# Pollen Expression of HRGP Genes in Arabidopsis Thaliana

# Introduction

Hydroxyproline-rich glycoprotein or HRGP genes present in *Arabidopsis Thaliana*

What cis-regulatory sequences control the pollen-specific expression of HRGP genes in *Arabidopsis thaliana*?

# Methods

We started with a set of all (###) known HRGP genes in *Arabidopsis Thaliana*. We then split the set into categories; those that were only expressed in pollen tissue, and those that were expressed in non-pollen tissue. We then discarded any genes in both the pollen tissue and the non-pollen tissue as they would not contrast differences in the sets of genes. This left 145 genes in total, 13 pollen specific genes and 132 non-pollen specific.

Given the two sets of genes we collected the promoter regions from tair10. Using these promoter regions, we then ran a motif discover tool to generate (3502) potential motifs of length greater than 18 nucleotides.

A screenshot of a cell phone

Description generated with very high confidenceWe then used another tool to categorize if a motif was present in the promoter region of each gene. Using this information we created a plot of all motifs and the number of pollen specific genes vs all the non-pollen specific genes the motif was found in.

Each point on this scatter plot is one motif. The x-axis and y-axis represents the number of times the motif is found in the pollen specific and non-pollen specific genes respectively. We then selected the motifs present in 11(84.5%) or more pollen specific genes and present in at most 30(22.7%) non-pollen genes.

# Results

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| true positives | true negatives | Accuracy | Sensitivity | Specificity | motif name |
| 11 | 116 | 0.88 | 0.85 | 0.88 | gimme\_143\_MEME\_4\_w12 |
| 12 | 114 | 0.87 | 0.92 | 0.86 | gimme\_104\_Improbizer\_GTGCAACGGAGAGT |
| 12 | 111 | 0.85 | 0.92 | 0.84 | gimme\_146\_MEME\_7\_w12 |
| 11 | 105 | 0.80 | 0.85 | 0.80 | gimme\_16\_BioProspector\_w12\_1 |
| 11 | 115 | 0.87 | 0.85 | 0.87 | gimme\_13\_BioProspector\_w10\_3 |
| 11 | 106 | 0.81 | 0.85 | 0.80 | gimme\_92\_MDmodule\_Motif.12.3 |
| 12 | 117 | 0.89 | 0.92 | 0.89 | gimme\_132\_MEME\_3\_w10 |
| 13 | 108 | 0.83 | 1.00 | 0.82 | gimme\_105\_Improbizer\_AACACACGTTTATTAGATGTTT |
| 12 | 112 | 0.86 | 0.92 | 0.85 | gimme\_102\_Improbizer\_GCGTTATACCCGAGGATCAG |
| 11 | 107 | 0.81 | 0.85 | 0.81 | DME\_GADGAYKAS |
| 11 | 105 | 0.80 | 0.85 | 0.80 | DME\_GRHTGDTGA |
| 11 | 103 | 0.79 | 0.85 | 0.78 | DME\_MARKGDSRGA |

* True positive - number of pollen genes the motif is found in. (out of the 13 identified)
* True negative - number of non-pollen genes the motif is not found in. (out of the 145 identified)
* Accuracy – (True Positives + True Negatives) / Total number of genes
* Sensitivity – Percentage of pollen genes motif is found in
* Specificity – Percentage of non-pollen genes motif is found in
* Motif Name – The name assigned to the motif

# Discussion

These genes are all likely candidates for binding sites for transcription factors. Of the list there are (2 or 3)

\*\*short list research information

# Conclusions

# Acknowledgements

# References