Red Spruce Transcriptomics - Zoe Portlas

Background

Red Spruce (*Picea rubens*) is a montane tree native to the Eastern United States and Canada. Its habitat is usually characterized as cool and moist (Blum, 1990), but it can also exist in warmer and drier sites. In this experiment, sites were separated by their climate of origin either being cool and wet or hot and dry based on the precipitation of the warmest quarter and the temperature of the warmest month. The expression of individuals from these habitats was assessed in response to a heat and a heat and drought treatment.

There were ten maternal families in the experiment, five families from each source climate, and there were three treatment groups in the experiment. Seedlings were grown in a common garden environment for 10 weeks and then moved to growth chambers under control conditions for 2 weeks before beginning the 3 treatments. The control group was watered every day and had a 16:8 L:D photoperiod, meaning a cycle of 16 hours of light at 23 °C followed by 8 hours of dark at 17 °C. The heat group experienced the same watering and photoperiod, but there was a 50% increase of temperature, so light/day was at 35 °C and dark/night was at 26 °C. Finally, the heat/drought group experienced the same photoperiod and temperatures as the heat group, but in addition, water was withheld.

Tissue was sampled on days 0, 5, and 10 for each of the groups. RNA was extracted from the whole seedlings, which includes root, stem, and needle tissue. This analysis focuses on the samples from day 10, which had 5 reps per source climate and treatment (total n = 30). Samples were quantified for RNA quantity and concentration on a Bioanalyzer, which is a chip based capillary electrophoresis machine. Samples with more than 1 ng/μL were sent to Cornell for 3’ tag sequencing. Library prep followed the LexoGen protocol for 3’ tag sequencing, which is useful for detecting short transcripts.

Bioinformatics

The quality of the reads was assessed using FastQC before and after cleaning. Trimmomatic was used to trim the raw reads (Bolger et al., 2014). Trimmomatic cut Illumina specific adaptor sequences from the raw reads, cut bases below a quality score of 20 off the leading and trailing ends, and did a head crop to cut 12 bases from the start of the read. The program performed a sliding window trimming approach to clip the read when the quality within a window of 6 bases drops below an average quality of 20. If the read was below 35 bases, it was dropped.

Cleaned reads were mapped to the *Picea abies* reference transcriptome (Nystedt et al., 2013) and abundance was quantified using Salmon (Patro et al., 2017). Because 3’ RNASeq includes part of the untranslated region (UTR), the reads were mapping at a very low rate to a reference transcriptome that only included the exome (~2%). Instead, we mapped them to a reference that combined the UTR and the exome, which increased mapping to 40-70% of reads mapping across samples, with a mean of about 52%.

The counts matrix from Salmon was read into R in order to perform differential gene expression analysis with DESeq2 (Love et al., 2014). The data was subset to contain only the Day 10 samples and genes with fewer than 30 reads were filtered out, so there was an average of 1 read per sample. After filtering, there were 24,300 reads, reduced from 66,408. The experimental design was defined as ‘design = ~ climate + treatment + climate:treatment’. The results of these models showed the differentially expressed genes for different contrasts, which were visualized using MA Plots (Figure 2). Principal components analysis was performed (Figure 3).

Results

The number of reads per sample were visualized using a barplot (Figure 1) and the mean and median number of reads were found. The mean number of reads across all samples was 1296.096 and the median number of reads was 10. This shows that there was dispersion across different genes and large difference in magnitude of expression.

A picture containing screenshot

Description automatically generated

Figure 1. Barplot showing the number of reads per sample, only for samples taken on Day 10. Blue line shows the average number of reads across all samples.

MA Plots show the differentially expressed genes between contrasts in a model. Contrasts obtained through the interaction term of the model and the contrast between source climate had no significant differentially expressed genes, so only MA Plots for the contrasts hot/dry vs. control, hot vs. control are shown (Figure 2).

A screenshot of text

Description automatically generatedA close up of a map

Description automatically generated

Figure 2. MA plots showing genes up and down regulated in the two treatment groups compared to the control group.

Principal components analysis was performed on the differential expression model (Figure 3). PC1 explained 17% of the variance and PC2 explained 13% of the variance. There appears to be no grouping based on source climate (colors), but there appears to be grouping based on the treatment (shape), with the Hot and Dry treatment (squares) separating from the control and hot treatments over PC1.

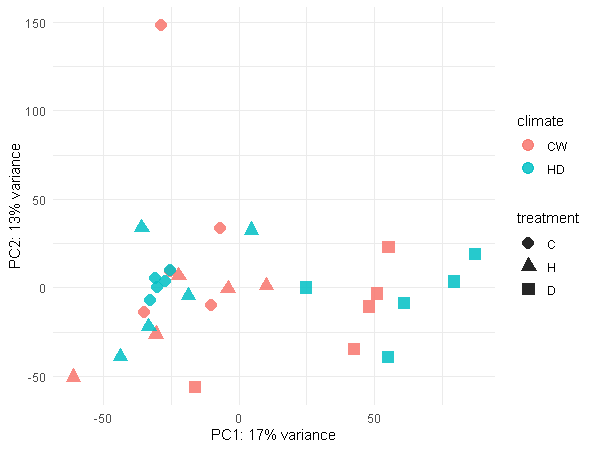


Figure 3. Principal components analysis of differential expression with color showing source climate (CW = Cool and Wet, HD = Hot and Dry) and shape showing treatment (C = Control, H = Hot, D = Hot and Dry).

Conclusion

While there was no significant differentially expressed genes depending on the source climate of the samples, there were significant differences in the expression between the treatments. 1 gene was significantly up regulated in the Hot treatment compared to the Control treatment, as shown in Figure 2. Between the Hot/Dry treatment and the Control treatment, however, there were many more differentially expressed genes, with 92 being up regulated and 174 being down regulated (Figure 2). Taken together, this suggests that the populations in different source climates are not differently adapted to those climates, because there is no differential expression when source climate is considered. However, Red Spruce did respond to the Hot and Hot/Dry treatments, but far more genes were differentially expressed between the Hot/Dry and Control treatments than the Hot and Control treatments.

The PCA (Figure 3) also supported those results. There was no separation of points by climate and the Control and Hot points were not separated, but the Hot/Dry points were separated along PC1 from the other two treatments. This could either suggest that Red Spruce is more tolerant of changes in temperature and is forced to respond to changes in water stress or that Red Spruce is more able to respond to changes in water stress with a plastic response than it is to changes in temperature.

One modification to this experiment that could better address the question of whether drought or heat is more stressful to Red Spruce would be to include a treatment group at the same temperature as the control, but with no water. Without this treatment group, it is not possible to tell whether the high levels of differential expression between Hot/Dry and Control are due to drought stress or due to a synergistic effect between drought and heat stress.

References

Bolger, A. M., Lohse, M., and Usadel, B. 2014. Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics, btu170.

Blum, Barton M. "Picea rubens Sarg. Red Spruce." Silvics of North America 1 (1990): 250-259.

Love MI, Huber W, Anders S. 2014. “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.” *Genome Biology*, 15, 550.

Nystedt, B., Street, N., Wetterbom, A. et al. 2013. The Norway spruce genome sequence and conifer genome evolution. *Nature* 497, 579–584.

Patro, Rob, Geet Duggal, Michael I. Love, Rafael A. Irizarry, and Carl Kingsford. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* 14: 417-419.