

## Resources for gene expression analysis in *Medicago truncatula*, and their use for profiling symbiotic development

K. A. VandenBosch<sup>1</sup>, G. Endre<sup>1</sup>, N. Sharopova<sup>1</sup>, M. Graham<sup>1</sup>, K. Silverstein<sup>1</sup>,  
and C.D. Town<sup>2</sup>.

<sup>1</sup>Department of Plant Biology, University of Minnesota, St. Paul, MN

<sup>2</sup>The Institute for Genomic Research, Rockville, MD

A goal of the NSF-sponsored *Medicago truncatula* genomics program is to create resources for investigating legume gene function through analysis of expression patterns. We have generated >60,000 expressed sequence tags (ESTs) from 17 libraries from vegetative and reproductive organs. The project emphasizes interactions of *M. truncatula* with microorganisms, and 2/3 of the ESTs are from tissues responding to symbionts, pathogens, or elicitors. At TIGR, publicly available ESTs from *M. truncatula*, totaling about 165,000, have been grouped into contigs to produce tentative consensus sequences (TCs). The current *Medicago* gene index (MtGI 5.0) predicts >33,000 unique sequences, including about 16,000 TCs and >17,000 singletons (<http://www.tigr.org/tdb/tgi/mtgi/>). *In silico* analysis of gene expression can be used as a means to assess gene expression patterns, especially for genes that are highly expressed. We sorted TCs to identify predicted genes with particular expression patterns, based on the libraries of origin of the ESTs in the TCs. For example, about 17% of TCs are composed of ESTs strictly from libraries of tissues responding to microbes, and 60% of these are symbiosis-specific. Ongoing bioinformatic analysis of ESTs emphasizes identification of legume-specific sequences by comparison, using BLAST algorithms, of *M. truncatula*, *Lotus japonicus* and *Glycine max* EST contigs with those of other angiosperms. Some of these novel sequences appear to be members of gene families, based on clustering analyses. Future work will highlight expression analysis and phylogenetic relationships of these genes.

Our second major aim is to use microarrays to profile gene expression during microbial interactions. Expression profiling assays to date have used a microarray of ~1,000 non-redundant cDNA clones, named the "kiloclone set", that contains tissue-specific markers, indicators of microbial responses, and genes of unknown function. The kiloclone set has been resequenced and annotated, and will soon be ready for public release. This set has been used in pilot microarrays to standardize hybridization conditions and data analysis. Examples will be presented from the analysis of early symbiotic responses, and genes that are co-regulated during microbial interactions will be highlighted. Assembly of a 6k set of clones is now underway as a step towards construction of a comprehensive unigene set. This set encompasses the kiloclone set, plus clones selected from EST contigs deemed to be reliably error-free. A discussion of the evaluation of contig quality will be presented. The unigene set is designed as a public resource for studying many aspects of *Medicago* biology. Options will be discussed for expression analysis via collaborative or independent projects.