From: **Kay Hodgins** kathryn.hodgins@monash.edu 📎
Subject: analysis
Date: May 10, 2017 at 3:18 AM
To: Katie Lotterhos k.lotterhos@gmail.com

Hey Katie,

I redid the analysis and as predicted I didn't find much after the fdr correction (just one molecular function was significant). However, I think there are still many interesting candidates. For example in the final cluster the only gene that was annotated was ACCLIMATION OF PHOTOSYNTHESIS TO ENVIRONMENT. In the second cluster there is a gene involved in sporopollenin biosynthesis which is important for desiccation resistant microspores in land plants. Several flowering time/floral development genes and drought/stress response genes are in the lists. I attached the lists with the TAIR homologs. Below are the top biological processes for each cluster. Only a few genes were annotated for each of the top terms. I guess you might expect more than one type of molecular pathway to be involved in adaptation to each climate combo.

Cluster 1 top terms

| | GO.ID | Term | Annotated | Significant | Expected | p | fdr |
|---|---|---|---|---|---|---|---|
| 17 | GO:0009653 | anatomical structure morphogenesis | 308 | 2 | 0.33 | 0.0395 | 0.2566727 |
| 23 | GO:0044262 | cellular carbohydrate metabolic process | 342 | 2 | 0.36 | 0.0479 | 0.2566727 |
| 25 | GO:0048856 | anatomical structure development | 842 | 3 | 0.89 | 0.0505 | 0.2566727 |
| 27 | GO:0001101 | response to acid chemical | 360 | 2 | 0.38 | 0.0526 | 0.2566727 |
| 28 | GO:0048513 | organ development | 370 | 2 | 0.39 | 0.0552 | 0.2566727 |

Cluster 2 top terms

| | GO.ID | Term | Annotated | Significant | Expected | p | fdr |
|---|---|---|---|---|---|---|---|
| 1 | GO:0006885 | regulation of pH | 18 | 2 | 0.07 | 0.0020 | 0.07880000 |
| 2 | GO:0006475 | internal protein amino acid acetylation | 21 | 2 | 0.08 | 0.0027 | 0.07880000 |
| 3 | GO:0016573 | histone acetylation | 21 | 2 | 0.08 | 0.0027 | 0.07880000 |
| 4 | GO:0018393 | internal peptidyl-lysine acetylation | 21 | 2 | 0.08 | 0.0027 | 0.07880000 |

Cluster 3 top terms

| | GO.ID | Term | Annotated | Significant | Expected | p | fdr |
|---|---|---|---|---|---|---|---|
| 13 | GO:0009628 | response to abiotic stimulus | 669 | 3 | 0.62 | 0.01852 | 0.21917 |
| 15 | GO:0009416 | response to light stimulus | 248 | 2 | 0.23 | 0.02027 | 0.21917 |
| 16 | GO:0009314 | response to radiation | 254 | 2 | 0.24 | 0.02121 | 0.21917 |

I also looked to see if your outlier clusters had an over-representation of differentially expressed genes using a fisher's exact test. I didn't find anything significant so far for pairwise comparisons of the climate treatments. However, I will also see if there are any co-expressed networks that were over-represented in your outlier clusters.

I just took a quick look at the LD for some of the outliers (Sam had the LD for the outliers calculated). As predicted, you can see for the few I looked at from cluster 1 the LD was pretty high - perhaps suggesting that selection is driving the associations ... if they are unlinked. Do you have this file? Sam has it on dropbox. I guess the only worry could be that they are actually nearby one another and the clusters are due to linkage so perhaps you wouldn't want to focus on this.

```
nput_data_and_intermediates kayhodgins$ grep comp90150_c0_seq1
pine_super_outlier_LD_p4_all_genemeans.txt | comp9553_c0_seq1
-bash: comp9553_c0_seq1: command not found
MU00063726:input_data_and_intermediates kayhodgins$ grep
comp90150_c0_seq1 pine_super_outlier_LD_p4_all_genemeans.txt | grep
comp9553_c0_seq1
comp90150_c0_seq1        comp9553_c0_seq1        0.182609947794215
```

```
comp9553_c0_seq1          comp90150_c0_seq1          0.182609947794215
MU00063726:input_data_and_intermediates kayhodgins$ grep
comp90150_c0_seq1 pine_super_outlier_LD_p4_all_genemeans.txt | grep
comp1565_c0_seq1
comp1565_c0_seq1          comp90150_c0_seq1          0.196499097314239
comp90150_c0_seq1          comp1565_c0_seq1          0.196499097314239
MU00063726:input_data_and_intermediates kayhodgins$ grep
comp90150_c0_seq1 pine_super_outlier_LD_p4_all_genemeans.txt | grep
comp24588_c0_seq1
comp24588_c0_seq1          comp90150_c0_seq1          0.291894701616989
comp90150_c0_seq1          comp24588_c0_seq1          0.291894701616989
```

Best,
Kay





tair_pine_cluster2







tair_pine_clusterall

limit.df$cluster==1, # aridity
limit.df$cluster==2, # geography
limit.df$cluster==3, # freezing
limit.df$cluster==4)) # multi