Introduction

The aim of the study *Nutrient levels control root growth responses to high ambient temperature in plants* was to investigate thermomorphogenesis in plants, which is how plants respond to high temperatures. Botanical biomarkers of thermomorphogenesis include quickened tissue growth (including roots), hyponastic growth (where plants curve due to this growth), and early flowering. The paper notes the societal importance of this research; botanists need to understand how plants respond to heightened temperatures, as climate change is a major environmental issue.

There are two transcriptional factors that serve as central hubs in thermomorphogenesis: phytochrome interacting factor 4 (PIF4) and elongated hypocotyl 5 (HY5). PIF4 plays a major role specifically in shoot thermomorphogenesis, while HY5 plays a major role in root thermomorphogenesis (Lee et al., 2024). Another key factor to consider in thermomorphogenesis is energy allocation. Nitrogen uptake is energy and resource intensive for plants, and heightened temperatures and accompanying heat stress can be associated with impaired nitrogen uptake (Mishra et. al, 2023). A critical and well researched botanical gene for nitrate sensing and transportation is nitrate transporter (NRT1.1).

It is important to note that the HY5 gene is a master regulator for light-mediated transcription. It is involved in many physiological and biological processes in plants such as photomorphogenesis, root growth, flavonoid biosynthesis, acquiring nutrients, and response to stress. Thus, the role of the HY5 is not limited to just root thermomorphogenesis (Xiao et al., 2022).

One of the many roles of the HY5 gene is to downregulate NRT1.1. NRT1.1 is downregulated by HY5 in a root specific manner at heightened temperatures in roots. HY5 will directly bind to the promoter NRT1.1 to make this happen. Thus, in order to make thermomorphogenesis happen, plants appear to allocate energy and resources into growing larger roots, and take it away from nitrogen sensing, transportation, and uptake.

Therefore, for the RNA-seq portion of *Nutrient levels control root growth responses to high ambient temperature in plants*, three *Arabidopsis thaliana* genotypes were sequenced to investigate their global gene expression. Results were compared to previous research. *Arabidopsis thaliana* is a common model organism in botany, with high quality reference genomes available. The three genomes of *Arabidopsis thaliana* that were sequenced were: Col-0, which represents wild-type arabidopsis, hy5-215, which is an arabidopsis mutant that lacks the HY5 gene, and chl1-5, which is an *Arabidopsis thaliana* mutant that lacks the NRT1.1 gene. The plants were grown at 21 degrees celsius (the control group) and 28 degrees celsius (the treatment group) and the genes that were differentially expressed under these different circumstances were investigated. 21 degrees celsius is considered a temperature that promotes homeostasis in *Arabidopsis thaliana*, while 28 degrees celsius promotes thermomorphogenesis (Lee et al., 2024).

Methods

For the RNA-seq portion of *Nutrient levels control root growth responses to high ambient temperature in plants*, the genotypes of three different arabidopsis plants were sequenced (Col-0, hy5-215, and chl1-5). First, all the arabidopsis plants were germinated for 4 days at 21 degrees celsius. Next, control groups continued to be grown at 21 degrees celsius for another 5 days. The treatment arabidopsis plant groups, on the other hand, were grown at 28 degrees celsius for 5 days after the initial period of germination at 21 degrees celsius. Each genotype, and each control and treatment group, had 3 biological replicates, for a total of 18 samples. That is to say, 3 genotypes, multiplied by 2 temperatures, multiplied by 3 replicants, would be equal to 18 runs.

Next, samples were taken. Root samples were extracted using the RNeasy Plant Mini Kit. Next, RNA quality and quantity was analyzed via a 2100 Bioanalyzer tape station from Agilent Technologies and Qubit Fluorometer from Invitrogen. Sequencing was conducted via the Illumina Novaseq6000 platform, and the sequencing libraries were generated through the Salk Next Generation Sequencing Core based on the Illumina manufacturer's specifications. Short-reads were mapped using the Arabidopsis Information resource web site www.arabidopsis.org and the Splice Transcripts Alignments to Reference (STAR) version 2.7.0a.

Afterwards, a differentially expressed genes analysis was conducted via the software package edgeR. To initially determine differentially expressed genes, the false discovery rate used was 0.05, and a log2FC cutoff of less than or greater 0 was used. However, next, a k-means clustering method was used with a log2FC cutoff of less than or equal to an absolute value of 1 and a false discovery rate of 0.05 to more strictly determine differentially expressed genes. These cutoffs of an absolute value of 1 (log2FC) and 0.05 (false discovery rate) were used for reporting and visualization, including the study's heatmap, which was made via the ComplexHeatmap software package.

A venn diagram was made of the three different genotypes. This way, how the genes that were differentially expressed overlapped across the three different genotypes could be visualized. The groups shown in the Venn diagram were the number of genes that were only differentially expressed in Col-0, chl1-5, and hy5-215, the number that were differentially expressed in all the genotypes, and the numbers that were differentially expressed in both Col-0 and chl1-5, both chl1-5 and hy5-215, and both Col-0 and hy5-215.

Gene Ontology (GO) analysis was conducted on the set of genes that were either differentially expressed in only the genotype Col-0, or both Col-0 and hy5-215. That is, genes that were differentially expressed in Col-0 but not also differentially expressed in chl1-5. This was done because chl1-5 is the genotype that is missing the NRT1.1 gene, and the paper aimed to isolate the effect of the NRT1.1 gene on gene ontology. This group was labeled the "NRT1.1 dependent group".

In addition, Gene Ontology (GO) analysis was conducted on the set of genes that were either differentially expressed in only the genotype Col-0, or both Col-0 and chl1-5. That is, genes that were differentially expressed in Col-0 but not also differentially expressed in hy5-215.

This is because hy5-215 is the genotype that is missing the HY5 gene, and the paper aimed to isolate the effect of the HY5 gene on gene ontology. This group was labeled the "HY5 dependent group".

Lastly, Gene Ontology (GO) analysis was conducted on the set of genes that were differentially expressed in both the genotype Col-0 and the genotype chl1-5. This way, the scientists can isolate the downstream effects of the HY5 on gene expression by seeing what genes are still differentially expressed without it, but with the NRT1.1 gene. This group was labeled the HY5-NRT1.1 dependent group.

No repositories were provided, but the data for RNA-seq analysis is available in the Gene Expression Omnibus at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE262197. All samples are single-end, rather than paired end (Lee et al., 2024).

Results

To replicate the experiment, FactQC (v.0.12.1) was conducted on all the raw data, then it was summarized through MultiQC (v1.15). The raw data was found through the Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE262197). All 18 samples passed Mean Quality Scores, Per Sequence Quality Scores, Sequence Quality Histograms, Per Base N Content, Sequence Length Distribution, Adapter Content, and Sequence Length Distribution. There were thus no obviously problematic samples.

