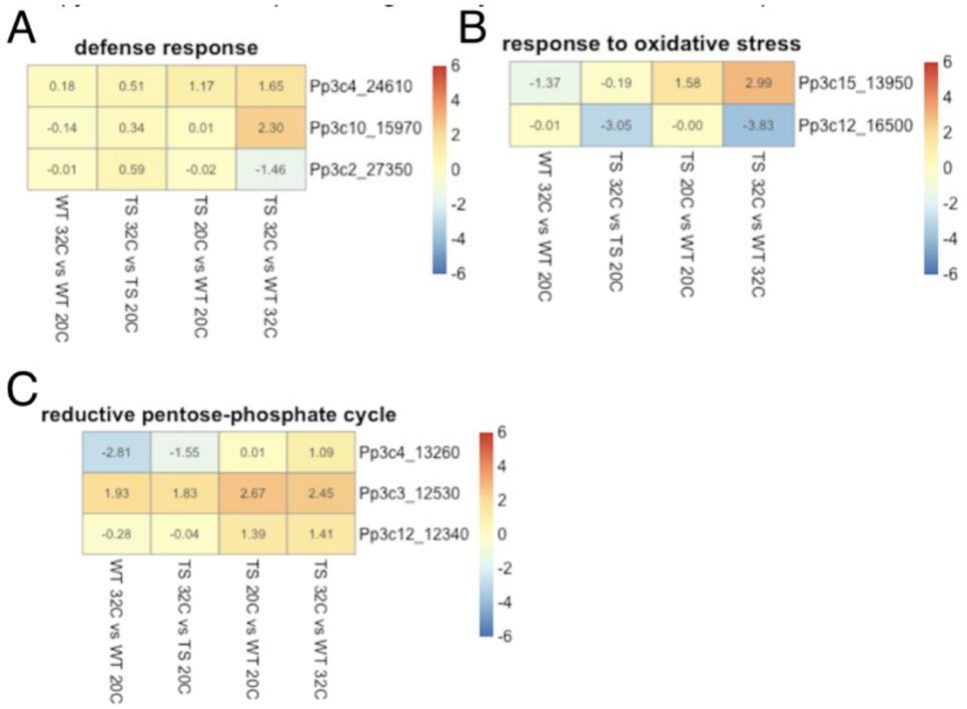

Figure 8: Log fold change heatmaps for DEGs in the 32°C comparison identified using process GO terms (a) defense response (b) response to oxidative stress and (c) reductive pentose-phosphate cycle.

Gene ID	Annotation	GO Process
Pp3c4_24610	(1 of 6) PTHR31942:SF15 - MLO-LIKE PROTEIN 11-RELATED	Defense response
Pp3c10_15970	(1 of 2) PTHR11017//PTHR11017:SF163 - LEUCINE-RICH REPEAT-CONTAINING PROTEIN	Defense response
Pp3c2_27350	(1 of 17) PF10604 - Polyketide cyclase / dehydrase and lipid transport (Polyketide_cyc2)	Defense response
Pp3c15_13950	(1 of 71) 1.11.1.7 - Peroxidase / Lactoperoxidase	Response to oxidative stress
Pp3c12_16500	(1 of 3) PTHR31235:SF11 - PEROXIDASE 21	Response to oxidative stress
Pp3c4_13260	(1 of 33) 4.1.1.39 - Ribulose-bisphosphate carboxylase / RuBP carboxylase	Reductive pentose-phosphate cycle
Pp3c3_12530	(1 of 33) 4.1.1.39 - Ribulose-bisphosphate carboxylase / RuBP carboxylase	Reductive pentose-phosphate cycle
pp3c12_12340	(1 of 33) 4.1.1.39 - Ribulose-bisphosphate carboxylase / RuBP carboxylase	Reductive pentose-phosphate cycle

Table 8: Genes from Figure 8 listed with annotations and GO process category.

Functional GO enrichment was also performed, but the vast majority of genes identified through functional categories were redundant with those from the previous process-based approach. In fact, only one additional gene was identified in this analysis, which was a Cytochrome P450 enzyme associated with heme binding.

