

IA2_Rmarkdown_Output_Bi610

Dane Dewees

Sun Nov 26 18:48:27 2017

Statistical Analysis

Obtained data from Dr. Bill Cresko's lab were compared using a two-factor fixed effects ANOVA model. Analysis was focused on the effects of a three-level protein concentrations (ERD10, NCED9, and SRK2D, respectively) via mass spectrometry in root tissues of Arabidopsis plants that are influenced by *water* and *genotype*. Water treatment had two levels: *control* and *stress*. Genotype had two levels: *mutant* and *wildtype*. This was followed by Post hoc analysis (Tukey) of the sublevels for the three protein mean distribution. A P value <0.05 was considered to be statistically significant.

Genotype

The Null hypothesis assumes that there is no significant difference between WT and mutant genotype. The alternative hypothesis assumes that there is a significant difference between the two levels of the categorical variable (i.e., WT and mutant genotype):

$$H_0(\text{Genotype}) : \mu_{WT} = \mu_{Mutant}$$

$$H_A(\text{Genotype}) : \mu_{WT} \neq \mu_{Mutant}$$

Water

The Null hypothesis assumes that there is no significant difference between control and stress water treatment. The alternative hypothesis assumes that there is a significant difference between the two levels of the categorical variable (i.e., stress and control):

$$H_0(\text{Water}) : \mu_{control} = \mu_{stress}$$

$$H_A(\text{Water}) : \mu_{control} \neq \mu_{stress}$$

Genotype_Water_Interaction

The Null hypothesis for the interaction assumes that there is no interaction between the two levels of both water and genotype. The alternative hypothesis assumes that there is an interaction between said levels of both the established factors in study (Water and Genotype):

$$H_0(\text{Genotype} : \text{Water}) : \mu_{ij} = \mu_i + \mu_j - \mu$$

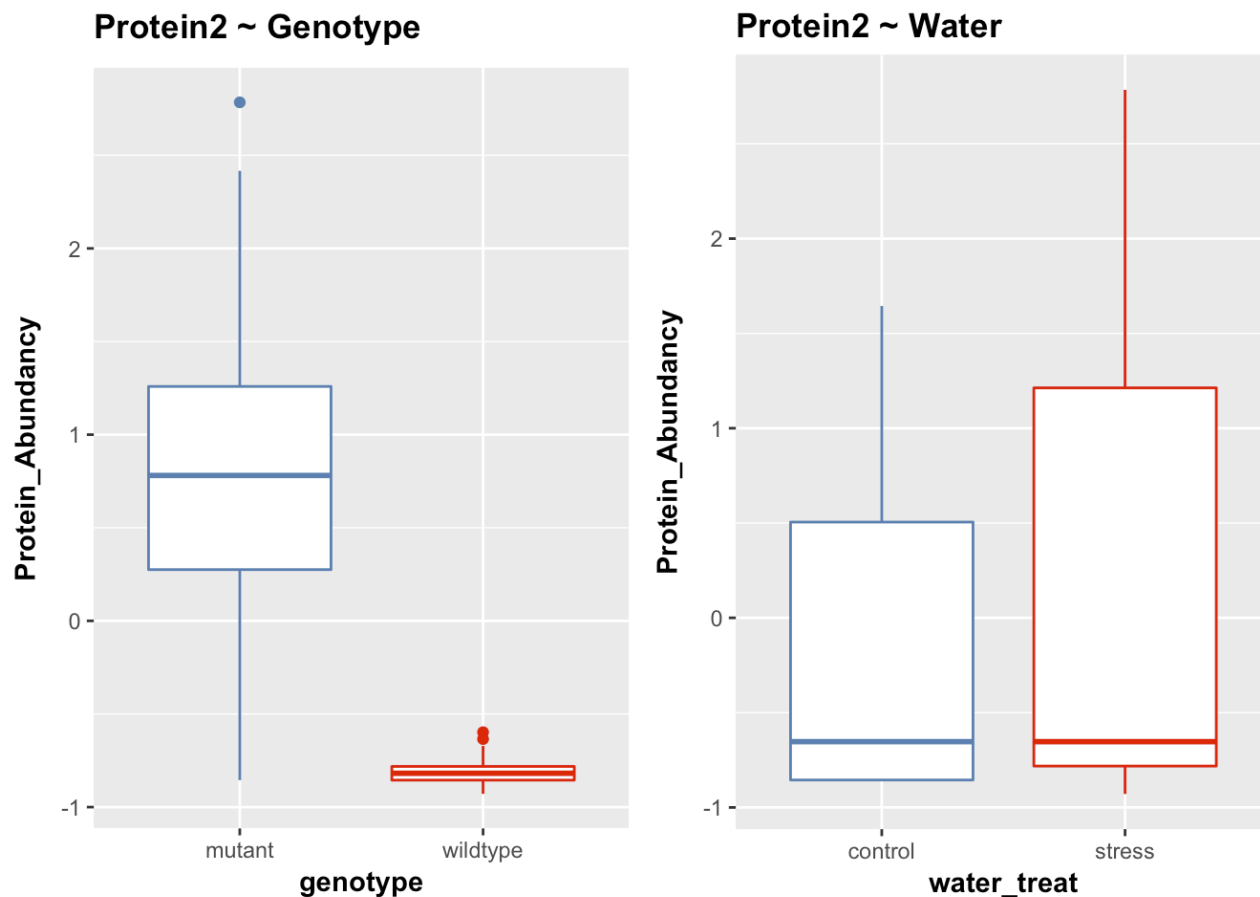
$$H_A(\text{Genotype} : \text{Water}) : \mu_{ij} \neq \mu_i + \mu_j - \mu$$