Figure 1: All 18 samples passed **Mean Quality Scores**, with Phred Scores above 35.

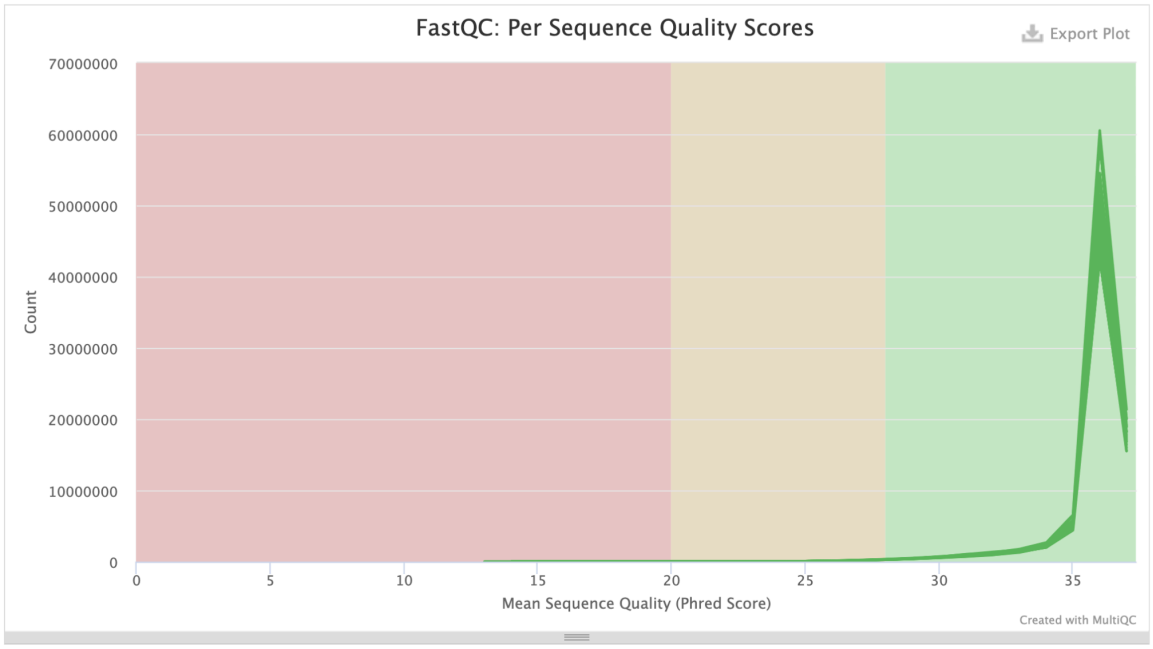

Figure 2: All 18 samples passed **Per Sequence Quality Scores**, with a peak of about 60,000,000 counts.

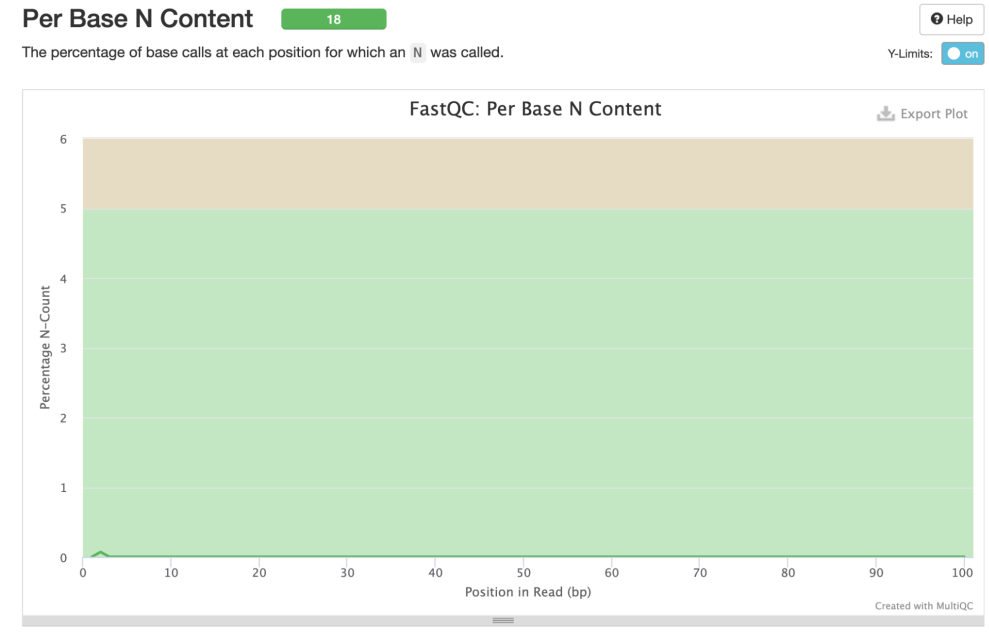

Figure 3: All 18 samples based **Per Base N Content**, with very low Percentage N-Counts.

