Background Red spruce (*Picea rubens*) is a species of conifers that thrives in humid, cool climates [1]. However, due to the ongoing climate change the average temperature has been increasing in eastern North America, where this species is native, and it is predicted to increase more [2]. It has been found in other species of plants that while heat can already significantly influence fidelity and trigger genome-wide transcriptional responses, the combined effect of heat with drought can be devastating as heat causes further evaporation [3]. Some populations of red spruce that are located at a lower latitude and/or elevation and/or on the south-facing slope experience longer periods of heat (and potentially drought), which resembles the predicted future climate, and thus individuals in these populations have potentially been selected for alleles causing a differential gene expression that can in turn enable higher resilience against these conditions. It is important to understand both how individuals of this species in general, and how individuals originating from populations experiencing different climates react to heat and drought in order to understand the molecular underpinnings of the evolution of resilience to future climate conditions and predict the ability of this already endangered species to survive.

In this study, we used ten maternal families, five from hot and dry and five from cool and wet climates. Seedlings were germinated and grown under common garden conditions for 10 weeks. Individuals originating from both climates were subjected to heat (50% increase in temperature compared to control) or heat + drought treatments (50% increase in temperature and complete water withholding). Tissues (root, stem, and needle) were collected on days 0, 5 and 10 for RNA extraction. Extracted RNA samples were quantified and quality checked using Bioanalyzer and concentrations above 1 ng/μl were sequenced using 3'-Tag RNA-seq on one lane of NextSeq500.

Bioinformatics Pipeline

Trimming To trim raw data and remove adapters Trimmomatic (version 0.33) was used [4]. As before, adapter and other illumina-specific sequences were cut out, bases below the threshold quality = 20 were cut off from the start and the end of a read and reads were scanned with a 6-base wide sliding window and were cut when the average quality per base dropped below 20. Finally, trimmed reads shorter than 35 were dropped. To further the quality of our data, we additionally removed 12 bases from the start of each read. Before trimming the average number of sequences per sample was 5,007,268.2, which decreased to 4,794,817.8 based on FastQC analysis (in the ASC population).

Mapping As reference transcriptome for *P. rubens* doesn't exist, cleaned reads from each sample were mapped to a reference transcriptome of a closely related species, Norway spruce (*P. abies*). Salmon, a "wickedly fast" command line tool was used to simultaneously map reads to the reference transcriptome and quantify the abundance of these mapped reads [5]. For this program to run, first we had to index the reference transcriptome using Salmon's indexing tool. The minimum length for a valid match was set to 31. We used the Selective alignment feature of the tool (set by --validateMappings), which enables a more sensitive scheme to be used and thus enhances quantification accuracy. Mapping to the reference containing only exons yielded low quality mapping, thus we decided to include 3'UTRs as well. This resulted in a much higher mapping quality, on average 52% (compared to 32.4% in the ASC population).

Differential analysis of count data (DESeq2) We used DESeq2, an R package that uses negative binomial distribution to tests for differential expression. This tool is excellent at accounting for small replicate numbers, large dynamic ranges and outliers and enables the users to calculate the strength, not just the presence of differential expression [6]. DESeq2 takes the counts data matrix as an input (matrix containing counts data from each sample in each gene), which was generated using the tximport R package [7]. Before running the program, we filtered the counts data matrix to contain only genes that have at least 1 read per sample on average. While this excludes gene from the analysis with very low levels of expression, it also cleans the data from potential noise and significantly increases the strength of the statistics as there are less variables to account for (number of genes to test decreased from 66408 to 23887). After the analysis, count data of the most significantly differentially expressed genes was visually inspected to confirm the pattern.

