

TRANSCRIPTIONAL RESPONSE TO  
DROUGHT STRESS IN *Eutrema salsugenum*

YASSER SALAMA

McMaster University  
Hamilton, Ontario, Canada, L8S 4E8

December 17, 2015

BIO 720 | Bioinformatics  
Submitted to:  
Drs. DWORKIN, EVANS, AND GOLDING

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Methods</b>	<b>3</b>
2.1	Plant Growth and Drought Simulation . . . . .	3
2.2	Sequencing and RNA-seq Pipeline . . . . .	4
<b>3</b>	<b>Results</b>	<b>5</b>
3.1	Sample Information and Read Statistics . . . . .	5
3.2	Batch Effects . . . . .	8
3.3	Transcriptional Response to Drought . . . . .	11
<b>4</b>	<b>Discussion</b>	<b>14</b>

## List of Figures

1	Brassicaceae Phylogeny . . . . .	2
2	Shandong and Yukon ecotype morphology . . . . .	3
3	Read Processing Statistics . . . . .	7
4	Cluster analysis with PCA . . . . .	9
5	Heatmap analysis . . . . .	10
6	Expression response plots . . . . .	12

## List of Tables

1	Sample Names and Information . . . . .	6
2	Differentially expressed gene counts . . . . .	11
3	Differentially expressed genes in response to drought . . . . .	13

## Preface

The GitHub repository for this report is located [here](#). The repository is set up to contain only the shell, perl, and R scripts used for generating all of the results, as well as the text files containing the count data. The repository does not contain the read files or any subsequent files associated with the reads. The repository is not a clone of the directory on the cluster, however, all the files on the cluster can be accessed (read permission) at /home-/yasser/bio720/final\_project.

When explaining any method or result generated through a script, I will refer to the script in brackets by its names in the scripts folder of the GitHub repository. All the scripts are commented and contain more information about the individual commands than the report will give.

## 1 Introduction

The prevention of crop loss in the horticulture and agriculture sectors has been a long-standing concern because of the significant economic and societal impacts these losses have. The agriculture industry makes up almost 7% of Canada's gross domestic product, and with a growing world population, food and water scarcity, and rising global temperatures, these effects will become magnified [1, 2, 3].

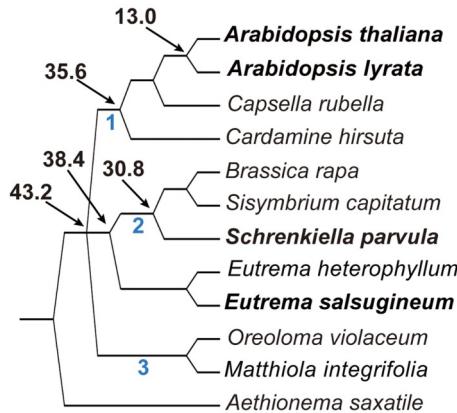
Abiotic stressors such as drought, salinity, and extreme temperature variations are major drivers of crop loss and can result in 50% to 80% loss of crop yield, depending on the crop and growth stage [4]. Drought is a major facet of abiotic stressors which accounts for approximately 40% of the indemnities paid for crop loss, larger than any other stressor. As well, in certain crops such as maize and rice, drought alone can account for up to 80% of the yield loss [5]. These are extravagant losses with far-reaching impacts. It is therefore imperative to examine the mechanisms by which drought stress can impact plants and their health, and how resistance mechanisms are adapted to such conditions.

The plant response to drought stress is well documented, albeit not fully characterized. In general, the reduction in water availability limits turgidity, inhibiting cell elongation and negatively impacting morphology [5]. As well, limited nutrient availability, photosynthetic deficits, and consequently diminished mitosis directly impact the health of the plant. The mechanisms by which plants can resist against drought can vary greatly, but typically, there is an attempt to reduce the water requirement through reduced leaf number and surface area, and maximize water uptake through root proliferation. Physiological responses also result in protection against oxidative stress which arise from the generation of reactive oxygen species. These mechanisms involve elevated expression of enzymatic constituents for oxidative stress protection [5]. Overall, the stress response and resistance mechanisms are extremely complex, and encompass a variety of biological mechanisms.

The complexity of the stress response is supplemented by the finding that over-expression of known stress-responsive genes can sometimes have little effect on mitigating the stressors' impact [6]. This is likely attributed to the network and co-regulatory nature of the stress response, and requires a more comprehensive understanding of the underlying molecular and physiological changes which occur in response to stress. A useful tool for understanding the stress response and the mechanisms by which it manifests is to examine the complete transcriptional regulation network and the changes that occur once a stress is applied. For the example of drought stress, restriction of