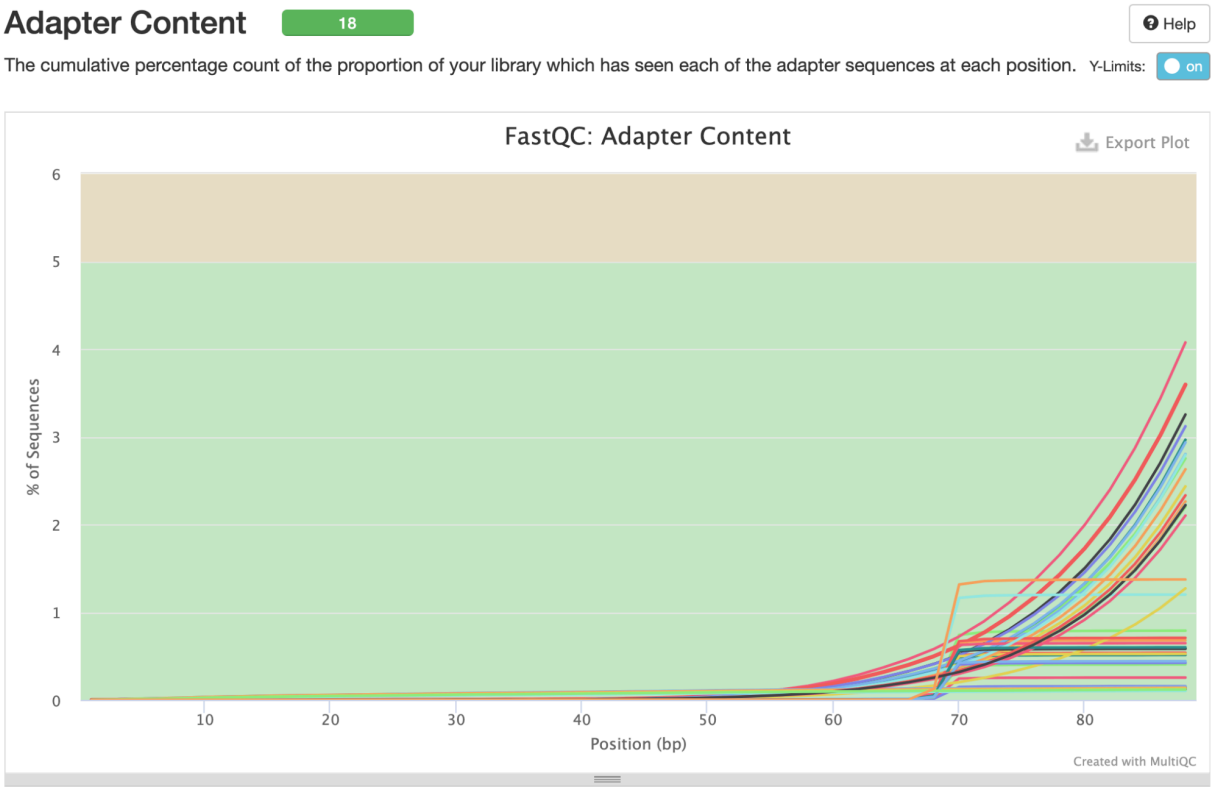

Figure 4: All 18 sequences passed the **Adapter Content** test, meaning that all 18 samples did not have marked issues with adapter contamination.

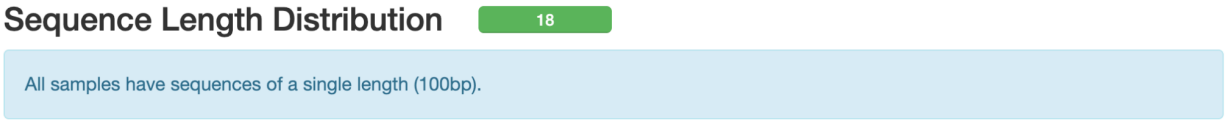

Figure 5: All 18 samples were the same length, and passed the **Sequence Length Distribution** test.

However, some statistics that needed to be considered further were that Per Base Sequence Content and Sequence Duplication Levels failed for all 18 samples, and 6 of my samples had warnings for Per Sequence GC content, and 16 had warnings for Overrepresented Sequences. However, since MultiQC treats the RNA-seq data as though it was genomic data, and RNA-seq data can skew in response to certain genes being up-regulated or down-regulated, this didn't seem like a strong cause for concern. Since all 18 samples passed the most important tests, one could be confident that they could continue the RNA-seq analysis with the samples.

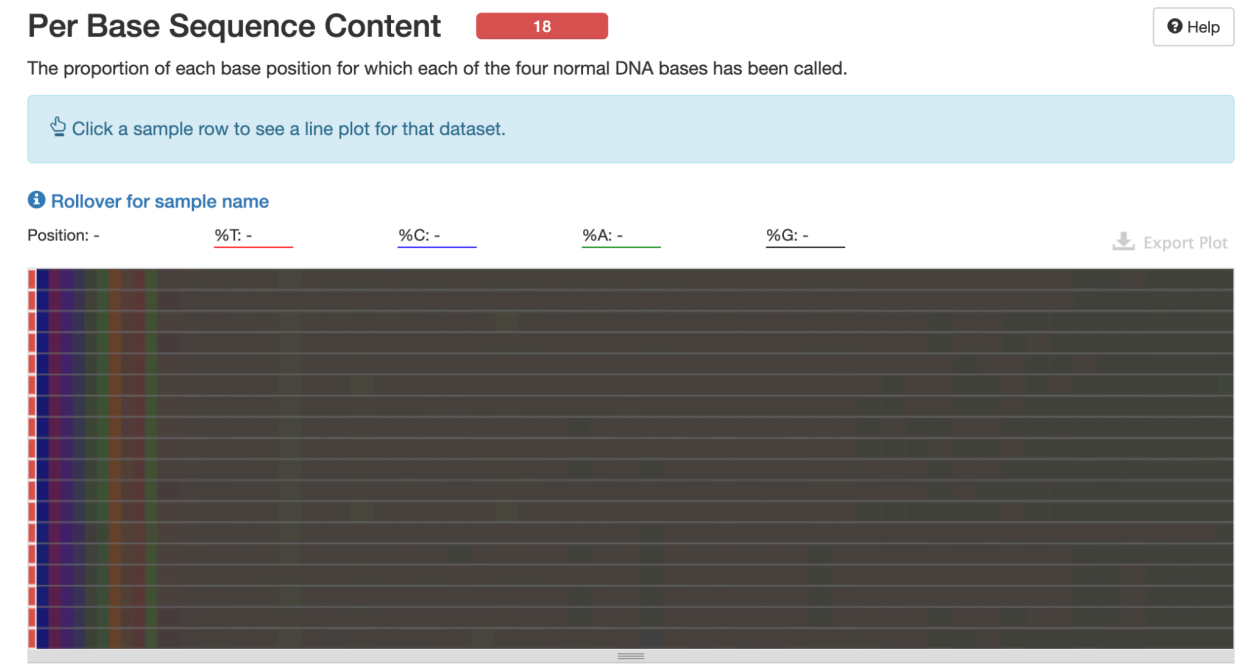

Figure 6: All 18 samples failed for **Per Base Sequence Content**. However, if one selects any individual sample, the percentage of reads evens out as the percentage (bp) increases. This is a good sign for high quality data.