Results When all data is considered (~ climate + treatment + day), PCA results showed that the data is clustered along PC2 such that day 10 samples are different from sample at other time points. While climate of origin doesn't seem to have an effect, the heat+drought treatment is separated by PC2 as well (overlaps with day 10). Thus, we decided to analyse data from day 10 and look at the effect of treatment and climate

in that subset of the data. Despite this decision, I decided to include an interesting observation from using the whole dataset (design ~ climate + treatment + climate:treatment). Comparing the effect of climate of origin across all treatments, one gene was found to be significantly differentially regulated, **Class VII chitinase (**$p = 4.97e-10$). This gene's expression was much higher in individuals from the hot and dry climate (15.96 log fold change). Interestingly, there was a significant interaction of the effect of climate and treatment on the expression of this gene: it was only upregulated in individuals from the hot and dry climate under the heat+drought treatment ($p = 1.585e-8$).

In the dataset including data only from day 10, PCA showed a similar pattern as before; climate of origin didn't seem to have an effect while PC1 (32% variance) clustered samples from the heat+drought treatment. The heatmap reinforced this observation (Figure 1a). When comparing **climates of origin** across treatments, again, one gene was significantly differentially expressed: MA_129323g0010, which was downregulated in individuals from the hot and dry climate (p=0.011, foldchange = -21.55). Unfortunately, this gene is labelled as unknown on the Congenie database and using NCBI BLAST of the CDS didn't result in any hit either. When comparing the samples from the **heat+drought treatment** to the control, 257 gene were significantly upregulated and 330 were significantly downregulated in heat+drought. When comparing **heat treatment** to control instead, there were only 6 genes that were differentially regulated – 4 up and 2 down compared to control. Remarkably, 3 out of the 4 gene upregulated in heat treatment were genes encoding heat shock proteins (two KDa class I heat shock -like, one KDa class IV heat shock -like). The upregulation of these genes was independent of climate of origin and fascinatingly only the class I heat shock proteins were also upregulated in the heat+drought treatment, making them the only two gene in the dataset that overlapped between treatments (Figure 1b). There were **no significant interactions** between climate and treatment in this reduced dataset.

Conclusion The results of our data analysis clearly showed that the effect of treatment on gene expression was much larger than the effect of climate of origin (PCA, heatmap and number of genes differentially expressed when single factor is considered). We found only one gene, a chitinase, that was differentially expressed in individuals from different climates of origin. Chitinases were found to be upregulated in response to fungal infection and drought in $P. abies$ [8] and in a highly drought tolerant wild tomato but less so in a susceptible species of the same genus [9]. Since individuals from the hot and dry climate upregulated this gene in the heat+drought treatment much more than individuals from the cold and wet climate in the same treatment, this could be a signature of local adaptation. However, this should be further investigated and overall there was little evidence for the effect of climate of origin. There was a much stronger effect of treatment, and heat+drought had much bigger effect than heat alone. Interestingly, only two genes were differentially regulated in response to both heat and heat+drought, which is similar to what was found in *Populus simonii* were the number of overlapping heat-/drought-responsive genes is also small [10]. These two proteins were heat shock proteins belonging to the same class. Upregulation of heat shock proteins under drought stress was observed in *Pinus halepensis*, thus drought stress alone can result in this response [11]. In a future study, it would be useful to increase the sample size to pick up small but potentially important differences based on climate of origin. Also, drought stress response can be very different in different tissue types [12], so separately analysing them could be more informative.
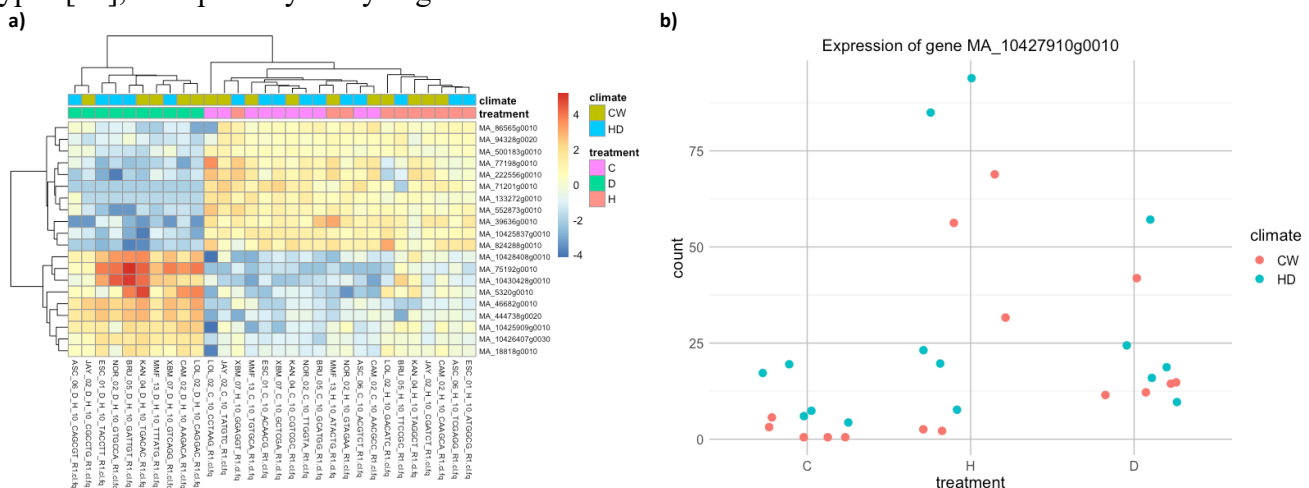