4.6 Other DEGs of Interest

In addition to highly differentially expressed genes and genes related to enriched GO terms, all differentially expressed genes in the “TS 32°C vs WT 32°C” comparison

were examined to search for genes relevant to tip growth. This includes endocytosis/exocytosis genes like clathrin and exocyst components like Sec3. These genes may not be highly up or downregulated but may still have important effects based on their known relevance to tip growth.

ID	Protein	LFC (TS 32°C vs WT 32° C)
A0A2K1L8G0	Exocyst complex component Sec3 PIP2-binding N-terminal domain-containing protein	-1.56
A0A2K1JMM6	Clathrin light chain	1.40
A0A2K1IJG0	Xyloglucan endotransglucosylase/hydrolase	-1.14
A0A2K1KDV3	Xyloglucan endotransglucosylase/hydrolase	-1.14
A0A2K1KLU8	Tubulin alpha chain	-1.12
A0A2K1KLU8	Tubulin alpha chain	-1.11
A0A2K1L9B7	GTD-binding domain-containing protein	-1.15

Table 9: Additional genes of interest with annotations and LFC values

5. Discussion

In this project, we applied bioinformatic methods to understand the effects of Myosin XI loss-of-function on gene expression by analyzing RNA-seq data of a Myosin XI TS mutant. The results partially support our hypothesis that Myosin XI loss-of-function causes widespread gene expression changes that are important for cell survival and growth. Regulatory changes were observed in pathways related to stress response and cell wall remodeling. Patterns of regulatory change were diverse, with many genes from similar pathways differing in expression. Using the expansins and beta-expansins as an example, some were differentially expressed at 32°C while one was only differentially expressed at 20°C and another was differentially expressed at both temperatures. We also saw differential expression across many pathways, ranging from protein folding to oxidative stress response. This is in line with our understanding of Myosin XI loss resulting in regulatory changes across multiple pathways. Further work is necessary to investigate the role of individual genes in this phenomenon. Given the ease of genetic experiments in *P. patens*, future experiments using knockdown mutants for key genes identified here may prove useful.

It is also necessary to study individual genes without confounding factors from heat response. While these results conclusively identify many genes directly affected by Myosin XI loss, it is also clear from the expression patterns of key heat response genes that the increased temperature causes significant transcriptome remodeling. This is in line with the results of previous studies looking at mild heat stress over short periods (30°C for 1hr), in which enriched processes included response to heat, protein folding, and protein folding chaperoning (Elzanati et al., 2020).

Many of the genes identified in this study appear to have clear connections to tip growth, such as the expansins. From this study, we identified 3 expansin A and 2 expansin B genes that were differentially expressed between WT and TS lines. Previous work in *P. patens* has identified 27 expansin A and 7 expansin B genes, so further investigation is necessary to understand why only a handful of these genes were differentially regulated (Carey & Cosgrove, 2007). A brief review of regulation of other expansins in this dataset showed many of these genes are not highly expressed and were even filtered out for low counts. Whether this is due to flaws in mapping or genuinely low counts for other expansins is unclear.

Stress-related genes were another key category differentially expressed between WT and TS cells. Notably, salt stress genes were not differentially expressed between lines despite being generally upregulated in both WT and TS cells, which matches past work identifying salt stress response genes as a component of overall heat response (Elzanati et al., 2020). This further contextualizes the differential regulation of related stress responses, such as oxidative stress. Salt stress can trigger production of reactive oxygen species (ROS), which should in turn trigger oxidative stress response (Kesawat et al., 2023). In our study, we identified multiple peroxidase and lactoperoxidase genes that were differentially expressed between WT and TS cells that were also differentially expressed in response to heat. This may suggest that Myosin XI loss-of-function is causing differences in cellular response to oxidative stress, which may be a factor in cell survival. However, it is also possible that lack of cell growth is itself exacerbating accumulation of ROS, which could lead to differential oxidative stress response and cell death (Mansoor et al., 2022). Regardless, oxidative stress genes were among the most

highly differentially expressed genes, motivating future work on their role in Myosin XI loss-of-function.