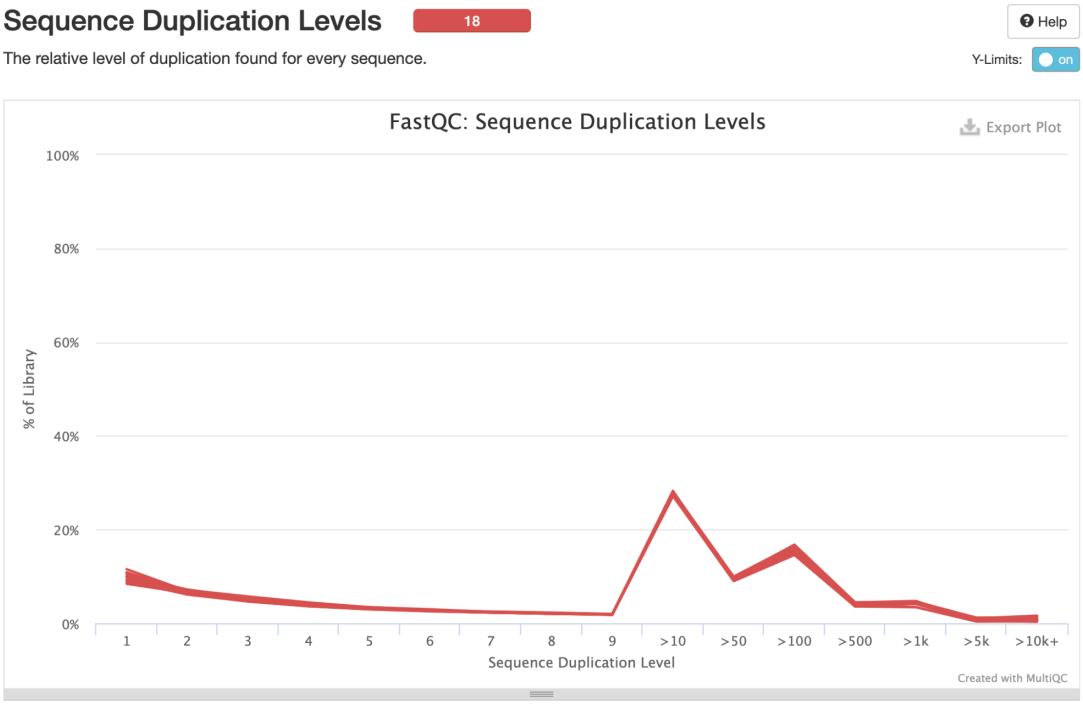

Figure 7: All 18 samples failed for **Sequence Duplication Levels**. However, given that FastQC and MultiQC treats the RNA-seq data as though it was genomic data, and RNA-seq data can

skew in duplication in response to certain genes being up-regulated or down-regulated, this didn't seem like a strong cause for concern. Given the pass rates for the other tests, this was not deemed a reason to discontinue or restart the experiment.

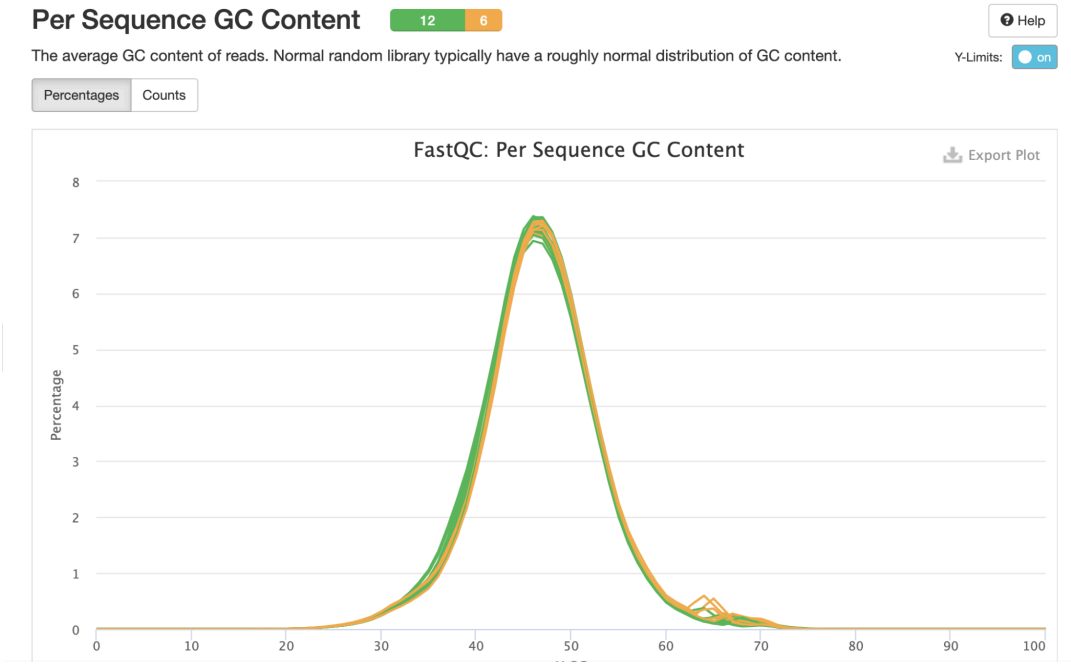

Figure 8: The majority of the samples passed **Per Sequence GC Content**. All samples followed a normal distribution closely.

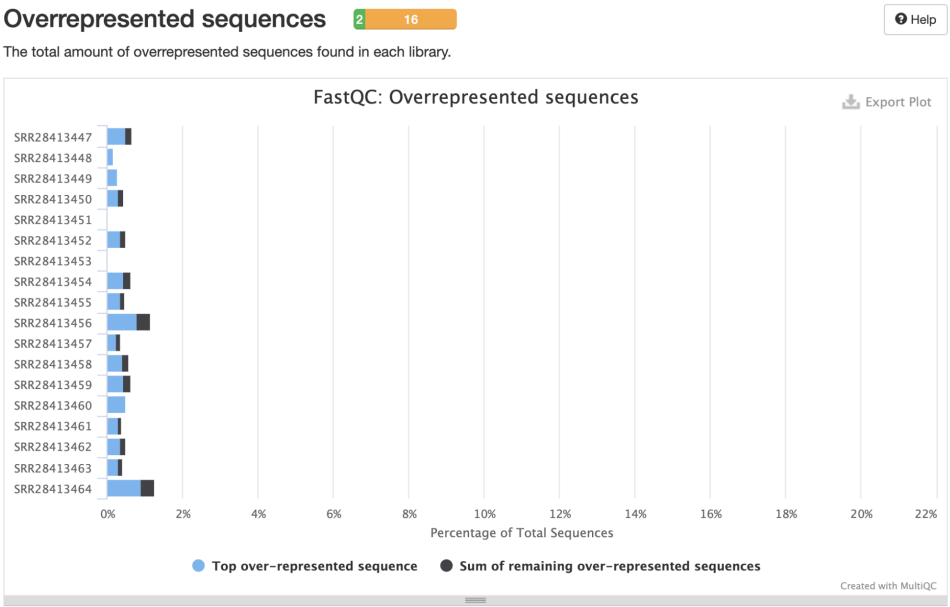

Figure 9: Most samples failed the **Overrepresented sequences** test. However, given that FastQC and MultiQC treats the RNA-seq data as though it was genomic data, and RNA-seq data can skew in duplication in response to certain genes being up-regulated or down-regulated, this didn't seem like a strong cause for concern. Given the pass rates for the other tests, this was not deemed a reason to discontinue or restart the experiment.