Figure 1: a) Heatmap of counts data of genes in control (C), heat (H) and heat+drought (D) treatments in individuals from coldwet (CW) and hotdry (HD) climates. There is a clear significant effect of the heat+drought treatment. In contrast, when testing the effect of heat alone or climate, there is not clear pattern. b) Expression of one of the KDa class I heat shock-like proteins in treatments and climates as mentioned in a). Climate of origin has no effect on expression, and it is upregulated in both heat and heat+drought treatments. Both a) and b) are of day 10.

References

1. Nowacki, Gregory, Robert Carr, and Michael Van Dyck. "The current status of red spruce in the eastern United States: distribution, population trends, and environmental drivers." (2010): 140-162.

2. Ribbons, Relena Rose. "Disturbance and climatic effects on red spruce community dynamics at its southern continuous range margin." *PeerJ* 2 (2014): e293.

3. Büntgen, U. L. F., et al. "Growth/climate response shift in a long subalpine spruce chronology." *Trees* 20.1 (2006): 99-110.

4. Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. "Trimmomatic: a flexible trimmer for Illumina sequence data." *Bioinformatics* 30.15 (2014): 2114-2120.

5. Patro, Rob, et al. "Salmon provides fast and bias-aware quantification of transcript expression." *Nature methods* 14.4 (2017): 417.

6. Love, Michael I., Wolfgang Huber, and Simon Anders. "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." *Genome biology* 15.12 (2014): 550.

7. Soneson, C., M. I. Love, and M. D. Robinson. "Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res 4: 1521." (2015).

8. Nagy, Nina Elisabeth, et al. "Effects of Rhizoctonia infection and drought on peroxidase and chitinase activity in Norway spruce (Picea abies)." *Physiologia Plantarum* 120.3 (2004): 465-473.

9. Yu, Long-Xi, et al. "Chitinase: differential induction of gene expression and enzyme activity by drought stress in the wild (Lycopersicon chilense Dun.) and cultivated (L. esculentum Mill.) tomatoes." *Journal of plant physiology* 153.5-6 (1998): 745-753.

10. Jia, Jingbo, et al. "Comparative transcriptomic analysis reveals the roles of overlapping heat-/drought-responsive genes in poplars exposed to high temperature and drought." *Scientific reports* 7 (2017): 43215.

11. Fox, Hagar, et al. "Transcriptome analysis of Pinus halepensis under drought stress and during recovery." *Tree physiology* 38.3 (2018): 423-441.

12. Haas, Julia Christa, et al. "Candidate regulators and target genes of drought stress in needles and roots of Norway spruce." *bioRxiv* (2019): 517151.