

Meeting report

Legume genomes and discoveries in symbiosis research

Michael K Udvardi

Address: Max Planck Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Golm, Germany.
E-mail: udvardi@mpimp-golm.mpg.de

Published: 21 August 2002

Genome **Biology** 2002, **3(9)**:reports4028.1–4028.3

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2002/3/9/reports/4028>

© BioMed Central Ltd (Print ISSN 1465-6906; Online ISSN 1465-6914)

A report on the First International Conference on Legume Genomics and Genetics: Translation to Crop Improvement, Minneapolis-St. Paul, USA, 2-6 June 2002.

The Leguminosae is one of the largest and most successful families of flowering plants, largely as a result of the ability of its members to enter into mutually beneficial symbioses with soil bacteria (rhizobia) and fungi (mycorrhiza) that provide the plant with nutrients, such as nitrogen and phosphorus, which are scarce in many soils. These attributes, together with the large, nutritious seeds that are produced by many legumes, have secured a major role for legumes in world agriculture. Legumes have been an integral part of agriculture for thousands of years and remain a key component of sustainable agricultural systems. This report describes the significant progress in symbiosis research resulting from legume genomics and functional genomics.

Legume genome sequencing

Despite their importance to agriculture, legume genomes have not been among the first to be sequenced; this honor was bestowed upon *Arabidopsis* and then rice. Two model legume species, *Lotus japonicus* (a relative of several forage legumes) and *Medicago truncatula* (a relative of alfalfa), each with a genome size of 450-500 megabases (Mb), are now the focus of major genome sequencing efforts in Japan and the USA, respectively. Genome sequencing of *L. japonicus* (accession MG-20) began in 2000 and, as described by Shusei Sato (Kazusa DNA Research Institute, Japan), is currently proceeding at the rate of between 3 and 4 million base-pairs (Mbp) per month using a genomic library contained in a transformation-competent artificial chromosome (TAC) vector. Seed clones for sequencing are being selected by PCR screening, using primers to expressed sequence tags (ESTs), cDNA, or DNA markers linked to mapped genes of interest. Approximately 12 Mb of genomic

sequence has already been deposited in GenBank, and further data, including mapping information, are available from the *Lotus japonicus* genome sequencing project [<http://www.kazusa.or.jp/lotus/>] and the Kazusa DNA Research Institute [<http://www.kazusa.or.jp/en/>]. A similar approach to sequencing the *M. truncatula* genome from bacterial artificial chromosome (BAC) clones was described by Bruce Roe (University of Oklahoma, Norman, USA). The *Medicago* project, which began this year, aims to sequence about 5 Mbp per month. Thus, a large fraction of the gene space or euchromatic regions of these two model legumes, estimated to constitute 20-50% of each genome, should be available to researchers by the end of 2003. Information about the *Medicago* project is available at the site for sequencing *Medicago truncatula* at the University of Oklahoma [http://www.genome.ou.edu/medicago_totals.html].

Targeted genome sequencing has already begun to pay significant dividends in *L. japonicus*. Two novel *Sym* genes were isolated by combining genetic and physical mapping with genome sequencing of TAC clones spanning the region of interest. *Sym* genes are essential for nodulation and/or symbiotic nitrogen fixation in legumes. Approximately 40 *sym* mutants have been identified to date in various legume species, including pea, soybean, *M. truncatula*, and *L. japonicus*. Transposon-tagging in *L. japonicus* led to the isolation of the first *Sym* gene (*LjNin* or *LjSym20*) three years ago by the group of Jens Stougaard (University of Aarhus, Denmark). Until recently, however, map-based cloning of genes in legumes has proved to be a severe bottleneck in gene isolation. It is therefore significant that not one, but two new *Sym* genes have recently been isolated using this map-base cloning approach in *L. japonicus*. The group of Martin Parniske (John Innes Centre, Norwich, UK) isolated the first of these genes, *LjSym2*; it is required for symbiosis involving both arbuscular mycorrhizal (AM) fungi and rhizobia in root nodules (RNs). The *LjSym2* gene encodes a receptor-like kinase (hence its name change to *LjSymRK*) that is implicated in an early signal transduction pathway

common to both types of symbiosis. The existence of this common signaling pathway indicates a possible evolutionary link between AM and RN symbioses. Cloning of the gene from *Lotus* facilitated the isolation of the orthologous gene, *PsSym19* from pea, which is also essential for nodule development. *NORK*, the ortholog in *M. sativa* of *LjSymRK* and *PsSym19*, was also isolated recently by György Kiss. This work was not presented at the meeting in Minneapolis but was recently published (Endre *et al.*, *Nature* 2002, **417**:962-966) back-to back with the work of the Parniske group.