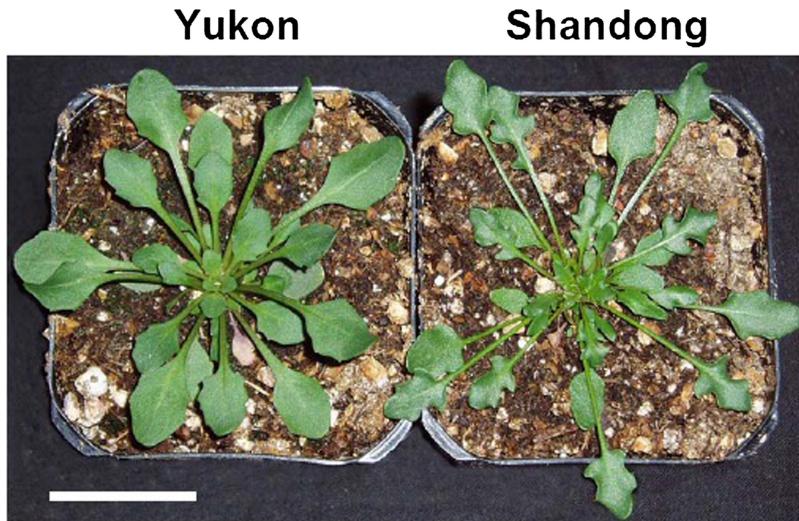
water and subsequent comparative examination of the transcriptional changes which occur may provide great insight into the mechanisms for the stress response, and in the case of a drought-tolerant organism, the drought resistance mechanism.

*Eutrema salsugineum* is a halophytic plant, capable of thriving in various abiotic stress conditions, and is an ideal organism to study the response and resistance mechanism to drought. *E. salsugineum* is closely related to the well studied model plant *Arabidopsis thaliana*, with an estimated divergence time of around 43 million years (Figure 1). An important physiological adaptation in *E. salsugineum* is the presence of a low permeability cuticle, which reduces the loss of water through the leaves. This waxy cuticle has been observed to be much more abundant in *E. salsugineum* than *A. thaliana*, suggesting that this is an adaptive trait for drought resistance.



**Figure 1** Brassicaceae phylogeny inferred with two markers, *ndhF* and *PHYA* [7].

Intriguingly, two different ecotypes of *E. salsugineum*, Shandong and Yukon, exist, with different standard environmental conditions. Whereas the Yukon ecotype is sustained in sub-arctic temperatures and a semi-arid climate, the Shandong region is more temperate, although both regions feature a high salinity soil. These two ecotypes show distinct responses to drought stress [8]. Previous findings have shown that the Shandong ecotype seems to have a delayed response to drought and differential changes in cuticle permeability traits between the two ecotypes [9].



**Figure 2** Shandon and Yukon accessions of *E. salsugenum* [9].

The underlying reasons for the differential stress response and resistance of these two ecotypes is not well characterized, and provides a great opportunity to possibly elucidate the molecular mechanisms through which drought is tolerated in plants. By examining the transcriptional adjustments *E. salsugenum* undergoes in response to stress, we may begin to understand what pathways are important for this process. By examining two different ecotypes, we can also identify common mechanisms and mechanisms unique to each accession, which is quite informative.

In this work, we examine the transcriptional response of both *E. salsugenum* ecotypes to drought through RNA-seq analysis of Shandon and Yukon ecotypes of *E. salsugenum* in well-watered and drought treated conditions (experimental design and processing provided by Weretilnyk lab). We also examine how the recovery of these plants manifests through a re-watering phase.

## 2 Methods

### 2.1 Plant Growth and Drought Simulation

*E. salsugenum* seeds corresponding to Yukon (Y) and Shandong (S) accessions were grown at 4°C for 4 and 7 days, respectively, to synchronize germination. Subsequently, pots were transferred to growth cabinets simulating standard growth conditions with an optimal watering schedule. The plants were grown under these conditions for 4 weeks after which the plants underwent exposure

to drought simulation through complete water-restriction. Effects of the water-restriction were measured gravimetrically using the fraction of transpirable soil water (FTSW) metric. When the fraction reached 0% (i.e. no remaining water in soil), the plants visibly wilted. Re-watering commenced on the day of wilting to 50% of the soil capacity, followed by full saturation the next day. Tissue from plants undergoing normal growth conditions (WW, well-watered), drought conditions (D, drought), and re-watering conditions (RW, re-watered) was collected and total RNA was then isolated and processed for sequencing.

the genome is shandong. checked reads mapped. only looked at leaves not roots which have the most response to drought 243 mb genome (check this) overexpression of gene induced in stress response has little effect, so its complex. critical reviews in plant sciences coverage of mapped reads?

## 2.2 Sequencing and RNA-seq Pipeline

Two replicates of each treatment and ecotype were used in this analysis, resulting in 12 total samples. Sequencing was completed on a HiSeq 2000 in four batches. A subset of the samples were split onto two lanes, while the rest were run solely on one lane (multiple lane samples are denoted by L1 and L2, see Table 1).

FastQC (v0.11.3) was used to validate read metrics prior to and after trimming [10] (script: Step1\_Concat\_trim.sh). A simple loop through all the generated FastQC output files using

```
more */summary.txt | grep FAIL
```

from the fastqc\_reports/pre\_process/ and fastqc\_reports/post\_process/ folders provides a crude way of seeing which metrics fail.

