Towards linking fungal genes to chemical spectra from soil organic matter using machine learning

Authors list
Martin-Bertelsen T(1), Nicolás C(2), Bentzer J(2), Johansson T(2), Smits M(3), Persson P(4), Tunlid A(2), Troein C(1)
Affiliations
(1) Computational Biology and Biological Physics, Department of Astronomy and Theoretical Physics, Lund University, SE-223 62 Lund, Sweden. (2) Microbial Ecology Group, Department of Biology, Lund University, SE-223 62 Lund, Sweden. (3) Centre for Environmental Sciences, Research Group Environmental Biology, Hasselt University, 3590 Diepenbeek, Belgium. (4) Centre for Environmental and Climate Research & Department of Biology, Lund University, SE-223 62 Lund, Sweden.

Abstract
Characterization of biological and chemical processes in soil organic matter (SOM) are increasingly being done by complementary high- throughput experimental techniques. Interpreting these diverse high-dimensional data types together poses challenges about data integration and analysis. A bioinformatics merge of machine learning and chemometrics may be a promising avenue to fully explore links between biological and chemical processes. Here, we propose computational methods for linking genes to decomposition mechanisms by integrating genome-wide transcription profiling data (RNA-Seq) with chemical changes in the organic compounds occuring during SOM decomposition as measured by chemical spectra (FT-IR, pyrolysis-GC/MS). The methods are intended to enable extraction of patterns that can be recognized and interpreted by domain experts. We also wanted to show proof of concept of integration of genome-wide gene expression data with chemical spectra from a study of two fungal species during SOM decomposition. This experiment was set up to measure effects of decreasing glucose levels over time on the decomposition of SOM. Two ectomycorrhizal fungi Paxillus involutus and Laccaria bicolor were grown in axenic cultures on SOM water extracts in a time series experiment. Species-specific co-expression networks were constructed from pairwise correlations between gene transcriptions across the four time points. Genes of similar functions between fungi were identified from orthology detection and spectral associations akin to chemometric 2D synchronous correlation analysis were computed and correlated with gene transcriptions to create gene-spectral association networks. Cross-species clustering was performed for discovery of conserved as well as species-specific modules of co-expressed genes and including correlated spectral patterns thus enabling characterization of the decreasing glucose responses in terms of modules of genes and associated spectral ranges. [Anticipated:] Modules of genes with similar expression profiles across time points were found and linked to certain spectral ranges of the FT-IR chemical spectra as well as certain compounds found with pyrolysis. These results demonstrate a[n anticipated] powerful application of machine learning clustering tools for integration of transcriptomics and chemical spectroscopy to leverage interpretation of data from controlled fungal SOM experiments.