Results

Once the two factor and the levels associated within each factor was established, a two-factor fixed ANOVA model was used to measure the distribution across the protein abundance relative to the two

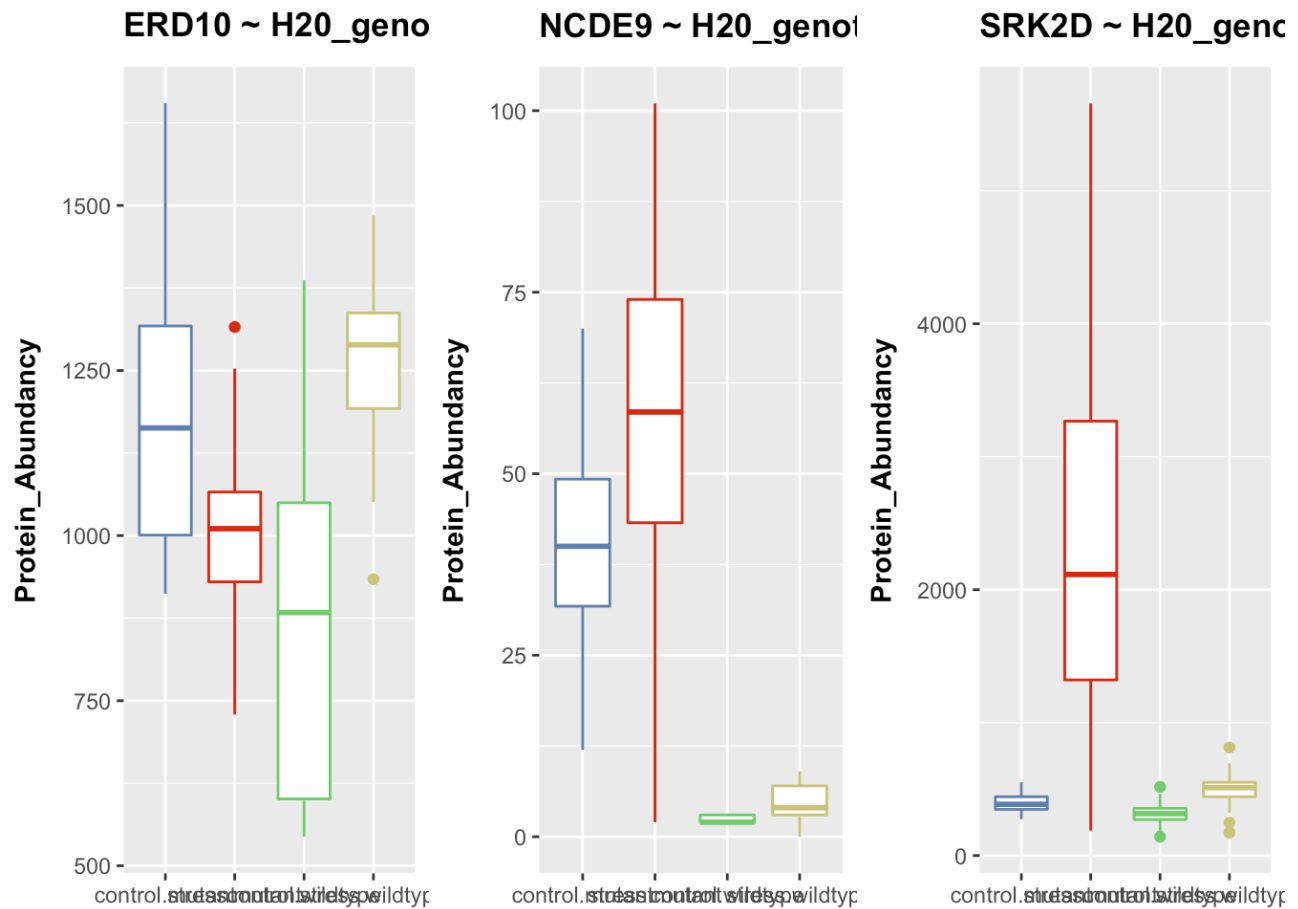
factors. Using exploratory analysis presented insight towards differences between the two factors when cross-referencing against the three proteins. Linear combinations was analyzed and determined the use of a two-way ANOVA model for further analysis. As shown in Figure 1, there was a spread in mean difference across genotype levels of *mutant* vs *wildtype* for protein NCED9. There was no difference (notable) between *water* and *genotype* for the other two proteins (Sup. Figure 1).