Quality trimming and adapter removal was accomplished with BBduk from the BBMap suite [11] (script: Step1\_Concat\_trim.sh). The reference genome and annotation file for *E. salsugenum* [7] were downloaded from Phytozyme (v10.3) [12]. The reference genome is 241 Mb and corresponds to the Shandong ecotype. STAR (v2.3.2) was used to index the reference genome of *E. salsugenum* and map the quality filtered reads to the reference [13] (scripts: Step2\_Index\_ref.sh, Step3\_Align.sh) (Figure 3). The alignment file was outputted as a bam file already sorted by coordinate through the flag:

```
--outSAMtype BAM SortedByCoordinate
```

The bam file generated is in proper format for use with htseq-count and mitigates the need for subsequent sam to bam conversion and sorting. HTSeq-count(v0.6.1) was used to obtain counts for genes in the genome [14] using both union mode (scripts: Step4\_Count\_union).

Read statistics for number of reads prior to and after trimming, as well as number of successfully mapped, were obtained (scripts: trim\_stats.sh, mapped\_stats.sh, read\_statistics.R).

Counts were imported into R (v3.2.2) [15] and differential gene expression analysis was conducted using the DESeq package (v1.10) [16] (script: diff\_gene\_exp.R). All adjusted p-value cutoffs were set to 0.01.

Gene id conversion to gene descriptions was accomplished with the annotation\_info file provided by Phytozyme. Genes of interest from the contrast analysis were put in a file and using grep, were searched for in the annotation file to obtain the gene description. To search for abscisic acid and dehydrin related genes, a simple grep command,

```
grep -e abscisic -e Abscisic -e dehydrin -e  
Dehydrin Esalsugineum_173_v1.0.  
annotation_info.txt >  
abscisic_and_dehydrin_genes.txt
```

was used. The gene ids were extracted using a simple cut -f 3 command and then searched for in the DESeq object.

## 3 Results

### 3.1 Sample Information and Read Statistics

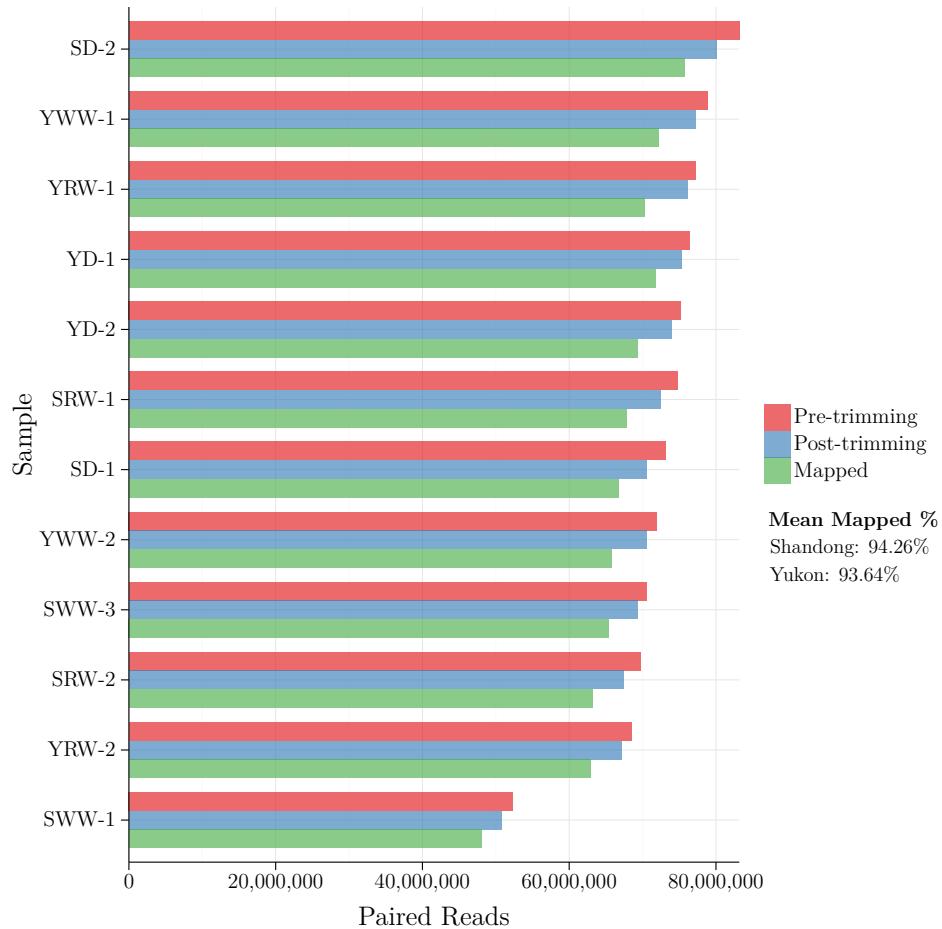
The samples used in this study are summarized in Table 1. Since lane effects may introduce a confounding factor and impact results, all the samples which were run on two lanes were left separated. Another potential confounding factor present in the data is the fact that all the samples weren't run together (i.e. run in different batches). Additionally, the only samples run in batch 4 were SRW-1 and SRW-2, which makes it very difficult to rule out whether this batch introduced a systematic variability not due solely to the sample conditions alone. Although running on different lanes may introduce a confounding factor, if all the samples within a batch were evenly divided between lanes, or all samples are run on the same lane, then that effect is essentially negligible. This is the case with these samples. Within every batch that has samples run on two lanes, all the samples in that batch were split between those lanes (e.g. batch 3 and 4). As well, batch 2 samples were all run in one lane. Therefore, this lane effect within each batch doesn't need to be accounted for and only the batch effect is what matters. There is a slight caveat to this; since batch one has two samples run on lane 3 and one sample run on lane 5, there might be a possible lane effect there. There isn't really anyway to test for this lane

effect because there isn't enough samples across the lanes, so it will just be ignored.

