M-08

Exploiting *Medicago* structural variation to discover novel genes for nodulation and symbiosis

*Nevin Young**, College of Biological Sciences, University of Minnesota, Minnesota, USA
Shaun Curtin, Cereal Disease Laboratory, USDA-ARS, Minnesota, USA
Joseph Guhlin, Plant Biological Science Graduate Program, University of Minnesota, Minnesota, USA
Diana Trujillo, Plant Biological Science Graduate Program, University of Minnesota, Minnesota, USA

Like soybean, *Medicago truncatula* has become a model for legume biology and genomic research. *M. truncatula* has multiple reference quality genomes, dozens of deeply sequenced and assembled accessions as well as hundreds of survey sequenced genotypes suitable for SNP discovery and GWAS. *M. truncatula* is especially well studied in terms of symbiosis, nodulation and mycorrhization, leading to many of the seminal discoveries in plant-microbe communication. Nodulation in the *Medicago / Ensifer (Sinorhizobium)* symbiosis differs from soybean in the fact that nodules are indeterminate and nodulation relies on a large gene family – the nodule-specific cysteine-rich peptides (NCRs) – found only in *Medicago* and its close taxonomic relatives. *De novo* assembly of 16 *Medicago* accessions enabled genome-scale exploration of structural variation (SV), especially the variant architecture of large gene families. This study demonstrated that 22% of the genome is involved in large structural changes and 42% of reference genomic sequence is missing from one or more accession. Different stress-related gene families also differ notably in the details of their SV architecture. *Medicago* re-sequencing also provided a framework for the discovery of novel gene products important in nodulation. SNP-based GWAS of nodulation in 255 accessions uncovered numerous gene candidates. Ten of these candidates were tested through carefully replicated, statistically robust validation experiments involving *Tnt1* transposon, RNAi-hairpin, or CRISPR knockout (knockdown) lines. Among the newly discovered nodulation genes was *Pho2* (which helps to control phosphate supply). A parallel GWAS based on read depth variants (essentially, copy number variants) revealed that NCRs found in a cluster on chromosome 4 play an important functional role in nodulation. Finally, detailed analysis of gene family expansions specific to different legume lineages led to the discovery of *Medicago*-specific PLAT proteins and *Phaseoleae*-specific nodulins. CRISPR-based knockouts of *Medicago* PLAT proteins confirmed their functional role in nodule development.