Another surprising result was the lack of DEGs known to interact closely with actin. The initial study using the TS line found that depolymerization of F-actin prevented cell death, which suggested a role for F-actin in causing cell death when Myosin XI function is lost (Galotto et al., 2021). This was contrary to other research that found a connection between actin depolarization and plant cell death in other model plants, though this did not include mosses (Smertenko & Franklin-Tong, 2011). The moss study also showed that F-actin dynamics in the TS line were not related to regulation of known cell death pathways, suggesting a novel role for F-actin and Myosin XI in cell survival (Galotto et al., 2021). However, while we did find some cytoskeleton components, such as tubulin, being differentially regulated, there was a relative lack of actin-associated proteins in the DEGs we found. We also looked for Myosin family and Myosin-binding genes that were differentially expressed, but did not find significant differential expression for those genes except for a single GTD-binding domain containing protein, which may be further investigated. This may suggest that Myosin XI loss-of-function is solely responsible for F-actin related cell death, or that other related processes, such as vesicle transport and accumulation, may be responsible.

It is possible that many key genes which would explain the role of F-actin or other components known to be associated with Myosin XI are not well-annotated, as many of the most highly differentially expressed genes were uncharacterized. These uncharacterized genes may fit into known key pathways, including F-actin organization. Alternatively, they may include additional cell wall or stress response genes that are yet

to be identified in *P. patens*. However, it is also possible that these genes serve functions not highlighted through this analysis. GO enrichment analysis depends on annotated genes, so it is prone to ignoring poorly annotated pathways, as without annotations it is not possible to quantify enrichment (Young et al., 2010). Therefore, further investigation of these uncharacterized genes, particularly those highly differentially regulated, is necessary to fully understand the effects of Myosin XI loss and potentially explain the lack of growth phenotype.

Future work, both computational and experimental, may be done to build off these findings. Firstly, as mentioned previously, further computational approaches may be useful in identifying the function of key uncharacterized genes. Software such as MEME may be useful in identifying functional motifs in the sequences of uncharacterized proteins from this study (Bailey et al., 2015). Orthology-based methods such as OrthoFinder may also help to place uncharacterized genes in known gene families with shared function (Emms & Kelly, 2019). Experimental work will also be needed to verify the importance of genes identified in this study. Unlike with Myosin XI, which necessitated a temperature-sensitive inactivation method due to its key role in development, other techniques such as RNA interference may be possible to investigate the role of less vital proteins at the tip growth stage. With such methods, we may hope to verify and further explain the results of this study.

Bibliography

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166-169. https://doi.org/10.1093/bioinformatics/btu638

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. *Online*. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Bailey, T. L., Johnson, J., Grant, C. E., & Noble, W. S. (2015). The MEME Suite. *Nucleic Acids Res*, *43*(W1), W39-49. https://doi.org/10.1093/nar/gkv416

Bibeau, J. P., Galotto, G., Wu, M., Tuzel, E., & Vidali, L. (2021). Quantitative cell biology of tip growth in moss. *Plant Mol Biol*, *107*(4-5), 227-244. https://doi.org/10.1007/s11103-021-01147-7

Bibeau, J. P., Kingsley, J. L., Furt, F., Tuzel, E., & Vidali, L. (2018). F-Actin Mediated Focusing of Vesicles at the Cell Tip Is Essential for Polarized Growth. *Plant Physiol*, *176*(1), 352-363. https://doi.org/10.1104/pp.17.00753

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120. https://doi.org/10.1093/bioinformatics/btu170

Carey, R. E., & Cosgrove, D. J. (2007). Portrait of the expansin superfamily in Physcomitrella patens: comparisons with angiosperm expansins. *Ann Bot*, *99*(6), 1131-1141. https://doi.org/10.1093/aob/mcm044

Elzanati, O., Mouzeyar, S., & Roche, J. (2020). Dynamics of the Transcriptome Response to Heat in the Moss, Physcomitrella patens. *Int J Mol Sci*, *21*(4). https://doi.org/10.3390/ijms21041512

Emms, D. M., & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*, *20*(1), 238. https://doi.org/10.1186/s13059-019-1832-y

Engel, P. P. (1968). The Induction of biochemical and morphological mutants in the moss Physcomitrella patens. *American Journal of Botany*, *55*.