Sample	Ecotype	Treatment	Replicate	Lane	Batch
SWW-1	Shandong	Well-watered	1	2	2
SWW-3.L1	Shandong	Well-watered	3	1	3
SWW-3.L2	Shandong	Well-watered	3	2	3
SD-1	Shandong	Drought	1	2	2
SD-2	Shandong	Drought	2	2	2
SRW-1.L1	Shandong	Re-watered	1	1	4
SRW-1.L2	Shandong	Re-watered	1	2	4
SRW-2.L1	Shandong	Re-watered	2	1	4
SRW-2.L2	Shandong	Re-watered	2	2	4
YWW-1	Yukon	Well-watered	1	5	1
YWW-2.L1	Yukon	Well-watered	2	1	3
YWW-2.L2	Yukon	Well-watered	2	2	3
YD-1	Yukon	Drought	1	3	1
YD-2.L1	Yukon	Drought	2	1	3
YD-2.L2	Yukon	Drought	2	2	3
YRW-1	Yukon	Re-watered	1	3	1
YRW-2.L1	Yukon	Re-watered	2	1	3
YRW-2.L2	Yukon	Re-watered	2	2	3

**Table 1** Sample names and information. Each combination of ecotype and treatment had two replicates, with some samples being run across two lanes. All samples were collectively run in four batches.

The samples averaged 72,653,152 paired reads prior to trimming (Figure 3) and FastQC analysis as described in Section 2 suggests that read quality was a “pass”, even prior to trimming. The only metrics which are considered as “fail” by FastQC are “Sequence duplication levels” and metrics associated with kmer and sequence content. The duplication levels are expected for RNA-seq data as many copies of the same transcript will be present. The kmer and sequence content also seems to be an artifact of the library preparation method, which introduces nucleotide biases at the beginning of the sequence. More importantly, the measures of sequence quality and adapter contamination pass in all of the samples.



**Figure 3** Read processing and mapping. Values are given as pairs of reads. Samples run on multiple lanes are left separated. The mean percentage of trimmed reads which were successfully mapped is indicated for each ecotype.

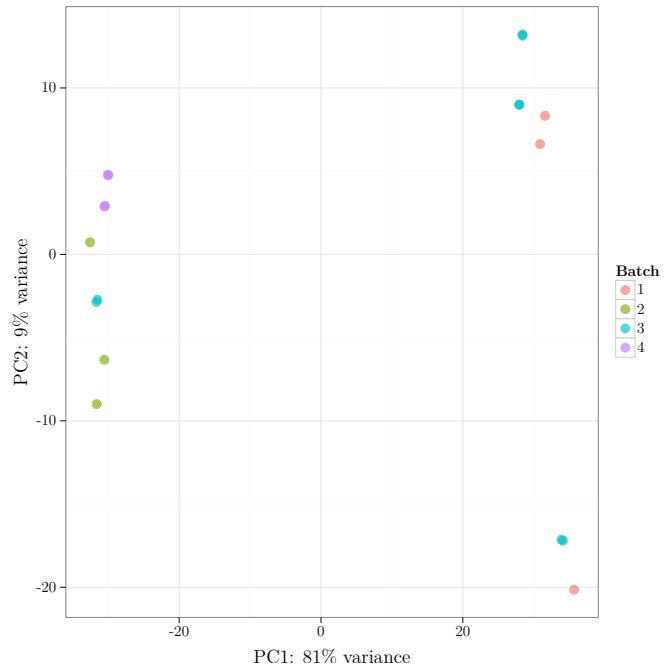
Quality trimming of the samples confirmed that the data was relatively high quality (only 2.4% reads lost due to trimming). Further, mapping was relatively successful, with an average of 94% of trimmed reads successfully mapping. Since the reference genome was generated using a Shandong ecotype plant, I checked whether there was a difference in the percentage of successfully mapped reads between the Shandong and Yukon samples. It might be expected that Shandong samples would show greater percentage of mapped reads, but there wasn't enough evidence to say that the two ecotypes mapped with different efficiency ( $p > 0.1$ ).

