Towards understanding of the ryegrass-endophyte symbiosis through the integration of transcriptome and metabolome data

Mingshu Cao, Linda Johnson, Albert Koulman, Geoff Lane and Susanne Rasmussen
AgResearch Grasslands, Private Bag 11008, Palmerston North, New Zealand
Email: *Mingshu.Cao@agresearch.co.nz*

Introduction

Fungal endophytes (*Neotyphodium lolli*) in perennial ryegrass (*Lolium perenne*) produce a range of bioactive alkaloids which are implicated in toxicity to grazing animals but also in resistance to insects. Better understanding of the regulatory and biochemical mechanisms of the symbiosis will provide clues for the genetic manipulation of beneficial alkaloid production. High throughput technologies in functional genomics can provide comprehensive information on a biological system. However, the integration of data from heterogeneous sources poses challenges for the effective formation of hypotheses. We report here data integration methods to gain insights into the complex biological system of the symbiosis of ryegrass and its fungal endophyte.

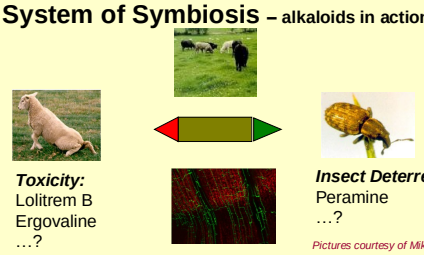
Objectives

Develop methods to manipulate data from high-throughput technology, such as microarray and tandem mass spectrometry.

Evaluate machine learning algorithms for effective feature selection and meaningful biological interpretation.

Long-term goal: identify regulatory and biochemical networks responsible for maintaining the symbiosis.

The System of Symbiosis – alkaloids in action



Toxicity:
Lolitrem B
Ergovaline
…?

Insect Deterrent:
Peramine
…?

Pictures courtesy of Mike Christensen

Data Generation and Preparation

Samples

24 ryegrass samples were examined in this study comprising three tissue types (Immature, Blade and Mature or Sheath) and four replicates of both endophyte-infected (E+) and endophyte-free (E-) isogenic ryegrass lines.

Microarray and direct infusion

About 15,000 ESTs were generated from 6 suppressive substractive hybridization (SSH) libraries and other sources. cDNAs were spotted in two-colour microarrays. Balanced incomplete block design was used for the original microarray experiments to make direct treatment comparisons robust. For the purpose of data integration, the original data were converted to log2 ratio values for each treatment.

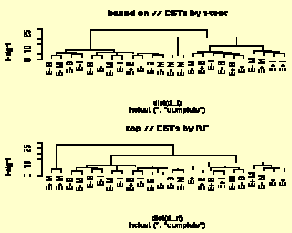
Direct infusion MS/MS

Direct infusion electrospray ionisation ion trap mass spectrometry (DIMS/MS) was used for metabolite profiling (m/z: 150-1000). The MS1 data were rounded to nominal mass, aligned and averaged. Additional MS2 and MS3 spectra were collected on the top 250 ions. Targeted analysis of the samples was conducted by GCMS, LC-PDA and GC-fluorescence.

Data normalisation

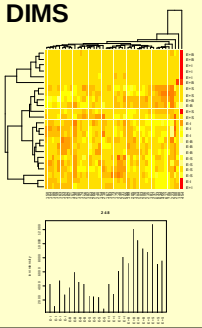
For comparison with expression values of ESTs, the concentration of each estimated metabolite was normalised against the average of the observations in all the samples, using log2 (**x**(i) / mean (**x**)), where vector **x** is the concentration measurements of each metabolite, and **x**(i) is the concentration of each individual treatment with i: 1 ~ 24.

EST



-- 77 differentially expressed ESTs based on adjusted p-value < 0.075 (Benjamin and Hochberg's false discovery rate) were compared with the top ranked 77 ESTs selected by RandomForest (RF). Although the differentiation of E+ and E- samples by RF is not better than by t-test, RF retains interactions among ESTs without overfitting (9.7% error rate of sample classification here).

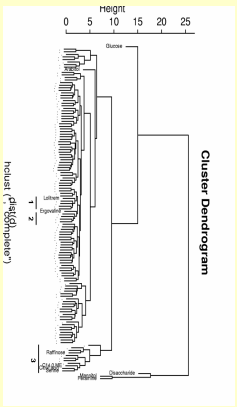
DIMS



-- E+ and E- samples can be differentiated by MS1 ions selected by RF (Fig. left above). Peramine (248) is present in the top list of ions. Its identity was confirmed by MS2. A plot of its intensity across samples is shown in Fig. left below.

-- Preliminary investigation of top ranked ions shows a number of highly correlated ions are C^{12} and C^{13} isotope pairs (203 and 204, 163 and 164, 633 and 634, for instance). 175 and 206 are chemically related (Ornithine derivative?).

Co-regulation of gene expression and metabolite accumulation – an example



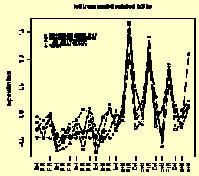


Fig. left: A clustering based on 119 selected ESTs and 11 target metabolites. Most selected ESTs are highly expressed in the E+ mature tissue. The expression values of lolitrem related ESTs (labelled as cluster 1 in the Fig.) are plotted in the Fig. above. Of these four ESTs, one is a gene for ascorbate peroxidase, one is DMAT (dimethylallyl tryptophan synthase), a gene likely to be involved in the lolitrem biosynthesis. The other two have unknown functions from current databases.

Discussion

The symbiosis of ryegrass and endophyte is a complex system which involves mutual communication and orchestration. The accumulation of different alkaloids is tuned by different genotypes of both fungi and the host. There is some knowledge about the biosynthesis of secondary metabolites in fungi but little about plant responses at the molecular level. Our analysis suggests that the general plant defence system is being triggered during the symbiosis. Many transcriptional factors and R related genes are up-regulated in the sheath. Increasing genomic information on ryegrass with effective data integration methods will improve our understanding of this complex system.

MS2 and MS3 from DIMS/MS provide rich information for structural inference of metabolites. Systematic approaches need to be developed to investigate and exploit the pattern of fragmentation of ions.

Acknowledgements

We thank Zaneta Park-Ng for pre-processing the microarray data; David Baird for advice on analysis of the gene expression experimental design; Alan McCulloch for EST assembly and blasting multiple gene databases; Karl Fraser and Brian Tapper for targeted analysis of metabolites data. References are omitted due to the limit of space.