Next, the data was then trimmed using Trimmomatic (v0.39). Next, FastQC (v0.12.1) was run on the trimmed files, then FastQC was again summarized by MultiQC (v1.15).

An *Arabidopsis thaliana* reference genome was downloaded from Ensembl. Hisat2 (v2.2.1) was used to create an index for the genome file. Samtools (1.16.1) was used for sequence alignment to the Ensembl reference genome, which created bam files as an output. Htseq (v0.13.5) was used to create the counts files. There were no outliers in the counts files.

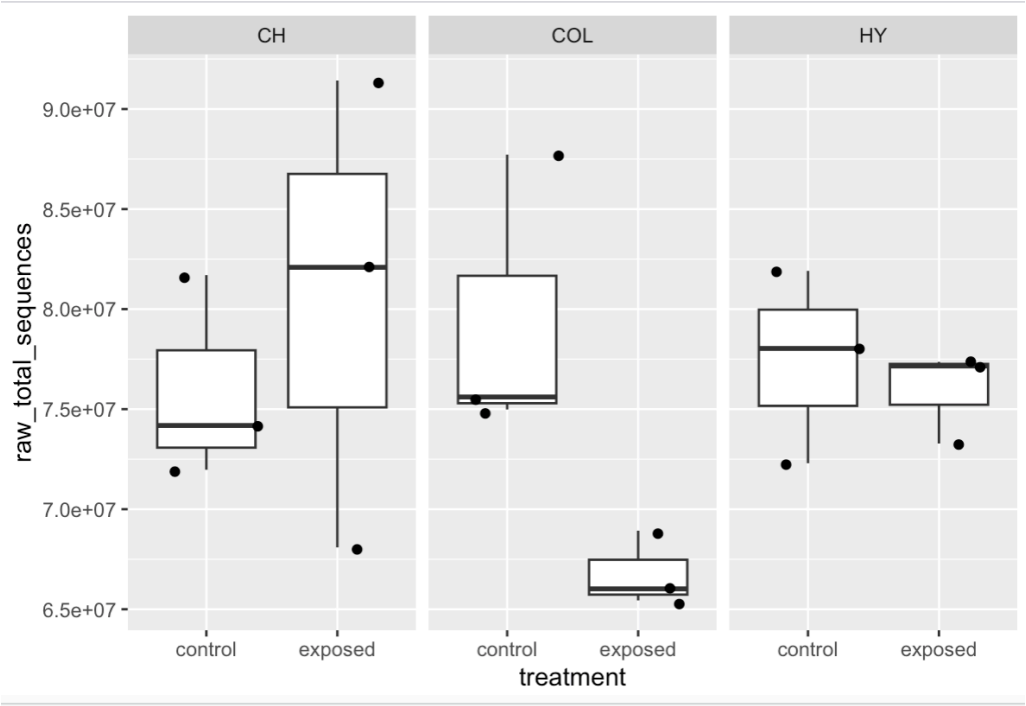

Figure 10: A boxplot was created to see if there were any outlier samples that needed to be discarded. However, no outliers were found. CH represents reads from the chl1-5 genotype, COL represents reads from the Col-0 genotype, and HY represents reads from the HY5-215 genotype.

Although the number of raw total sequences for the Col-0 exposed group is smaller than for the other groups, all three samples cluster together. In addition, the number of raw sequences for chl1-5 varies a lot, but its minimum is still higher than the number of reads for most sequences for the Col-0 genotype, and its maximum is not considerably higher than the

maximum number of reads for the Col-0 genotype. Thus, the counts boxplot analysis gave no reason to discard any of the samples.

The differential expression analysis reflected this high quality genetic data. A log2fold change with an absolute value of at least 1 was used, and a false discovery rate of 0.05 was used. The software package edgeR was used, all in accordance with the paper. A table was created for the results using the tibble library, and a Venn diagram for the results was created using the ggVennDiagram library.

```
  genotype   low  high total
  <chr>    <int> <int> <int>
1 COL        711   802  1513
# A tibble: 1 × 4
  genotype   low  high total
  <chr>    <int> <int> <int>
1 CH         972   823  1795
# A tibble: 1 × 4
  genotype   low  high total
  <chr>    <int> <int> <int>
1 HY         540   746  1286
```

Figure 11: A table was created using the Tibble library, which displayed the number of underexpressed, overexpressed, and total genes for the 3 different genotypes. COL represents DEGs from the Col-0 genotype, CH represents DEGs from the chl1-5 genotype, and HY represents DEGs from the HY5-215 genotype.

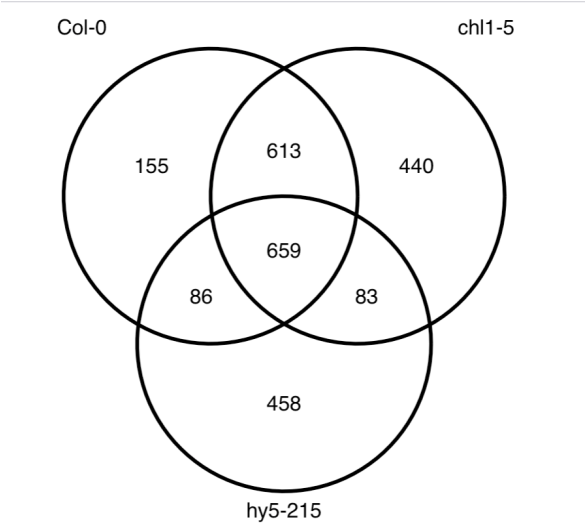

Figure 12: A Venn diagram was created, showing the overlapping and non-overlapping DEGs for the Col-0 genotype, the chl1-5 genotype, and the hy5-215 genotype.

The paper arrived at a result of 1,512 genes differentially expressed for heat-exposed Col-0 compared to the control, 1,761 genes differentially expressed for heat-exposed chl1-5 compared to the control, and 1,295 genes differentially expressed for heat-exposed hy5-215 compared to the control. The replication results were 1,513 genes, 1,795 genes, and 1,286 genes respectively.

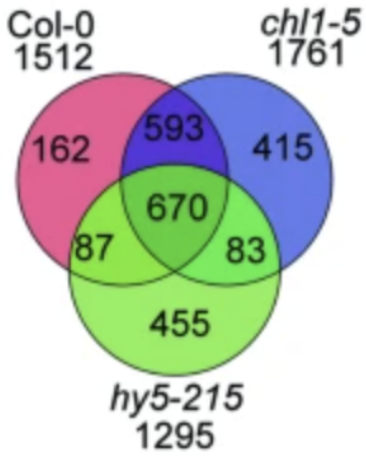

Figure 13: This Venn diagram is from the paper (Lee et al., 2024). The results of the paper were quite similar to the results of the replication, as discussed further below.