### 3.2 Batch Effects

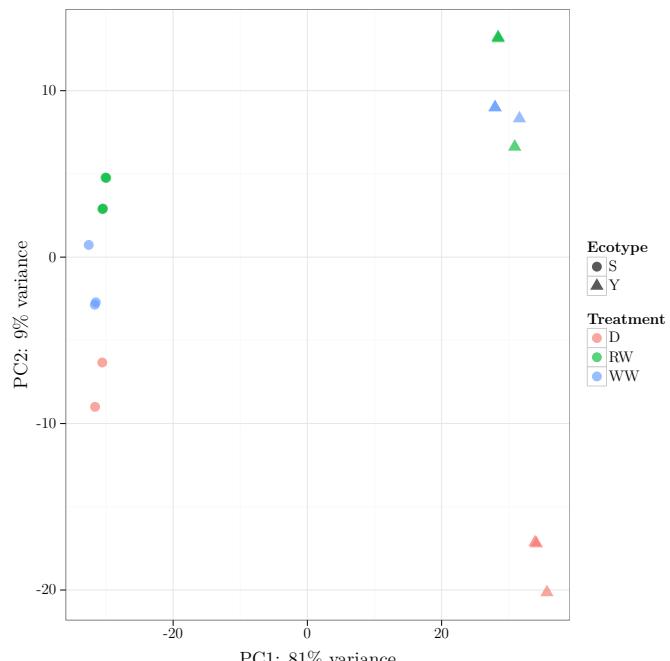
Since the samples for this study were not all run at the same time, there may be a batch effect which introduces bias within the samples. Unfortunately, the batches are not well mixed (i.e. there isn't a good representation of different samples within each batch), which makes it hard to delineate the true differences from the batch differences. However, we can still perform a batch effect analysis and use a conservative approach by removing any genes which show differences due to the batch (script: `diff_gene_exp.R`). When comparing all the samples and batches, we see quite a high number of genes showing differential expression due to batch ( 23% with adjusted p-value < 0.1). However, since batch 4 consists of only Shandong re-watered plants, this might be inflating this value because the batch effect is inherently intertwined with the possible effects of the actual variables we are interested in (ecotype and treatment). I generated a subsetted dataset which removes the SRW samples from the analysis, and re-ran the batch effects just to see how much of a difference these samples would make. When the SRW samples (which are the only samples in batch 4) are removed from the analysis, 1.3% genes are showing a difference in expression due to batch. This suggests that the effect is likely magnified because batch 4 contains samples of one type. A way to approach this is to remove the 1.3% genes showing differential expression from further analysis, and then to be wary of interesting genes which are part of the 23% of genes found earlier.

Before removal of the trouble genes, we can perform some cluster analysis to get an idea of our samples and crudely visualize whether there is batch effect. Figure 4a provides a principal components analysis (PCA) of the samples, labelled by batch. Although the samples aren't distinctly clustering due to batch, there isn't even mixing between the batches either. However, this would be expected due to the lack of mixing between the samples. For instance, if all the samples in one batch correspond to one ecotype, it would not be expected that this batch's samples would cluster with a batch containing samples of the other ecotype. This is more evident in Figure ?? where the samples are labelled based on ecotype and treatment. We can see a clear separation of the ecotypes on PC1, and provides a more reasonable explanation for the clustering pattern seen in Figure 4a. PC2 is able to slightly separate the samples based on treatment condition, but the low variance explained suggests that the ecotype is the major factor in the differences between the samples.

The samples can also be visualized using a heatmap (Figure 5.) This further presents the separation of the samples based on ecotype, with the Shandong and Yukon ecotype samples clustering together. As well, the SRW



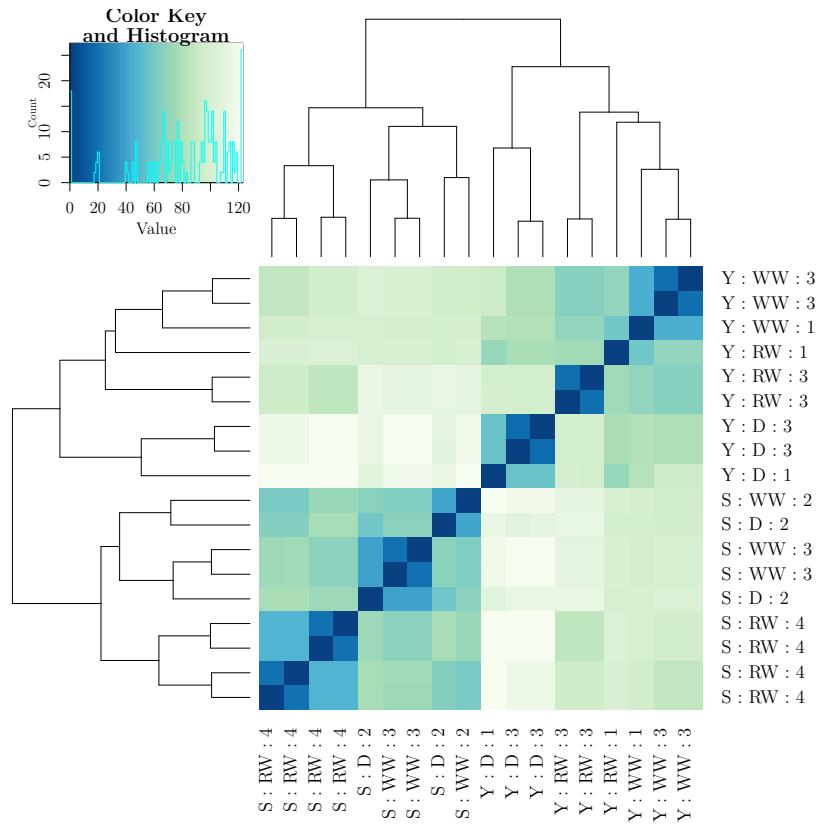
(a)