Fernandez-Pozo, N. (2023). PEATmoss: A Gene Expression Atlas for Bryophytes. *Methods Mol Biol*, *2703*, 91-107. https://doi.org/10.1007/978-1-0716-3389-2_8

Galaxy, C. (2022). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Res*, *50*(W1), W345-W351. https://doi.org/10.1093/nar/gkac247

Galotto, G., Wisanpitayakorn, P., Bibeau, J. P., Liu, Y. C., Furt, F., Pierce, E. C., Simpson, P. J., Tuzel, E., & Vidali, L. (2021). Myosin XI drives polarized growth by vesicle focusing and local enrichment of F-actin in Physcomitrium patens. *Plant Physiol*, *187*(4), 2509-2529. https://doi.org/10.1093/plphys/kiab435

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., & Rokhsar, D. S. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res*, *40*(Database issue), D1178-1186. https://doi.org/10.1093/nar/gkr944

Hartwell, L. H., Hopfield, J. J., Leibler, S., & Murray, A. W. (1999). From molecular to modular cell biology. *Nature*, *402*(6761 Suppl), C47-52. https://doi.org/10.1038/35011540

Huang, H. C., Niu, Y., & Qin, L. X. (2015). Differential Expression Analysis for RNA-Seq: An Overview of Statistical Methods and Computational Software. *Cancer Inform*, *14*(Suppl 1), 57-67. https://doi.org/10.4137/CIN.S21631

Kesawat, M. S., Satheesh, N., Kherawat, B. S., Kumar, A., Kim, H. U., Chung, S. M., & Kumar, M. (2023). Regulation of Reactive Oxygen Species during Salt Stress in Plants and Their Crosstalk with Other Signaling Molecules-Current Perspectives and Future Directions. *Plants (Basel)*, *12*(4). https://doi.org/10.3390/plants12040864

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

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Mansoor, S., Ali Wani, O., Lone, J. K., Manhas, S., Kour, N., Alam, P., Ahmad, A., & Ahmad, P. (2022). Reactive Oxygen Species in Plants: From Source to Sink. *Antioxidants (Basel)*, *11*(2). https://doi.org/10.3390/antiox11020225

Rensing, S. A., Goffinet, B., Meyberg, R., Wu, S. Z., & Bezanilla, M. (2020). The Moss Physcomitrium (Physcomitrella) patens: A Model Organism for Non-Seed Plants. *Plant Cell*, *32*(5), 1361-1376. https://doi.org/10.1105/tpc.19.00828

Schaefer, D. G., & Zryd, J. P. (1997). Efficient gene targeting in the moss Physcomitrella patens. *Plant J*, *11*(6), 1195-1206. https://doi.org/10.1046/j.1365-313x.1997.11061195.x

Smertenko, A., & Franklin-Tong, V. E. (2011). Organisation and regulation of the cytoskeleton in plant programmed cell death. *Cell Death Differ*, *18*(8), 1263-1270. https://doi.org/10.1038/cdd.2011.39

UniProt, C. (2023). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res*, *51*(D1), D523-D531. https://doi.org/10.1093/nar/gkac1052

Vidali, L., Burkart, G. M., Augustine, R. C., Kerdavid, E., Tuzel, E., & Bezanilla, M. (2010). Myosin XI is essential for tip growth in Physcomitrella patens. *Plant Cell*, *22*(6), 1868-1882. https://doi.org/10.1105/tpc.109.073288

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, *10*(1), 57-63. https://doi.org/10.1038/nrg2484

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., & Yu, G. (2021). clusterProfiler 4.0: A universal enrichment tool for

interpreting omics data. *Innovation (Camb)*, *2*(3), 100141. https://doi.org/10.1016/j.xinn.2021.100141

Young, M. D., Wakefield, M. J., Smyth, G. K., & Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol*, *11*(2), R14. https://doi.org/10.1186/gb-2010-11-2-r14