In addition, a PCA plot was created using the ggplot2 library. It used the 500 most differentially expressed genes for the variance calculation, and all 18 samples successfully clustered into groups based on control and treatment, and their genotype.

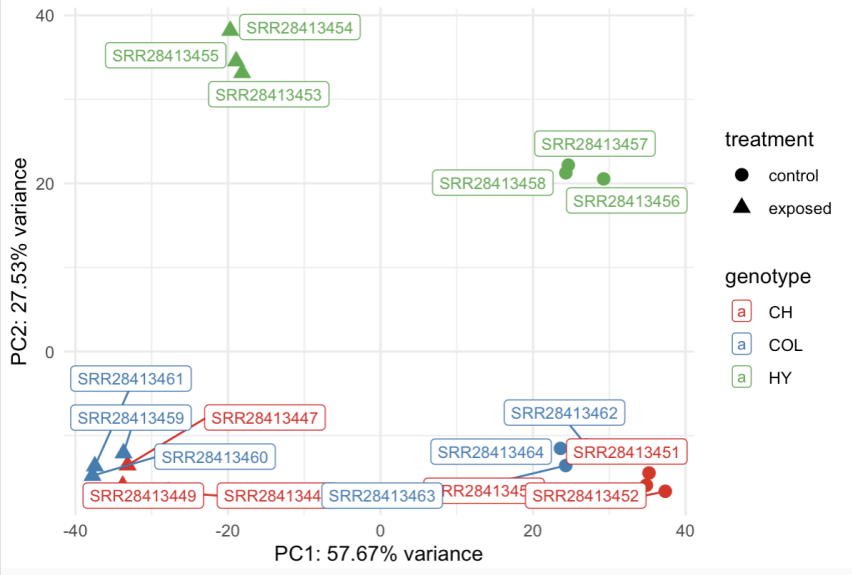

Figure 14: A PCA plot was created, with the control being the plants grown at 21 degrees celsius, and the treatment group being the plants grown at 28 degrees celsius. COL represents the Col-0 genotype, CH represents the chl1-5 genotype, and HY represents the HY5-215 genotype.

Similarly, a heatmap was created, using the pheatmap library package. It used the 50 most differentially expressed genes, and differences in coloration can be seen based on genotype, and whether the sample came from a control or treatment group:

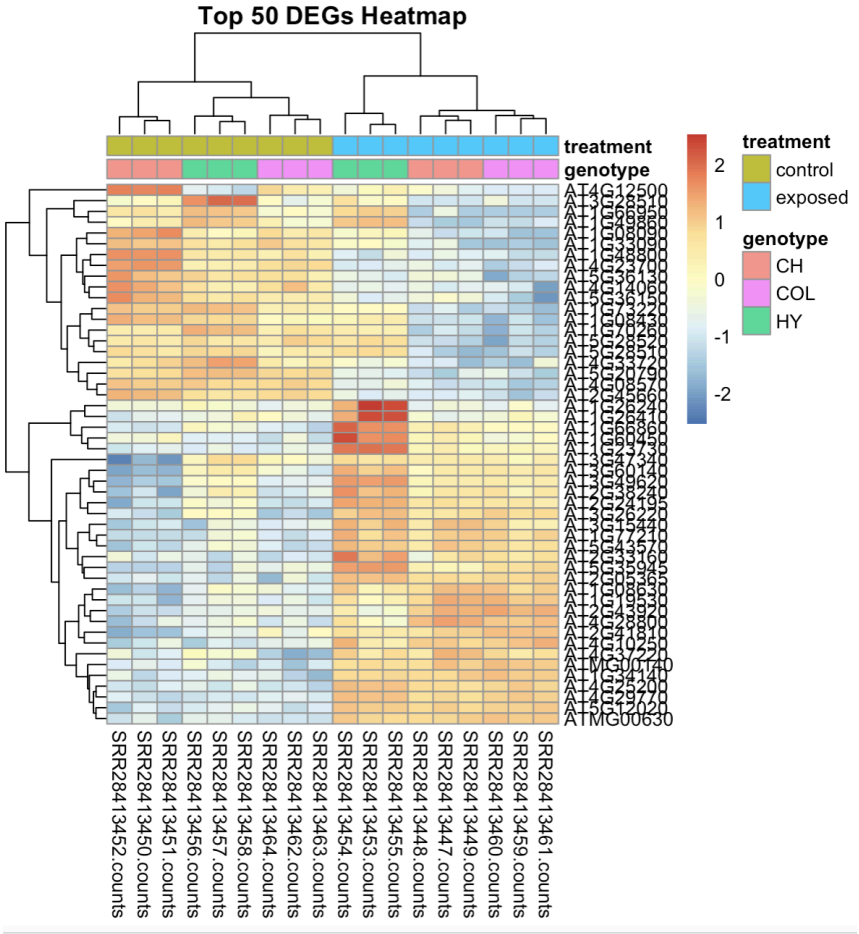

Figure 15: A heat map was created, with the control being the plants grown at 21 degrees celsius, and the treatment group being the plants grown at 28 degrees celsius. COL represents the Col-0 genotype, CH represents the chl1-5 genotype, and HY represents the HY5-215 genotype.

During the study replication, the gene IDs for the top ten most differentially expressed genes by log2 fold change for each genotype were printed. The study *Nutrient levels control root growth responses to high ambient temperature in plants* itself did not make a list of gene IDs of the most differentially expressed genes from its RNA-seq; supplementary tables were used to display other information.

[1] "Top DEGs for Col-0"

	gene_id	logFC
AT1G77210	AT1G77210	7.081692
AT1G73220	AT1G73220	-6.868246
AT5G12020	AT5G12020	6.833703
AT1G70260	AT1G70260	-6.705922
AT4G25200	AT4G25200	6.582449
AT2G43920	AT2G43920	6.110367
AT5G28520	AT5G28520	-5.846836
AT5G20790	AT5G20790	-5.721607
AT1G34140	AT1G34140	5.686374
AT1G08090	AT1G08090	-5.465913

Figure 16: This table shows the top ten most differentially expressed genes by log2 fold change for the Col-0 genotype.