(b)

**Figure 4** Principal components analysis of samples labelled based on **a)** batch or **b)** ecotype and treatment

samples show a high degree of similarity to each other and form a distinct cluster from the other samples. Again, we cannot be certain of the degree to which batch effects may be playing a role in the separation of the SRW samples, but the pattern of samples clustering together based on the variables of interest and not batch is a good sign. Interestingly, an SWW samples which was run in batch 2 clustered more closely with an SD sample which was also run on batch 2 rather than on the other SWW samples run on batch 3. This hints at the possibility of a batch effect and supports the removal of the 1.3% genes discussed earlier.



**Figure 5** Heatmap analysis. The heatmap shows a measure of similarity between samples and provides a tree based on the similarity of the samples. The ecotype, treatment, and batch are indicated for each sample.

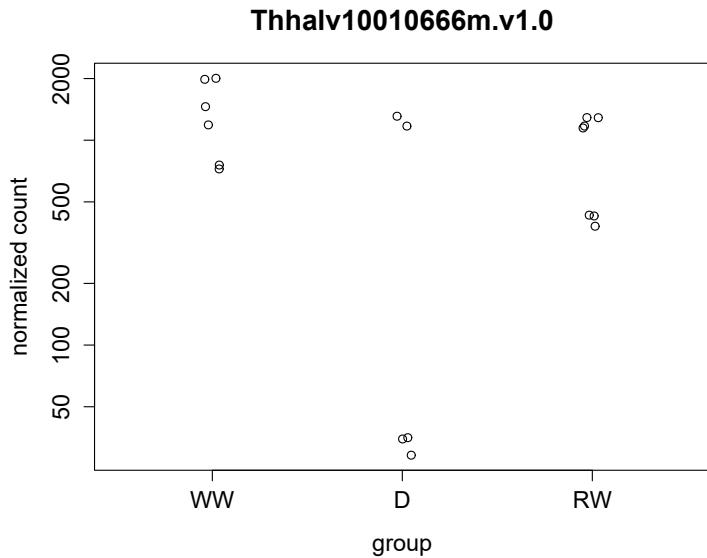
### 3.3 Transcriptional Response to Drought

Since there are two ecotypes and three treatments, there are a few things to explore in terms of differential gene expression. Firstly, the difference between the Shandong and Yukon ecotype can be examined to see basal differences in gene expression across all treatments (Table 2 row 1). As well, we can examine the overall effect of drought and re-watering conditions in both ecotypes (rows 2 and 3). More interestingly, however, we can examine the difference in response between the ecotypes to drought or re-watering (Table 2 rows 5 and 6). These final two comparisons are more informative because a gene may show significant differential expression in response to a treatment in only one ecotype, but not both (see Figure 6). Although there may be genes which show differential expression in response to drought (row 2), these genes may be called as differentially expressed even when only one ecotype shows a great response while the other does not.

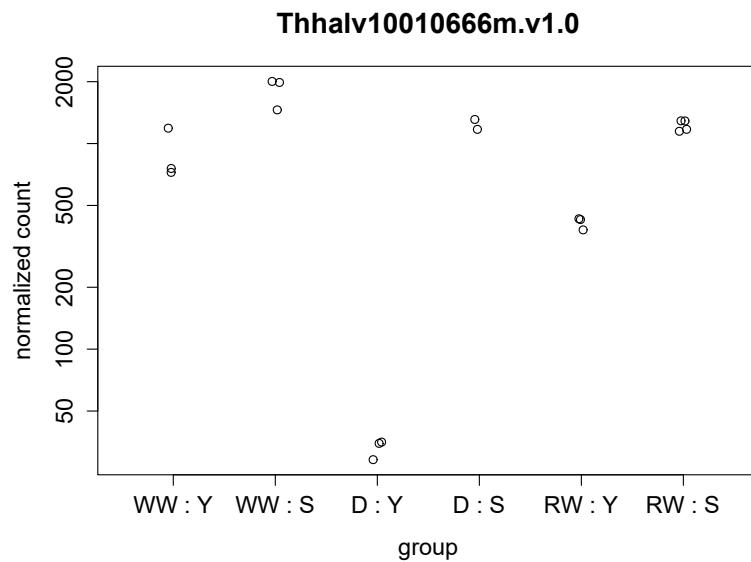
Comparison	LFC > 1	LFC < 1
Shandong vs Yukon	1824	1697
Drought vs Well-watered	1740	2553
Re-watered vs Well-watered	364	631
Shandong vs Yukon response to:		
Drought	661	269
Re-watering	209	313

**Table 2** Differentially expressed gene counts. Number of genes showing log 2 fold changes greater than 1 (upregulated) or less than 1 (downregulated) for multiple comparisons at an adjusted p-value cutoff of 0.01. Comparisons are relative to Yukon and well-watered values. Gray coloured rows show the number of genes which showed a different response between Shandong and Yukon in reaction to either treatment.

