

TRANSCRIPTIONAL RESPONSE TO DROUGHT
STRESS IN *Eutrema salsugenum*

YASSER SALAMA

McMaster University
Hamilton, Ontario, Canada, L8S 4E8

December 15, 2015

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

Abstract

TEST TEST TEST

Contents

1	Introduction	1
2	Methods	4
2.1	Plant Growth and Drought Simulation	4
2.2	Sequencing and RNA-seq Pipeline	4
2.3	Differential Gene Expression Analysis	6
3	Results	6
3.1	Sample Information and Read Statistics	6
3.2	k	8
3.3	Transcriptional Response to Drought	8
4	Discussion	8

List of Figures

1	Brassicaceae Phylogeny	2
2	Shandong and Yukon ecotype morphology	3
3	Read Processing Statistics	7

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 typically has little effect on mitigating the stressors' impact. 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.

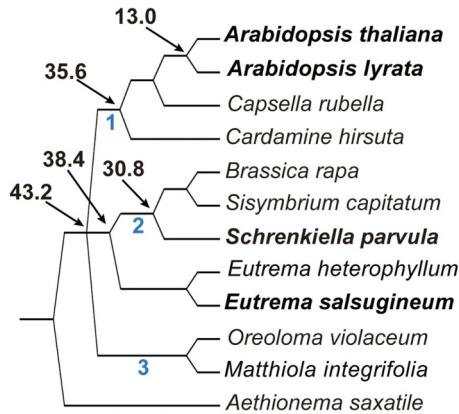


Figure 1 Brassicaceae phylogeny inferred with two markers, *ndhF* and *PHYA* [6].

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 [7]. Previous findings have shown that the Shandon ecotype seems to have a delayed response to drought and differential changes in cuticle permeability traits between the two ecotypes [8].

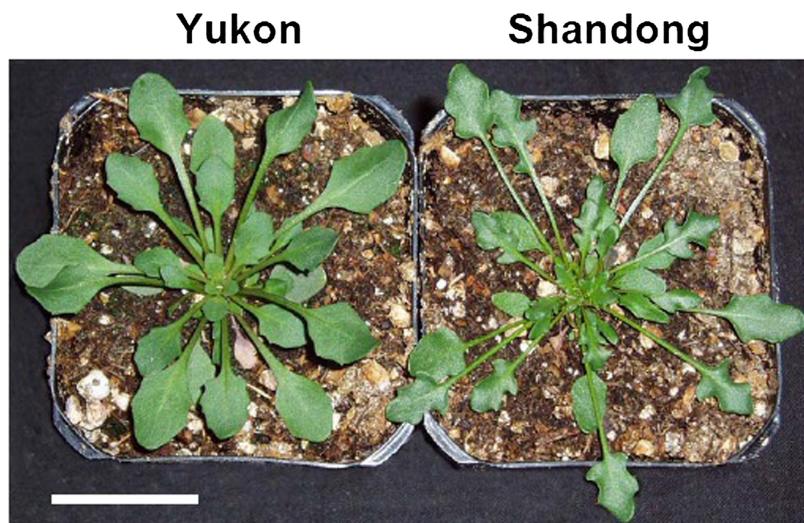


Figure 2 Shandon and Yukon accessions of *E. salsugenum*.

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. We also examine how the recovery of these plants manifests through a re-watering phase.

Establish challenges and the impact of these challenges. Explain that having a model that may provide a foundation of understanding for bypassing these challenges is obviously extremely useful and is needed. Introduce *E. salsugineum* as such a model. Describe it and then the ecotypes and then what is known already. Then describe about what is not known (how the ecotypes differ) and how understanding this can help such agricultural challenges. Describe how you did it, and then a brief overview of the findings.

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).

Sample	Ecotype	Treatment	Replicate	Lane	Batch
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
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
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
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

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.

FastQC (v0.11.3) was used to validate read metrics prior to and after trimming [9] (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 [10] (script: Step1_Concat_trim.sh). The reference genome and annotation file for *E. salsugenum* [6] were downloaded from Phytozyme (v10.3) [11]. 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 [12] (scripts: Step2_Index_ref.sh, Step3_Align.sh) (Figure 3). The alignment file was outputted as a bam file already sorted by coordi-

nated 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 [13] using both union and intersection-nonempty modes (scripts: Step4_Count_*).

