

Gene-Ontology-based analysis of gene expression changes in early development in *Ceratopteris*

Ann E. Loraine
University of Alabama at Birmingham
aloraine@uab.edu

Mari L. Salmi, Stephen C. Stout, and Stanley
J. Roux
University of Texas at Austin
mari@mail.utexas.edu,
sroux@uts.cc.utexas.edu

Abstract

Following exposure to light, single-celled spores of the aquatic fern Ceratopteris richardii undergo a series of rapid developmental changes similar to germination in higher plants. To investigate this process, Salmi, et al. tracked gene expression changes over 48 hours post-light exposure using a microarray printed with partially-sequenced Ceratopteris clones [1]. Here, we extend this work by analyzing over-representation of Gene Ontology terms among differentially up-regulated genes from each time point.

1. Introduction

Early development of the single-celled, aquatic fern *Ceratopteris* spore serves as an experimental model system for studying germination in higher plants, a process by which previously dormant cells such as pollen or seeds undergo rapid developmental changes in response to an activating signal. For *Ceratopteris* spores, the signal is exposure to light, which triggers a series of well-characterized events, including: a polar calcium current at 6 hours, nuclear migration at 24 hours, a polar cell division at 48 hours, and rhizoid (root analog) formation at 72 hours in a position determined by the prior nuclear migration.

To investigate this process in germinating *Ceratopteris* spores, Salmi and coworkers tracked gene expression changes over 48 hours post-light exposure using a custom microarray printed with partially-sequenced clones from a *Ceratopteris* early-development cDNA library [1]. They interrogated several time points (0 h, 6 h, 12 h, 24 h, and 48h post-germination) and identified *Ceratopteris* genes that showed evidence for differential gene expression relative to at least one other time point. Using functional annotations associated with most closely-related *Arabidopsis* homologs (best hits in a blastx search), they proposed theories explaining the role of

individual genes in germination. Here, we extend this work by looking at the behavior of many genes at once, using Gene Ontology classifications transferred from *Arabidopsis*. Specifically, we attempt here to create a broad strokes picture of *Ceratopteris* development by analyzing over-representation of Gene Ontology terms among differentially up-regulated genes from each time point.

2. Methods

Over-representation analysis was performed using version 2.04 of the ErmineJ software [2], which uses the chi-square test of proportions to assess significance and compute p-values. Terms with p-values equal to or less than 0.0001 were counted as significantly over-represented. This software requires a chip annotations file that relates array identifiers (Genbank ids) to Gene Ontology codes. To create this file, we performed a provisional GO annotation of the *Ceratopteris* cDNAs using results from a prior blast analysis in which the *Ceratopteris* sequences were searched against an *Arabidopsis* protein sequence database [1,3]. GO terms associated with the putative *Arabidopsis* homologs were transferred to the *Ceratopteris* clones. GO annotations were obtained from the Gene Ontology Web site in March, 2005 and included annotations from TAIR and TIGR [4]. The GO annotations file is available at http://www.transvar.org/roux_collab.

3. Results

We find that at 0 hours, terms related to seed storage (GO:0045735, nutrient reservoir activity), metal transport (GO:0005506, iron binding protein), and disease resistance (GO:0006915, apoptosis) are over-represented among the up-regulated clones. The term "rhodopsin-like receptor activity" (GO:0001584) is also over-represented. The clones annotated with this term appear to encode homologs of signal transduction

molecules that might act downstream of a rhodopsin-like GPCR. Taken together, these GO terms create a picture of an organism well-stocked with mRNAs encoding nutrient storage proteins, transporters for the import of metal ions, and signal transduction proteins.

At 6 hours following the light signal, one term: GO:0005744, "mitochondrial inner membrane presequence translocase enzyme," was over-represented. At 12 hours, no categories were over-represented at p-values below 0.0001; the most significant categories included genes encoding putative peroxidases (GO:0016209, antioxidant activity, p-value 0.000321), proteases (GO:0006510, ATP-dependent proteolysis, p-value 0.000957), and enzymes involved in starch and sugar mobilization (GO:0005975, carbohydrate metabolism, p-value 0.00099.)

At 24 hours, 14 GO categories were over-represented. These categories included genes involved in lipid-related functions, proteolysis, and cellular differentiation. Taken together, these terms suggest that by 24 hours after germination, the spore is engaged in the breakdown and recycling of proteins as well as lipid biosynthesis.

At 48 hours, 17 categories were over-represented. These included GO terms related to protein synthesis, response to hormones and external signals, and glyoxylate metabolism. The next most significant terms (p-value < 0.001) related to cell biosynthesis and carbohydrate metabolism.

4. Conclusions and Discussion

Using Gene Ontology annotations transferred from *Arabidopsis*, we identified functional categories that are over-represented among the differentially-expressed genes at each time point and therefore are most likely to be relevant to underlying biological processes. The relevant terms include: terms for seed storage proteins, signal transduction, and metal-binding at 0 hours; terms for mitochondrial functions at 6 and 12 hours; fatty acid biosynthesis, lipid transport, and protein degradation at 24 hours; and terms for protein synthesis, glyoxylate metabolism, fatty acid oxidation, cell wall biosynthesis, and response to hormones at 48 hours. Taken together, these results evoke a developmental time course in which the newly-activated gametophyte establishes signal transduction pathways (0-6 hours), activates mitochondrial functions (6 and 12 hours), breaks down proteins and synthesizes fatty acids (24 hours), and then undergo a burst of protein and cell wall biosynthesis (48 hours) that apparently coincides with setting up a new

sensitivity to hormone-related signal transduction pathways.

Table 1. GO categories over-represented at 24 hours following germination

Lipid-related
GO:0004315, 3-oxoacyl-[acyl-carrier protein] synthase activity
GO:0000038, very-long-chain-fatty acid metabolism
GO:0004312, fatty-acid synthase activity
GO:000828, lipid binding
GO:0006869, lipid transport
GO:0044255, cellular lipid metabolism
GO:0006631, fatty acid metabolism
Proteolysis
GO:0004175, endopeptidase activity
GO:0004194, pepsin A activity
Misc
GO:0031225, anchored to membrane
GO:0044274, organismal biosynthesis
GO:0016747, transferase activity, transferring groups other than amino-acyl groups
GO:0030154, cell differentiation

Table 2. GO categories over-represented at 48 hours following germination

Protein synthesis
GO:0015935, small ribosomal subunit
GO:0005830, cytosolic ribosome (sensu Eukaryota)
GO:00162283, eukaryotic 48S initiation complex
GO:00162282, eukaryotic 43S preinitiation complex
Signal transduction
GO:0009733, response to auxin stimulus
GO:0009605, response to external stimulus
Misc
GO:0006097, glyoxylate cycle

References

- [1] M.L. Salmi, T.J. Bushart, S.C. Stout, and S.J. Roux, "Profile and analysis of gene expression changes during early development in germinating spores of *Ceratopteris richardii*, in press.
- [2] <http://microarray.genomecenter.columbia.edu/ermineJ/>
- [3] http://www.sbs.utexas.edu/roux/Ceratopteris%20Page/ceratopteris_research.htm
- [4] <http://www.geneontology.org>