From Table 2, we can see that there are clear differences between the Shandong and Yukon plants in terms of overall gene expression, with Shandong showing about an even number of genes which are upregulated and downregulated compared to Yukon. As well, the response to drought seems to show more downregulation of genes than upregulation, and similarly with the re-watered condition. There is a possibility that genes which are downregulated due to drought persist in that state after re-watering. This could be checked by examining the intersect of genes which are down regulated in both conditions (not done in this analysis).



(a) Drought vs well-watered



(b) Shandong vs Yukon response to drought

**Figure 6** An example of the utility of the interaction term and looking at differential response to a treatment. In this example, the same gene is shown to be differentially expressed in the drought vs well-watered model, but when looking at the interaction term of ecotype and treatment, we can see that the gene is only responsive in the Yukon ecotype and not Shandong.

To examine how Shandong and Yukon react differently to the drought treatment, I looked at the top 20 genes (based on p-value) which showed differential expression between the two ecotypes to drought (Table 3). Since only 20 genes are looked at here, it is hard to draw general trends about the differences. It would be better to examine the complete list of genes and some form of functional enrichment analysis to categorize what functions are altered. However, we can still see some genes which are known to function in the drought response, such as a glutaredoxin family protein and a LOB domain-containing protein. The glutaredoxin protein fits in line with the idea of oxidative stress being a consequence of drought stress, and has been shown to regulate the plant response to stress [17]. As well, previous characterization of a LOB-like transcription factor was shown to play a part in salt stress [18]. These are potential pathways or mechanisms by which the drought response differs between the two ecotypes, as these are genes which were shown to show greater expression in Shandong than Yukon in response to drought.

Gene ID	LFC	Gene Description
Thhalv10028838m	1.71	ATP synthase delta-subunit gene
Thhalv10028074m	3.37	(bHLH) DNA-binding family protein
Thhalv10026359m	2.08	Glutaredoxin family protein
Thhalv10025559m	-3.76	leucoanthocyanidin dioxygenase
Thhalv10019053m	2.61	Pathogenesis thaumatin superfamily protein
Thhalv10018763m	3.91	Galactose oxidase repeat superfamily protein
Thhalv10018663m	2.20	B-box type zinc finger protein with CCT domain
Thhalv10017988m	3.09	Unknown
Thhalv10014990m	2.56	Unknown
Thhalv10014734m	5.70	Unknown
Thhalv10013774m	7.20	Eukaryotic aspartyl protease family protein
Thhalv10010666m	4.25	LOB domain-containing protein 38
Thhalv10009301m	-6.47	myo-inositol oxygenase 1
Thhalv10009104m	3.17	photosystem II reaction center W
Thhalv10008980m	4.67	Glutaredoxin family protein
Thhalv10008115m	3.03	myb-like tf family protein
Thhalv10007302m	-4.51	Major facilitator superfamily protein
Thhalv10005532m	4.27	basic leucine-zipper 1
Thhalv10004468m	-3.48	Adenosylmethionine decarbox family protein
Thhalv10003191m	-2.87	dual specificity kinase 1

**Table 3** Differentially expressed genes between ecotypes in response to drought

Known mechanisms of drought response include the abscisic acid and dehydrin response [19]. I wanted to examine whether there was any evidence for these response differing between the two ecotypes. The only gene showing a significant log 2 fold change (p-value < 0.01) was Thhalv10004906m, a dehydrin family protein. This protein had a log 2 fold change of -2.4, indicating that it was expressed greater in the Yukon ecotype than the Shandong ecotype in response to drought. It would be interesting to perform a time series analysis of the various dehydrin proteins between the ecotypes to see whether it is expressed earlier and in greater quantity in the Yukon ecotype, which might explain the delayed response to drought seen in Shandong.

## 4 Discussion

The ability to detect differentially expressed genes between samples reliably depends heavily on a non-confounding experimental design [20]. In this dataset, samples were run in different batches and sometimes within separate lanes in the same batch. These factors introduce unwanted variability and makes it hard to deduce true differences from differences due to technical variations. In this analysis, I attempted to control for this by removing genes which showed differential gene expression due to batch effect. Another approach to this problem is to put each sample into a unique factor and use that in the formula design of DESeq. In future analysis, the DESeq pipeline should be re-run with this modification. Despite this, the cluster analysis provided in Figures 4 and 5 in combination with the removal of problematic genes provided a foundation with which to identify differentially expressed genes relatively accurately.

The differential expression analysis revealed various genes which showed variability due to ecotype, treatment, and the interaction between these two variables. The importance of using the interaction term is emphasized in 6 where the same gene show variability due to both ecotype and treatment more so than can be explained by either factor alone. This type of interaction analysis revealed various genes which distinguished the Yukon and Shandong ecotypes when exposed to drought conditions. An important step in providing a more comprehensive analysis of these differences would be an examination of functional enrichment using all the significant differentially expressed genes. This might yield more clues for the discrepancy in drought response between the two ecotypes because looking at a small subset of genes such as in Table 3 is very limited.