Figure 1



Interaction plots were used for each of the three proteins to demonstrate mean differences when looking at the interaction of Water treatment when applied to wildtype Genotype plant in ERD10 (Figure 2). Furthermore, when referencing the figure, one can see a distinction between the two levels of genotype, despite the presence of water treatments for the protein NCED9. The SRK2D gene expression frequency showed difference when compared with mutant plants & the stress water treatment from all other linear combinations of factors.

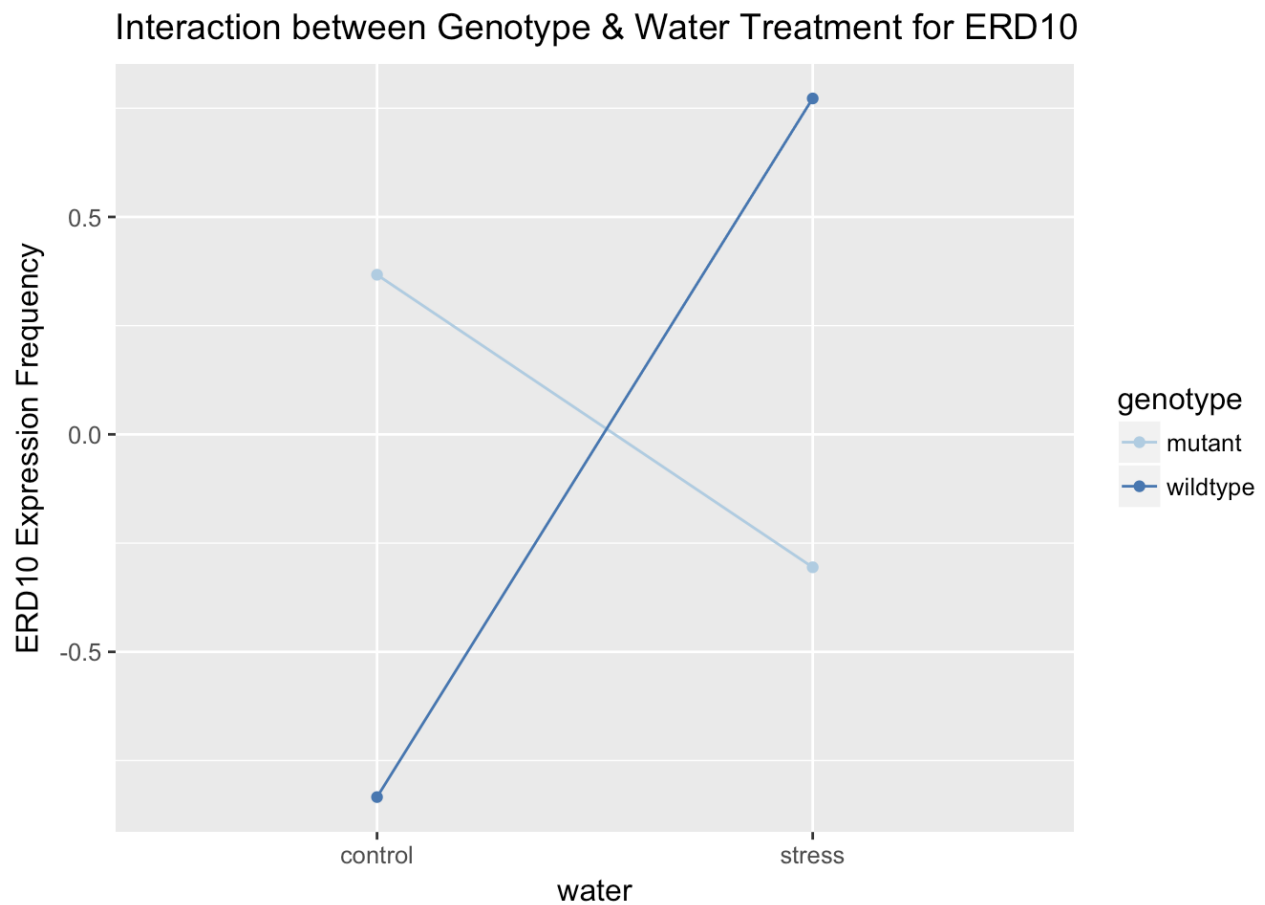
Figure 2



Following assumptions test and exploratory analysis, patterns indicated that normality assumption was not present for 95% confidence interval for both SRK2D & NCED9. No alterations were made to the dataset, but the output was something to note. Assumption for homoskedasticity of error variances also showed some violation for all three of the listed proteins. Determining the two categorical variables as being fixed enabled the use of factorial ANOVA model (fitted) which was demonstrated on the three protein concentrations. All levels of expression were normalized for comparisons sake.

