Ben Camber

Ecological Genomics

March 27, 2020

Differential gene expression analysis of the effects on transcription of red spruce (*Picea rubens*) climate transplantation.

Introduction

Simulating global climate change's anticipated effects on living things, particularly drought and heat stress is an important step in understanding and preparing for our changing environment. Common garden experiments are the premier way to disentangle the effects of genetics (G), environment (E), and GxE interaction effects. Red spruce (*Picea rubens*) populations in eastern North America have dwindled since the end of the last ice age due to human activity. Since then, the cool, wet conditions it needs to survive ceased to exist in the southern portion of its range, making recolonization of these sites unlikely. However, rather than assume *P. rubens* is incompatible with its former southern range, we measured differential gene expression of red spruce saplings from two populations with contrasting climates that had been treated with either control (C), hot (H), or hot and dry (D) conditions. In this way, we can test whether trees from different climates have different gene expression profiles in a common environment, characterize the transcriptional response to heat stress and drought, determine any interaction effects of climate and stress treatment, and discover genes important to these responses to crucial environmental stressors.

Methods

Individuals from ten maternal families across two source climates (abbreviated CW for cool and wet, HD for hot and dry) were raised in a common garden environment for ten week before being placed in control (C) conditions. After two weeks, two-thirds of the seedlings were placed in the two treatment groups hot (H) and hot/dry (D). RNA was extracted from needle, root, and stem tissues after 0, 5, and 10 days of treatment.

Samples were quantified for RNA concentration and quality on a Bioanalyzer and all samples with >1 ng/uL RNA were sent to Cornell for 3' tag sequencing. Library prep followed the LexoGen protocol   (LexoGen, Vienna, Austria) and sequencing occurred on one lane of a NextSeq500 (Illumina, San Diego, CA) with a resolution of 86bp per fragment. FastQC [1] was used for visualization of quality scores, adaptor sequences were removed with Trimmomatic [2], then the cleaned reads were verified by FastQC once more. Cleaned reads were simultaneously quantified and mapped to a reference transcriptome [3] using Salmon [4]. The mapped and quantified reads were passed to the R-based transcriptome importer tool tximport [5] to create a counts matrix. The counts matrix was put in a table then had all count values rounded to integers to match DESeq2 [6] requirements. We then calculated counts per sample to ensure sufficient coverage.

After filtering the counts table and samples description table for only the day 10 transcriptomes, both objects were passed to DESeqDataSetFromMatrix() to construct a generalized linear model design matrix based on the climate and treatment variables. Additionally, a second model design matrix including an interaction effect of climate and treatment. Genes with >1 read per sample on average were removed from each model design matrix before they were passed to DESeq(), which tested for differential gene expression between samples from different climates and treatments. Top differentially expressed genes and principle components of differential gene expression were then extracted from the DESeq dataset.

Results

The sequenced transcriptomes had a mean read count of 2743475 (Fig. 1) and a mean read count per gene of 3139. However, the median read count per gene was only 24, indicating that some genes have extremely high relative expression.

Heat stress alone elicited upregulation of 82 genes and downregulation of 9 vs. control, while heat and drought resulted in the upregulation of 1457 genes and downregulation of 1196 genes (Fig. 2), while samples from different climates had zero significantly differentially expressed genes.

Principle component analysis of the day ten transcriptome samples revealed no clustering by source climate, but it did reveal that hot/dry treatment elicits a strong but variable signature in the transcriptome (Fig. 3). Interestingly, including an interaction effect in the DESeq model did not change the clustering of the samples in the PCA. A heatmap demonstrating the similarity of the transcription profiles of C and H compared to D can be seen in Figure 4.

Differentially expressed genes for C vs. D with the lowest adjusted p-values (*padj*) and greatest absolute log fold changes were identified on Congenie.org. Of the five genes with lowest *padj*, three are listed as phosphatases, including two different results for phosphatase 2C. The four genes with the highest absolute log fold changes seemed to have no common function, but, perplexingly one "low temperature induced" gene was highly upregulated (LFC=7.96, *padj*<0.000001) in the hot and dry treatment.

Conclusions

It is clear that *Picea rubens* has a greater transcriptional response to hot and dry conditions than to heat stress alone. This could indicate that drought is a common challenge for red spruce and they have been selected to have highly plastic environmental thresholds. On the other hand, if heat stress alone can be ameliorated by high frequency, efficient, constitutively active mechanisms, then the comparatively small transcriptional response may indicate that acute heat stress--but not drought--was common throughout the evolution of *P. rubens*. Of these two scenarios, the latter seems more likely based on the species' southern expansion during the last ice age. If the red spruce were endemic to a region with inconsistent precipitation patterns as opposed to New England and southern Canada, plasticity in water requirements might be expected, but this robust differential gene expression between C and D for trees from different source climates suggests that drought is a costly challenge to *P. rubens*.