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).

2.3 Differential Gene Expression Analysis

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.

FastQC analysis as described in Section 2 suggests that even prior to trimming, read quality was a "pass". 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 prepartion method, which introduces nucleotide biases at the beginning of the sequence. More importantly, the measures of sequence quailty and adapter contamination pass in all of the samples.

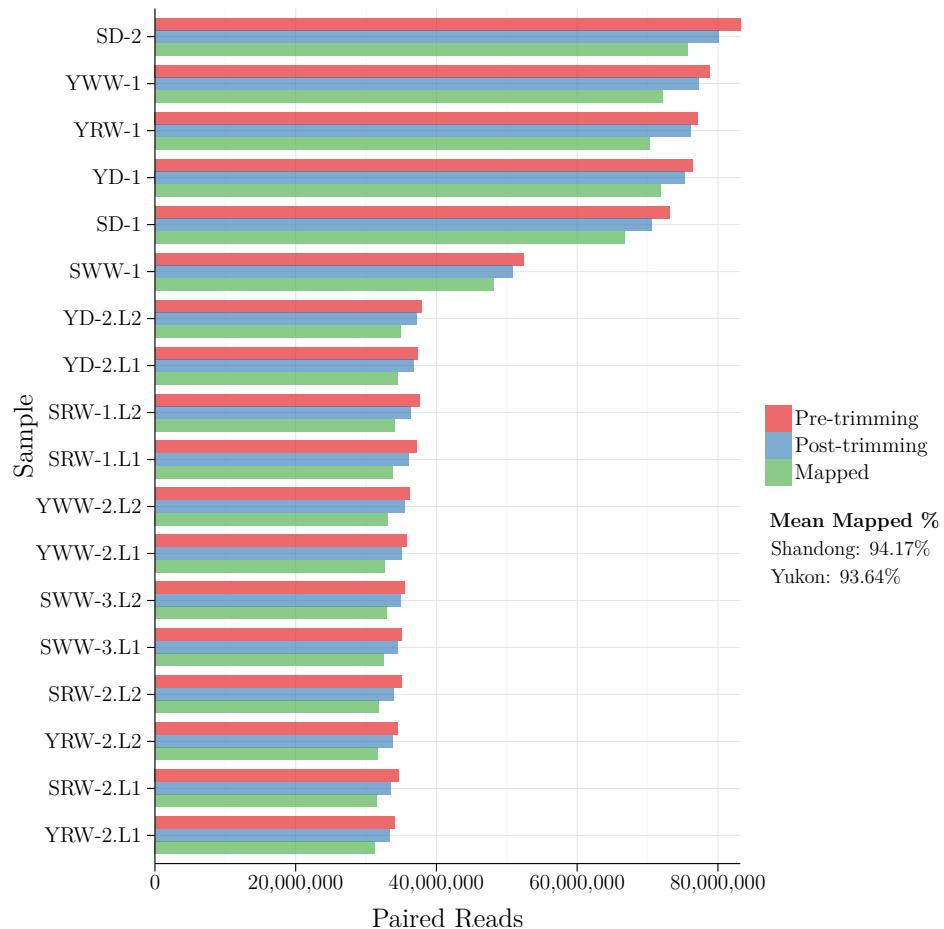


Figure 3 Read processing and mapping. Values are given as pairs of reads. Samples run on multiple lanes are left separated. The mean percentage of trimmed reads which were successfully mapped is indicated for each ecotype.

Quality trimming of the samples revealed 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 (Figure 3). 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 ($p = 0.09$).

3.2 k

3.3 Transcriptional Response to Drought

4 Discussion

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] 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.
- [7] 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.
- [8] 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.
- [9] S Andrews et al. Fastqc: A quality control tool for high throughput sequence data. *Reference Source*, 2010.
- [10] Brian Bushnell. Bbmap suite, 2014.

- [11] 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.
- [12] Alexander Dobin, Carrie A Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R Gingeras. Star: ultrafast universal rna-seq aligner. *Bioinformatics*, 29(1):15–21, 2013.
- [13] Simon Anders, Paul Theodor Pyl, and Wolfgang Huber. Htseq—a python framework to work with high-throughput sequencing data. *Bioinformatics*, page btu638, 2014.