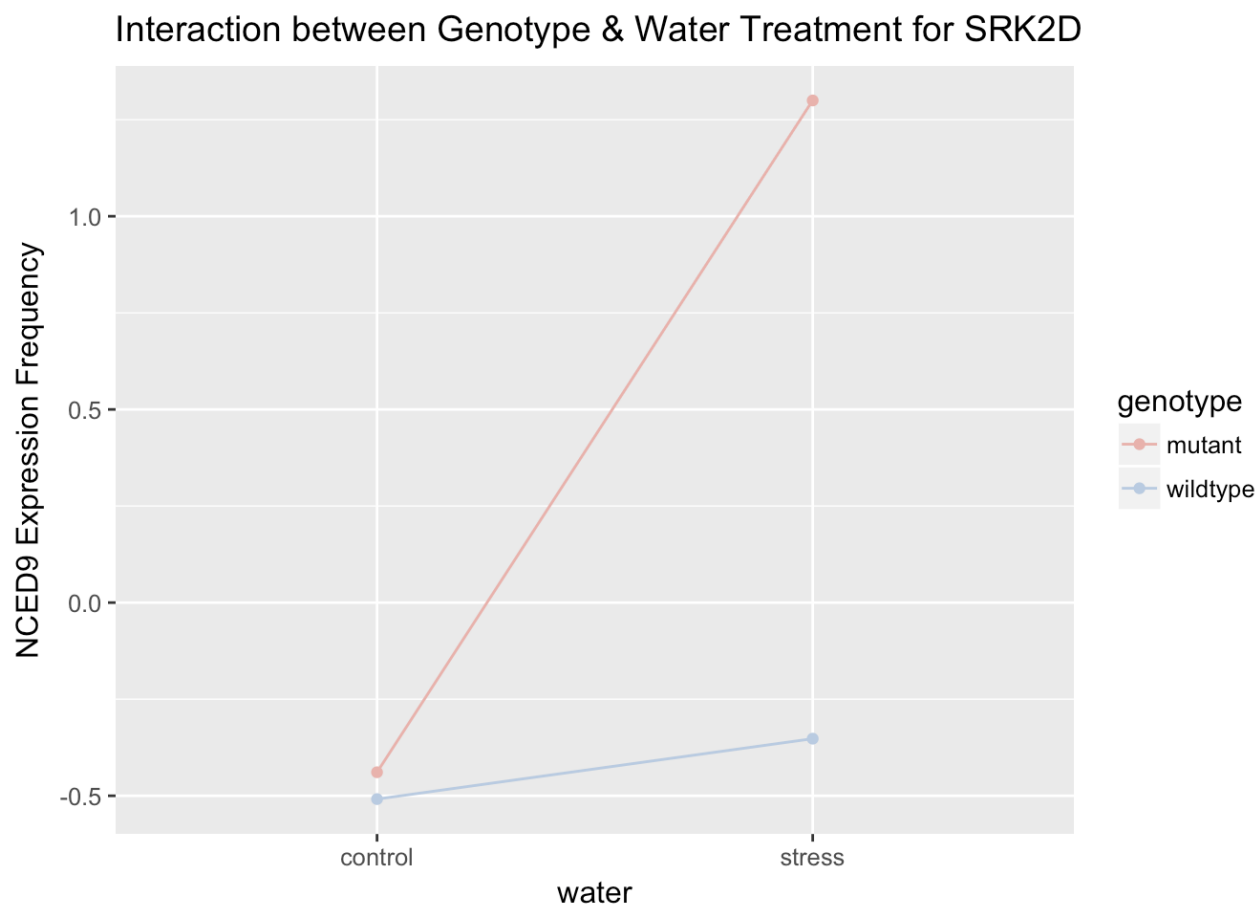
As seen in Figure 3 (Supp. Table 1), there was a significant interaction between genotype & water ($\alpha = 0.05$ comparison level, $F_{3,76} = 40.61$, $P = 1.299e-08$).

Figure 3



The use of a Tukey post-hoc test indicated that there was in fact mean differences between mutant and wildtype *genotypes* when cross-referenced against *water* treatment 'control' ($P = 0.00005$, Supp. Table 2a) when looking at the protein ERD10. Same was shown when looking at the mutant and wildtype *genotypes* when cross referenced with the *water* treatment 'strees' ($P = 0.0003$, Supp. Table 2a). When looking at NCED9, there was no statistically significant interaction between the two facotrs: *genotype* and *water*. The use of Tukey post-hoc was used to determine the mean differences between the *water* treatment levels (stress & control). The main effects model, which including factor variables Water and Genotype explain 67.5% of the variation in NCED9 normalized gene expression (Supp. Table 2b). Same was applied for SRK2D and there was a statistical interaction between *genotype* & *water* with regards to gene expression levels (Figure 4).