While the functions of the significantly differentially expressed phosphatases in *P. rubens* may not be understood, phosphatases' well documented roles as post-translational modifiers of proteins and modulators of cell signaling [7] . While post-translational modification may be expected to be an acute rather than long-term response to environmental stress, the activity of phosphatases and their kinase counterparts is ubiquitous and dynamic, and mechanisms that create rapid yet sustainable responses to temperature or water availability changes may be sufficient for phenotypic plasticity without the transcription of specific heat or drought-stress genes.
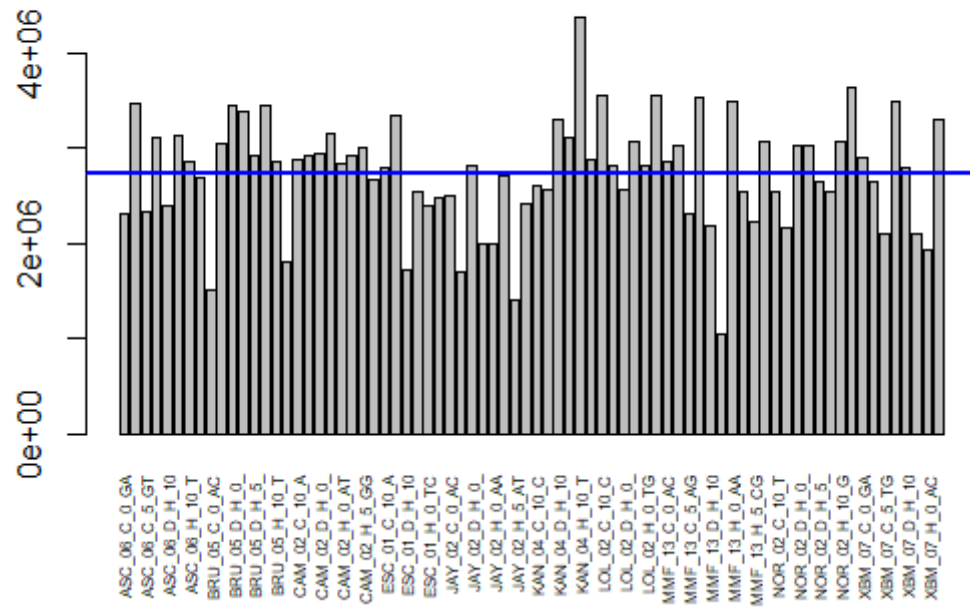
Figure 1: Mapped reads per sampled red spruce. Note that this data contains RNA samples from all three time points.
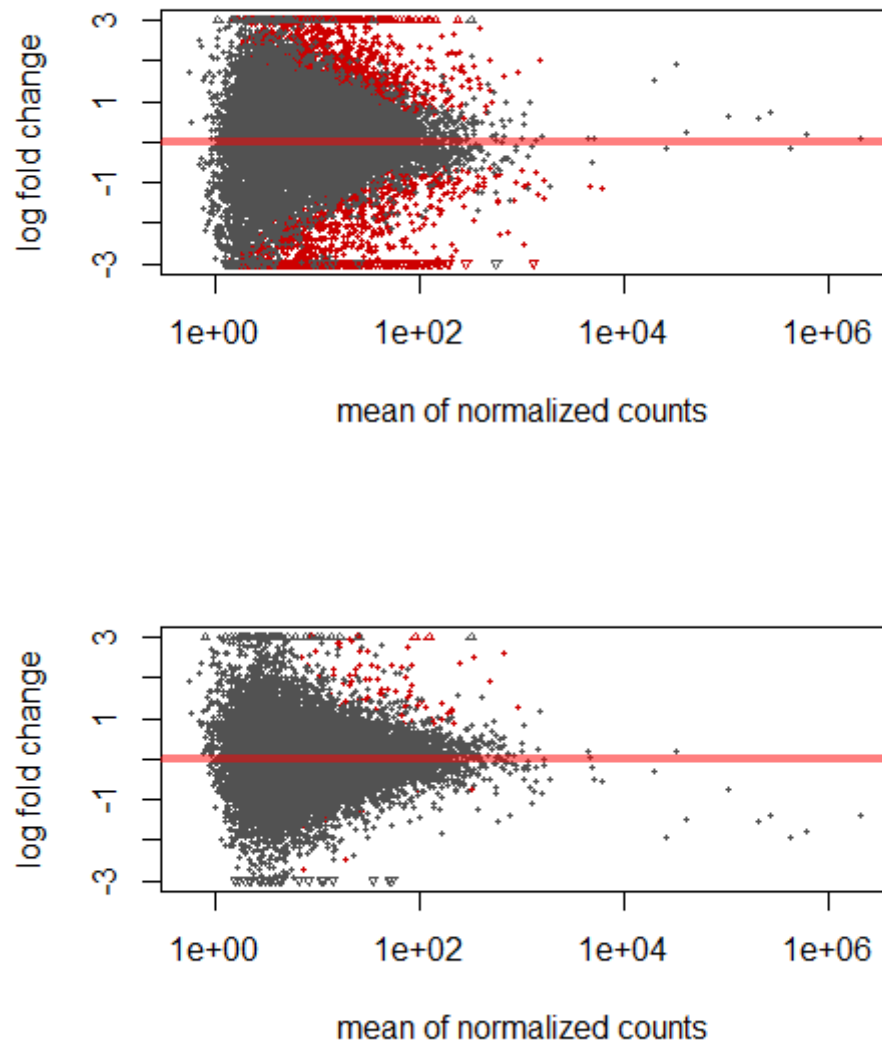
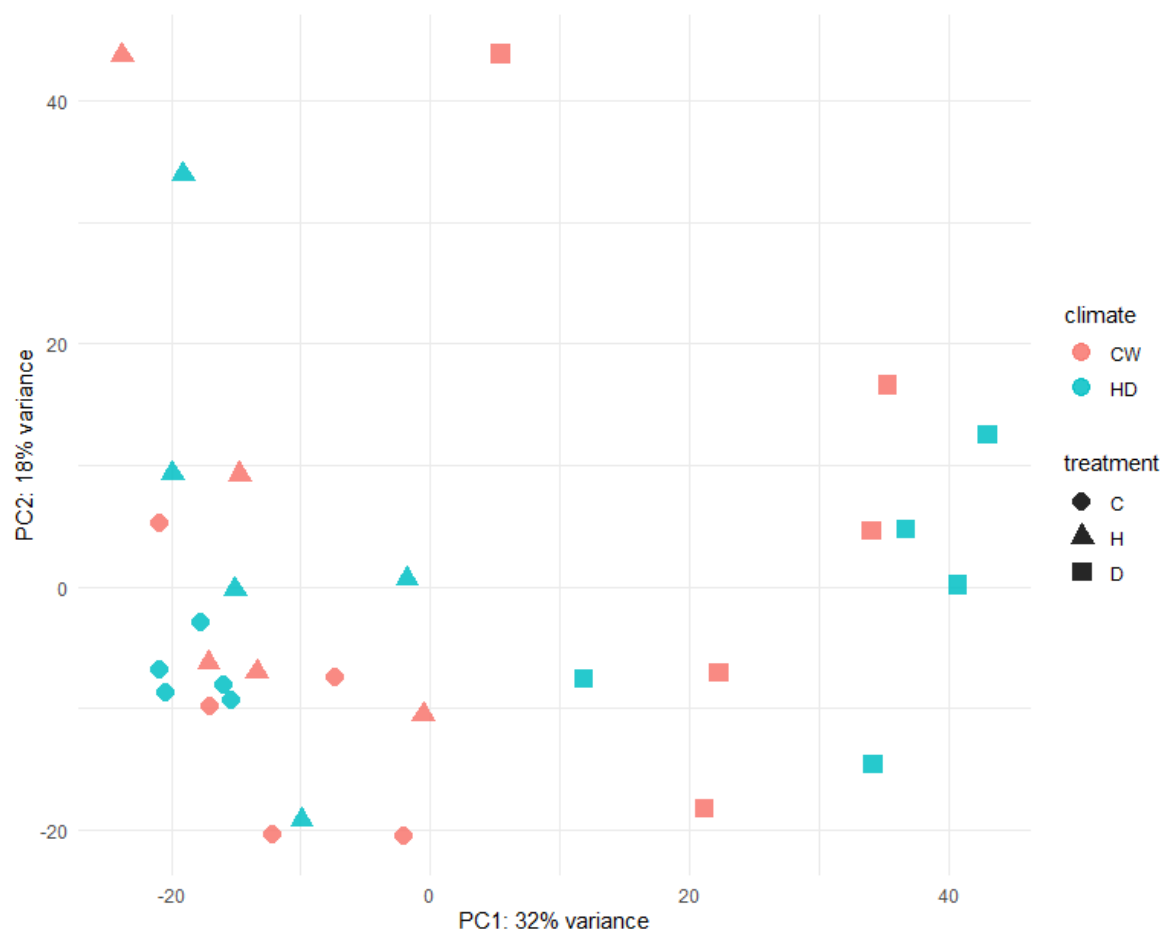Figure 2: MA plots of differential gene expression for C vs. D (top) and C vs. H (bottom).