The dataset presented in this report is quite extensive. Not all the analysis I wanted to accomplish could be completed for the report, and many things

can still be explored. It would be interesting to see the overlap in genes which were similarly regulated in the drought condition and re-watered condition in comparison to the well-watered condition. This would reveal genes which show a more persistent change even in the re-emergence of water and may reveal genes which could protect against future droughts. Analyzing the interaction term for ecotype and the re-watering condition would also reveal how the Yukon and Shandong ecotypes might differ in their recovery from drought. This might, for instance, show that one plant can recover to the normal state faster, but the other ecotype might be more resistance for the next drought.

This report has shown the power and possibilities of RNA-seq analysis

## References

- [1] Government of Canada. An overview of the canadian agriculture and agri-food system 2015, August 2015.
- [2] S Asseng, F Ewert, P Martre, RP Rötter, DB Lobell, D Cammarano, BA Kimball, MJ Ottman, GW Wall, JW White, et al. Rising temperatures reduce global wheat production. *Nature Climate Change*, 5(2):143–147, 2015.
- [3] Jacob Schewe, Jens Heinke, Dieter Gerten, Ingjerd Haddeland, Nigel W Arnell, Douglas B Clark, Rutger Dankers, Stephanie Eisner, Balázs M Fekete, Felipe J Colón-González, et al. Multimodel assessment of water scarcity under climate change. *Proceedings of the National Academy of Sciences*, 111(9):3245–3250, 2014.
- [4] John S Boyer. Plant productivity and environment. *Science*, 218(4571):443–448, 1982.
- [5] M Farooq, A Wahid, N Kobayashi, D Fujita, and SMA Basra. Plant drought stress: effects, mechanisms and management. In *Sustainable Agriculture*, pages 153–188. Springer, 2009.
- [6] Dorothea Bartels and Ramanjulu Sunkar. Drought and salt tolerance in plants. *Critical Reviews in Plant Sciences*, 24(1):23–58, 2005.
- [7] Ruolin Yang, David E Jarvis, Hao Chen, Mark A Beilstein, Jane Grimwood, Jerry Jenkins, ShengQiang Shu, Simon Prochnik, Mingming Xin, Chuang Ma, et al. The reference genome of the halophytic plant eutrema salsugineum. *Frontiers in plant science*, 4, 2013.
- [8] Xiaojing Xu, Jinchao Feng, Shiyou Lü, Greg T Lohrey, Huiling An, Yijun Zhou, and Matthew A Jenks. Leaf cuticular lipids on the shandong and yukon ecotypes of saltwater cress, eutrema salsugineum, and their response to water deficiency and impact on cuticle permeability. *Physiologia plantarum*, 151(4):446–458, 2014.
- [9] Mitchell JR MacLeod, Jeff Dedrick, Claire Ashton, Wilson WL Sung, Marc J Champigny, and Elizabeth A Weretilnyk. Exposure of two eutrema salsugineum (*thellungiella salsuginea*) accessions to water deficits reveals different coping strategies in response to drought. *Physiologia plantarum*, 2015.

- [10] S Andrews et al. Fastqc: A quality control tool for high throughput sequence data. *Reference Source*, 2010.
- [11] Brian Bushnell. Bbmap suite, 2014.
- [12] David M Goodstein, Shengqiang Shu, Russell Howson, Rochak Neupane, Richard D Hayes, Joni Fazo, Therese Mitros, William Dirks, Uffe Hellsten, Nicholas Putnam, et al. Phytozome: a comparative platform for green plant genomics. *Nucleic acids research*, 40(D1):D1178–D1186, 2012.
- [13] Alexander Dobin, Carrie A Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R Gineras. Star: ultrafast universal rna-seq aligner. *Bioinformatics*, 29(1):15–21, 2013.
- [14] Simon Anders, Paul Theodor Pyl, and Wolfgang Huber. Htseq—a python framework to work with high-throughput sequencing data. *Bioinformatics*, page btu638, 2014.
- [15] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.
- [16] Michael I Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome Biology*, 15:550, 2014.
- [17] Yushuang Guo, Changjun Huang, Yan Xie, Fengming Song, and Xueping Zhou. A tomato glutaredoxin gene slgrx1 regulates plant responses to oxidative, drought and salt stresses. *Planta*, 232(6):1499–1509, 2010.
- [18] Federico D Ariel, Anouck Diet, Martin Crespi, and Raquel L Chan. The lob-like transcription factor mtlbd1 controls medicago truncatula root architecture under salt stress. *Plant signaling & behavior*, 5(12):1666–1668, 2010.
- [19] Jill M Farrant, Norman W Pammenter, Patricia Berjak, Elizabeth J Farnsworth, and Christina W Vertucci. Presence of dehydrin-like proteins and levels of abscisic acid in recalcitrant (desiccation sensitive) seeds may be related to habitat. *Seed Science Research*, 6(04):175–182, 1996.

- [20] Jeffrey T Leek, Robert B Scharpf, Héctor Corrada Bravo, David Simcha, Benjamin Langmead, W Evan Johnson, Donald Geman, Keith Baggerly, and Rafael A Irizarry. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics*, 11(10):733–739, 2010.