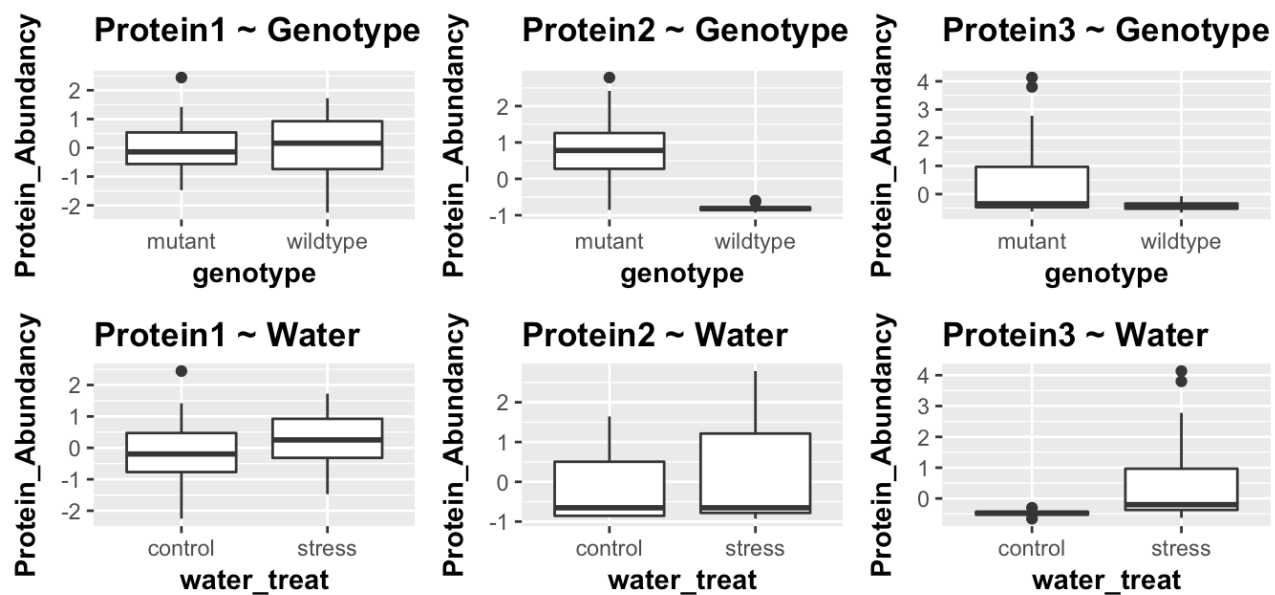
Figure 4



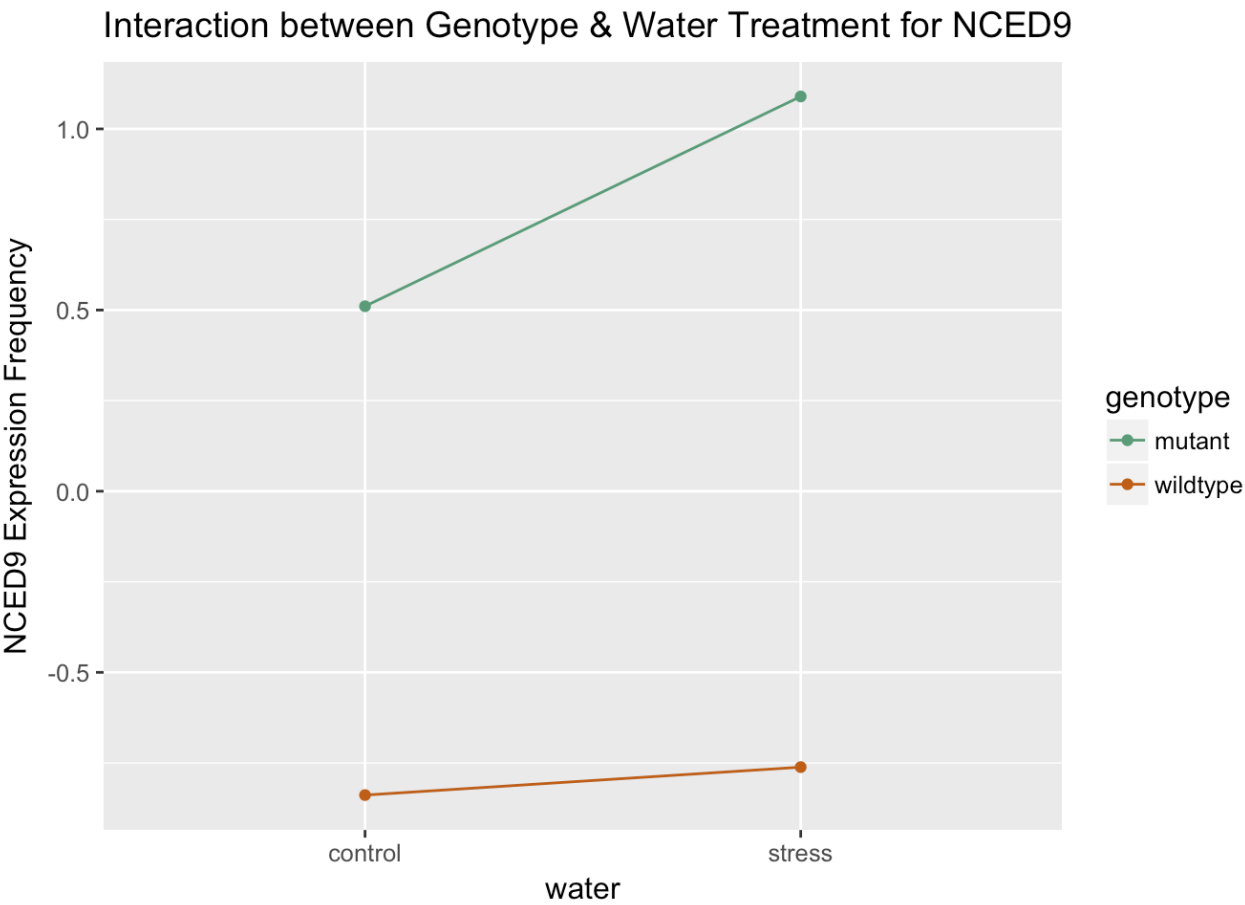
Variation seen in Figure 4 for SRK2D is explained with the interaction effects model with factor variables Water and Genotype. The overall mean-difference can be shown with regards to the linear trajectory in Figure 4 when referencing Plant differentiation. Variation in *genotype* plants showed to be statistically significant different with regards to the effect on normalized SRK2D gene expression values under the stress Water treatment ($P < 0.0001$, Supp. Table 3b).