[1] "Top DEGs for chl1-5"

	gene_id	logFC
AT1G77210	AT1G77210	9.789294
AT1G73220	AT1G73220	-8.046986
AT2G43920	AT2G43920	7.739200
AT3G47340	AT3G47340	7.646656
AT1G48800	AT1G48800	-7.378588
AT4G33720	AT4G33720	-7.149230
AT5G12020	AT5G12020	7.065590
AT1G08630	AT1G08630	6.417172
AT1G08090	AT1G08090	-6.395026
AT3G49620	AT3G49620	6.042421

Figure 17: This table shows the top ten most differentially expressed genes by log2 fold change for the chl1-5 genotype.

[1] "Top DEGs for hy5-215"

	gene_id	logFC
AT1G60450	AT1G60450	9.084138
AT1G26240	AT1G26240	7.074107
AT4G29770	AT4G29770	6.389318
AT4G33720	AT4G33720	-6.334975
AT5G12020	AT5G12020	6.202378

AT5G35945	AT5G35945	6.159708
AT4G25200	AT4G25200	6.011895
AT1G26410	AT1G26410	5.862986
AT2G33160	AT2G33160	5.625791
AT5G20790	AT5G20790	-5.599470

Figure 18: This table shows the top ten most differentially expressed genes by log2 fold change for the hy5-215 genotype.

Lastly, an over-representation analysis was conducted for functional analysis. All genes that were in the count matrix (all genes in the dataset) were used for the background. Ensembl gene IDs were mapped to the The Arabidopsis Information Resource (TAIR) IDs using the *Arabidopsis thaliana* dataset from Ensembl Plants via the biomaRt package. TAIR is considered the gold standard for *Arabidopsis thaliana* gene ontology research (Rhee et al). Gene Ontology enrichment was conducted with clusterProfiler, with significance cutoffs of $p < 0.05$ and $q < 0.1$. In accordance with the paper, an over-enrichment analysis was conducted on three gene sets: NRT1.1-dependent genes, HY5-dependent genes, and HY5-NRT1.1-dependent genes. These gene sets were each used as a foreground. Dotplots were created for the three different categories to visualize the results.

For the HY5-dependent group, 25 enriched terms were found. For the top 15 enriched terms by gene ratio, see the following dotplot:

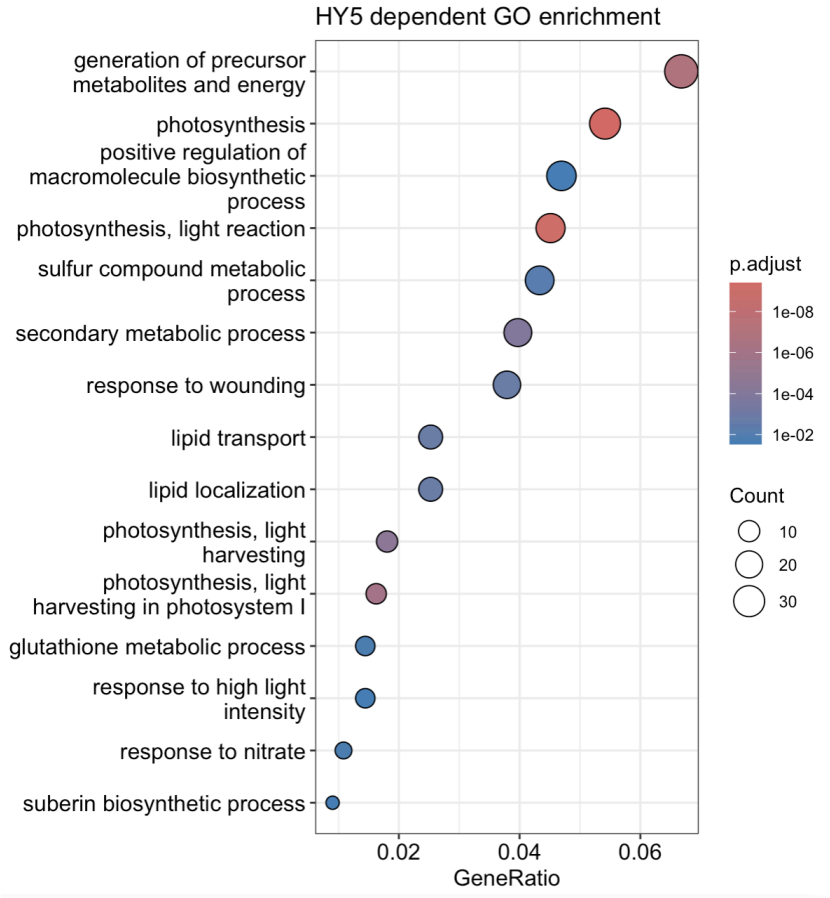

Figure 18: A dot plot was created of the 15 most enriched terms by gene ratio for the HY5-dependent group. There were 25 enriched terms found.

In the paper, the enriched terms listed were “photosynthesis”, “response to nitrate”, “response to sucrose”, and “glucosinate biosynthetic process”.

For the HY5-NRT1.1 dependent group, 16 enriched terms were found. See the following dotplot for the top 15 enriched terms:

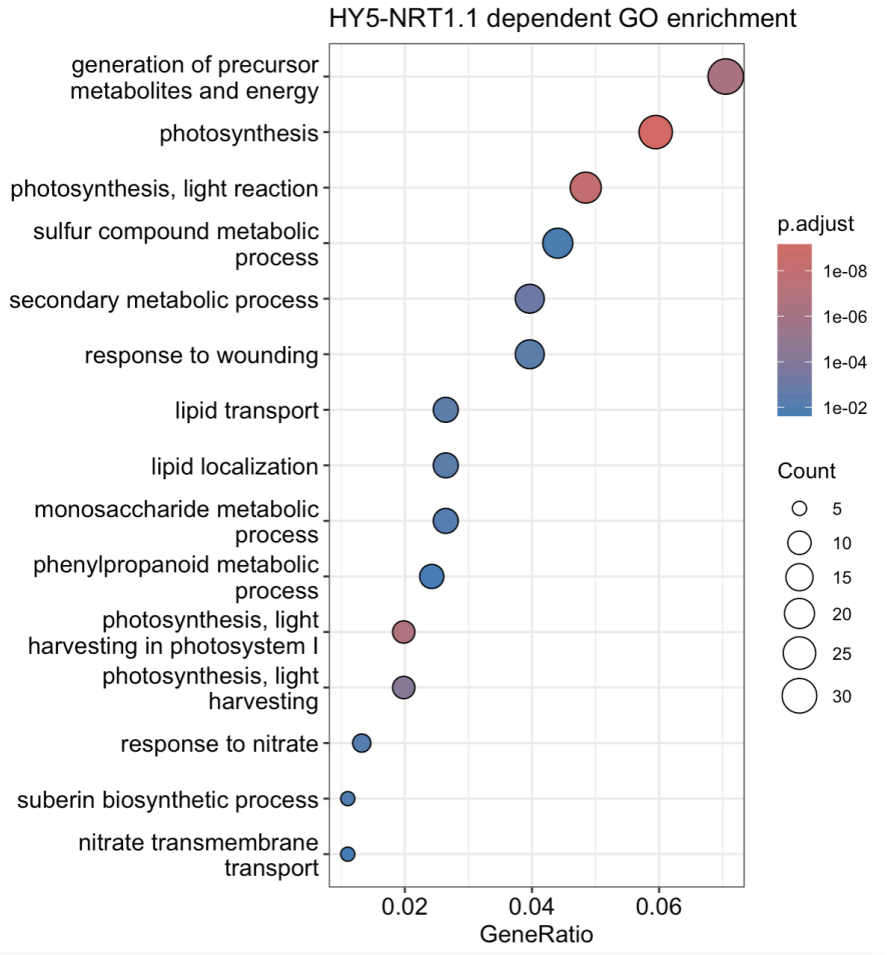

Figure 19: A dot plot was created of the 15 most enriched terms by gene ratio for the HY5-NRT1.1 dependent group. There were 16 enriched terms found.

In the paper, the enriched terms listed were “photosynthesis”, “suberin biosynthetic process”, “response to nitrate”, and “response to sucrose”.

Lastly, for the NRT1.1 dependent group, no enriched terms were found to be significant. However, the paper only listed one enriched term, which was “response to chemical” (Lee et al., 2024).

Discussion

One can be quite confident that the results were well replicated.

For differential gene expression, all three results were within 34 genes of the original, which represents common variation caused by rounding, software versions, and other small differences. Similarly, all subsections of the Venn diagrams were within 25 genes from the paper. An issue I ran into was that I had to switch software packages to achieve this result, as the paper uses edgeR. This uses different modeling mathematics than DESeq, which can affect results.

Another issue I ran into was that of the log2 fold change. I originally used a log2 fold change cutoff of greater than or less than 0, however, my results were very far off the paper. In addition, the wording of the paper was confusing, as they themselves originally claimed to use a log2 fold change of greater than or less than 0, then used an absolute value of 1 afterwards for a more stringent analysis. The wording was slightly ambiguous; however, given how closely my results matched the paper in the end, I am confident that my false discovery rate, log 2 fold change, and the use of edgeR were all necessary ingredients to properly match the results of the paper.

In addition, the PCA plots clustered very well by genotype and control versus treatment. It is interesting that the hy5-215 samples clustered so far away from the other two genotypes, but it makes sense, as the HY5 gene is considered the master gene for thermomorphogenesis and photosynthetic processes, and the NRT1.1 gene is downstream to it. It makes sense that a plant that lacks this gene would have biological responses that are very different from the other genotypes. In addition, the heat map shows a similar pattern; it clusters very well by genotype and control versus treatment, with the hy5-215 plants, again, showing a more extreme response than the other genotypes.

Similarly, my over-representation analysis matched the paper quite nicely. For the HY5 dependent group, the GO terms “photosynthesis” and “response to nitrate” were enriched in both our papers. In the paper, “response to sucrose” and “glucosinolate biosynthetic process” were also recovered. This is similar to the GO terms enriched in my results, such as “metabolites and energy”, and “glutathione metabolic process”. For the HY5-NRT1.1 dependent group, the GO terms “photosynthesis”, “suberin biosynthetic process”, and “response to nitrate” were enriched in both analyses. In the paper, “response to sucrose” was also recovered. However, again, I recovered the term “secondary metabolic process”, which is quite similar.

The paper notes that more than 83% of root Col-0 differentially expressed genes were shared between Col-0 and chl1-5, meaning that the mutant lacking the NRT1.1 gene shares a very similar thermomorphogenetic response to that of the wild type. I got a result indicating that 84% of Col-0 differentially expressed genes were shared between Col-0 and chl1-5, which is a very similar result to the 83% from the paper. In addition, the paper got the same GO terms for the HY5 dependent group, and the HY5-NRT1.1 dependent group, just in different orders. For my results 12 of the top 15 enriched GO terms were shared for both groups. Thus, both these two pieces of information were well replicated, and are in line with the mutant lacking the NRT1.1 gene sharing a very similar thermomorphogenetic response to the wild type.

I originally thought it was a problem that I did not recover any GO terms for the NRT1.1 dependent group. However, eventually, I realized this was actually biologically meaningful. The only GO term that the paper recovered for this group was “response to chemical”. What is more important is what they did not recover, which was a lack of GO terms related to photosynthesis and metabolic processes. This stresses the critical importance of the HY5 gene in thermomorphogenesis, and its role in downregulating the NRT1.1 gene. Since the NRT1.1 dependent group took out GO terms based on the genes dependent on the HY5 gene, which is the master gene for thermomorphogenic and photosynthetic processes, this result makes perfect

sense. Thus, I would strongly argue that this also replicated the results of the paper. In addition, this group only represents 249 genes (paper) and 241 genes (replication). It makes sense that, even with a very strong replication, one GO term, that of "response to chemical", might not show up due to expected mathematical differences.

Thus, since NRT1.1 is downstream to HY5, it makes biological sense that the hy5-215 plant, which lacks the HY5 gene, would produce the most markedly different thermomorphogenetic response. This was also the common story in both the heat map and the PCA plot I produced.

Thus, all in all, my results recapitulated the original findings, which makes sense, as this is a *Nature* paper, which represents a very stringent study with little room for errors.

Citations:

Lee, S., Showalter, J., Zhang, L. *et al.* Nutrient levels control root growth responses to high ambient temperature in plants. *Nat Commun* 15, 4689 (2024). https://doi.org/10.1038/s41467-024-49180-6

Mishra S, Spaccarotella K, Gido J, Samanta I, Chowdhary G. Effects of Heat Stress on Plant-Nutrient Relations: An Update on Nutrient Uptake, Transport, and Assimilation. Int J Mol Sci. 2023 Oct 27;24(21):15670. doi: 10.3390/ijms242115670. PMID: 37958654; PMCID: PMC10649217.

Rhee SY, Beavis W, Berardini TZ, et al. The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. *Nucleic Acids Research.* 2003;31(1):224–228.

Xiao Y, Chu L, Zhang Y, Bian Y, Xiao J, Xu D. HY5: A Pivotal Regulator of Light-Dependent Development in Higher Plants. Front Plant Sci. 2022 Jan 17;12:800989. doi: 10.3389/fpls.2021.800989. PMID: 35111179; PMCID: PMC8801436.