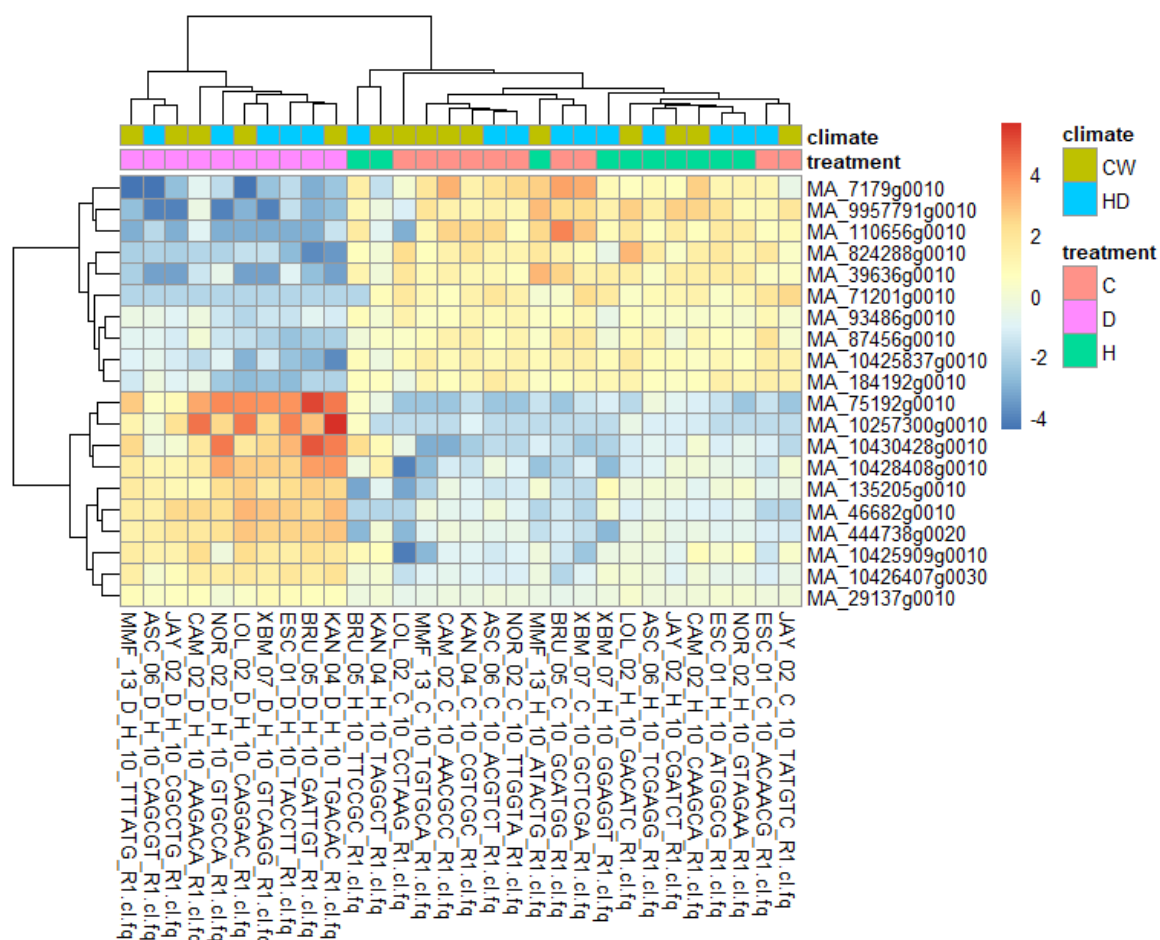Figure 3: Principle component analysis of sampled *P. rubens* transcriptomes.

Figure 4: Heatmap and clustering of differentially expressed *P. rubens* genes by sample.

1. Wingett SW, Andrews S. FastQ Screen: A tool for multi-genome mapping and quality control. F1000Res. 2018 Aug 24 [revised 2018 Jan 1] ↩

2. Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, btu170. ↩

3. Congenie.org ↩

4. Patro, Rob, et al. "Salmon provides fast and bias-aware quantification of transcript expression." *Nature Methods* (2017). Advanced Online Publication. ↩

5. Charlotte Soneson, Michael I. Love, Mark D. Robinson (2015): Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* ↩

6. Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, *15*(12), 550. ↩

7. Cohen, P. (2002). The origins of protein phosphorylation. *Nature cell biology*, *4*(5), E127-E130. ↩