Supplemental Material:

Supp. Figure 1



Supp. Figure 2



Supp. Table 1

## Analysis of Variance Table						
##						
## Response: plants_df\$scailed_zscore_ERD10						
##	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
## water	1	4.359	4.3595	6.8179	0.01087	*
## genotype	1	0.076	0.0757	0.1184	0.73170	
## water:genotype	1	25.969	25.9689	40.6133	1.299e-08	***
## Residuals	76	48.596	0.6394			
## ---						
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						

Supp. Table 1b

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = plants_df$scaled_zscore_ERD10 ~ water + genotype + water
:genotype)
##
## $water
##           diff           lwr           upr           p adj
## stress-control 0.4668774 0.1107579 0.822997 0.0108684
##
## $genotype
##           diff           lwr           upr           p adj
## wildtype-mutant -0.0615314 -0.4176509 0.2945882 0.731701
##
## $`water:genotype`
##           diff           lwr           upr
## stress:mutant-control:mutant -0.6726164 -1.3368477 -0.008385118
## control:wildtype-control:mutant -1.2010252 -1.8652565 -0.536793917
## stress:wildtype-control:mutant 0.4053460 -0.2588853 1.069577290
## control:wildtype-stress:mutant -0.5284088 -1.1926401 0.135822486
## stress:wildtype-stress:mutant 1.0779624 0.4137311 1.742193693
## stress:wildtype-control:wildtype 1.6063712 0.9421399 2.270602492
##           p adj
## stress:mutant-control:mutant 0.0460382
## control:wildtype-control:mutant 0.0000549
## stress:wildtype-control:mutant 0.3831038
## control:wildtype-stress:mutant 0.1657664
## stress:wildtype-stress:mutant 0.0003289
## stress:wildtype-control:wildtype 0.0000001
```

Supp. Table 2