Genome sequencing was also instrumental in the identification, by Stougaard's group, of candidate genes within a region of 100 kb encompassing the *sym16* or *har1* (hypernodulating and aberrant root formation) mutation in *Lotus*. Sequencing of several independent mutant alleles showed that *Sym16* encodes a receptor kinase. In a tribute to perseverance, Peter Gresshoff (representing his group and that of Bernie Carroll at the University of Queensland, Brisbane, Australia) was able finally to report the isolation of the soybean *Nts* (nitrate tolerant symbiosis) gene, which is also involved in negative-feedback regulation of nodulation. As fate would have it, the soybean gene appears to be orthologous to *Sym16/Har1* in *Lotus*. An important difference between the work in *Lotus* and soybean was the time it took to clone each gene: 4 years in the case of *Lotus* and 17 in the case of soybean. This reflects the difference in size and complexity of the two genomes, and clearly vindicates the choice of *L. japonicus* and *M. truncatula* for sequencing as models for other legumes. In fact, there was great optimism amongst scientists at the meeting that genome sequencing and genetic mapping of these two species will facilitate the isolation of many agriculturally important genes in the more recalcitrant crop and pasture legumes over the next few years.

Legume EST projects

Plant genome sequencing is a 'rich man's sport'. For a more modest outlay, ESTs can provide valuable information about the coding regions of tens of thousands of genes, indicate their patterns and levels of expression throughout the plant, and generate raw material for the production of cDNA arrays for transcriptome analysis. Three legume species in particular have been the focus of large EST projects over the last few years - *Glycine max* (soybean), *M. truncatula*, and *L. japonicus*. As of 31 May 2002, GenBank contained 255,291, 162,741, and 31,670 ESTs for these species, respectively, which fell into approximately 50,000, 30,000, and 10,000 different tentative consensus sequences (TCs). Thus, a large fraction of all the genes in these three species are represented in the current EST collections. Bearing in mind that the EST collections represent cDNA libraries derived from specific organs harvested at defined developmental stages, or plants subjected to defined biotic and abiotic stresses, it is possible to use bioinformatics to infer something about the relative expression of thousands of different genes under a

variety of conditions. Chris Town (The Institute of Genomic Research, Rockville, USA) described the TIGR Gene Index Database [<http://www.tigr.org/tdb/tgi/>], which has collected all publicly available EST data from soybean, *M. truncatula*, *L. japonicus*, and many non-legume species into a relational database; it clusters ESTs into TCs and provides information including the relative expression level of each transcript cluster in different tissues. Pascal Gamas (CNRS-Institute National de la Recherche Agronomique, Castanet-Tolosan, France) described a similar *M. truncatula*-specific database that his group has developed.

Transcriptome analysis

'Electronic northern', which predict RNA expression levels from EST data (rather than by direct hybridization) are a useful starting point for transcriptome analysis. They are limited, however, by the number and variety of cDNA libraries and the depth of EST sequencing, and are probably distorted by factors including non-random sampling of mRNA resulting from differential efficiencies of cDNA synthesis and preferential cloning of short cDNAs. Numerous groups are now using cDNA arrays to carry out transcriptome analysis in legumes. In the area of symbiosis research, Maria Harrison (The Noble Foundation, Ardmore, USA) used cDNA arrays and *in silico* analysis of EST collections to identify novel mycorrhizal-inducible *M. truncatula* genes, including a novel phosphate transporter that may be involved in phosphate uptake at the plant-fungal interface. I described a systems biology approach to understanding metabolism in nitrogen-fixing nodules of *L. japonicus*, which integrates transcriptome, proteome, and metabolome analyses. A striking similarity was noted between changes in transcript levels for genes involved in carbon metabolism, especially organic acid synthesis, during nodulation of *L. japonicus* and cluster root formation on phosphorus-depleted white lupin (Claudia Uhde-Stone, University of Minnesota, St. Paul, USA), which suggested that nodules may operate under severe phosphorus limitation. Thus, transcript profiles that integrate data from thousands of genes will likely be of value as 'fingerprints' not only of development, but also of the physiological state of plants.

Other groups presented preliminary results of cDNA array analyses aimed at a better understanding of the symbiosis between *M. truncatula* and the soil bacterium *Sinorhizobium meliloti* (Kathryn VandenBosch, University of Minnesota, St. Paul, USA; and Helge Küster, University of Bielefeld, Germany). Gary Stacey (University of Missouri, Columbia, USA) has begun to use cDNA arrays from soybean, *M. truncatula*, and the non-legume *Arabidopsis* in an attempt to identify unique components of legume signaling pathways that enable them to develop nitrogen-fixing root nodules.

Finally, as a short-term substitute for genome sequence, ESTs and especially the TCs that they form are a valuable

resource for proteomics. This was nicely illustrated by Ulrike Mathesius (The Australian National University, Canberra, Australia), who showed how hundreds of proteins from *M. truncatula* could be identified by comparing peptide-mass fingerprints of isolated proteins with protein sequences encoded by the TCs of this species. Although it has been postulated that rhizobia may manipulate their plant hosts by altering hormone balances, changes in the proteome of *M. truncatula* roots following addition of specific hormones showed little overlap with changes resulting from inoculation with rhizobia.

The potential pay-offs from legume genome and EST sequencing in the areas of genetics, and structural, functional and comparative genomics were abundantly clear from the meeting in Minneapolis. It will be interesting to see how much of this potential is realized by the time of the Second International Conference on Legume Genomics and Genetics, which will take place in France in 2004.