```
## Analysis of Variance Table
##
## Response: plants_df$scaled_zscore_NCED9
##           Df Sum Sq Mean Sq F value Pr(>F)
## water      1  2.153    2.153   6.7115 0.01148 *
## genotype   1 51.206   51.206 159.6168 < 2e-16 ***
## water:genotype 1  1.259    1.259   3.9247 0.05120 .
## Residuals  76 24.381    0.321
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Supp. Table 2b


```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = plants_df$scaled_zscore_NCED9 ~ water + genotype + water
:genotype)
##
## $water
##          diff          lwr          upr      p adj
## stress-control 0.3281078 0.07586094 0.5803547 0.011479
##
## $genotype
##          diff          lwr          upr p adj
## wildtype-mutant -1.6001 -1.852347 -1.347853      0
##
## $`water:genotype`
##          diff          lwr          upr
## stress:mutant-control:mutant      0.57901377 0.108525 1.0495026
## control:wildtype-control:mutant -1.34919398 -1.819683 -0.8787052
## stress:wildtype-control:mutant -1.27199215 -1.742481 -0.8015034
## control:wildtype-stress:mutant -1.92820775 -2.398697 -1.4577190
## stress:wildtype-stress:mutant -1.85100592 -2.321495 -1.3805171
## stress:wildtype-control:wildtype 0.07720184 -0.393287 0.5476906
##          p adj
## stress:mutant-control:mutant      0.0096180
## control:wildtype-control:mutant 0.0000000
## stress:wildtype-control:mutant 0.0000000
## control:wildtype-stress:mutant 0.0000000
## stress:wildtype-stress:mutant 0.0000000
## stress:wildtype-control:wildtype 0.9729669
```

Supp. Table 3

```
## Analysis of Variance Table
##
## Response: plants_df$scaled_zscore_SRK2D
##          Df Sum Sq Mean Sq F value    Pr(>F)
## water      1 17.958 17.9584   40.493 1.353e-08 ***
## genotype   1 14.815 14.8153   33.406 1.560e-07 ***
## water:genotype 1 12.521 12.5209   28.232 1.039e-06 ***
## Residuals 76 33.705  0.4435
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Supp. Table 3b

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = plants_df$scaled_zscore_SRK2D ~ water + genotype + water
:genotype)
##
## $water
##              diff          lwr          upr p adj
## stress-control 0.9475857 0.6510028 1.244169      0
##
## $genotype
##              diff          lwr          upr p adj
## wildtype-mutant -0.8606775 -1.15726 -0.5640946 2e-07
##
## $`water:genotype`
##              diff          lwr          upr
## stress:mutant-control:mutant      1.73881490 1.1856308 2.2919989
## control:wildtype-control:mutant -0.06944836 -0.6226324 0.4837357
## stress:wildtype-control:mutant      0.08690817 -0.4662759 0.6400922
## control:wildtype-stress:mutant -1.80826325 -2.3614473 -1.2550792
## stress:wildtype-stress:mutant -1.65190673 -2.2050908 -1.0987227
## stress:wildtype-control:wildtype 0.15635653 -0.3968275 0.7095406
##              p adj
## stress:mutant-control:mutant      0.0000000
## control:wildtype-control:mutant 0.9875292
## stress:wildtype-control:mutant 0.9761324
## control:wildtype-stress:mutant 0.0000000
## stress:wildtype-stress:mutant 0.0000000
## stress:wildtype-control:wildtype 0.8795823
```

Supp. Table 3c

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = plants_df$scaled_zscore_NCED9 ~ water + genotype + water
:genotype)
##
## $`water:genotype`
##
```

	diff	lwr	upr
stress:mutant-control:mutant	0.57901377	0.108525	1.0495026
control:wildtype-control:mutant	-1.34919398	-1.819683	-0.8787052
stress:wildtype-control:mutant	-1.27199215	-1.742481	-0.8015034
control:wildtype-stress:mutant	-1.92820775	-2.398697	-1.4577190
stress:wildtype-stress:mutant	-1.85100592	-2.321495	-1.3805171
stress:wildtype-control:wildtype	0.07720184	-0.393287	0.5476906

```
##
## p adj
## stress:mutant-control:mutant 0.0096180
## control:wildtype-control:mutant 0.0000000
## stress:wildtype-control:mutant 0.0000000
## control:wildtype-stress:mutant 0.0000000
## stress:wildtype-stress:mutant 0.0000000
## stress:wildtype-control:wildtype 0.9729669
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = plants_df$scaled_zscore_NCED9 ~ water + genotype + water
:genotype)
##
## $water
##
```

	diff	lwr	upr	p adj
stress-control	0.3281078	0.07586094	0.5803547	0.011479

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = plants_df$scaled_zscore_NCED9 ~ water + genotype + water
:genotype)
##
## $genotype
##
```

	diff	lwr	upr	p adj
wildtype-mutant	-1.6001	-1.852347	-1.347853	0