Meeting report
Ways to get from plant genomes to phenomes: via yeast
Pablo D Rabinowicz and Willem Rensink

Address: The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA.

Correspondence: Pablo D Rabinowicz. E-mail: pablo@tigr.org

A report on the Cold Spring Harbor Laboratory meeting 'Plant Genomes: From Sequence to Phenome', Cold Spring Harbor, USA, 9-12 December 2004.

Whole-genome sequencing, although not yet a routine laboratory technique, is certainly becoming more affordable, and increasing numbers of nearly complete eukaryote genomes are relentlessly being added to the list. Making sense of the resulting overwhelming amount of sequence may require an equal effort. Several high-throughput tools for automated identification of genes at the structural level are available, but functional annotation can only be tentatively inferred on the basis of sequence motifs or sequence similarity. 'Gold standard' structural and functional annotation still requires extensive human intervention to eliminate frequent errors. The next challenge is to investigate how a genome sequence determines the phenotype of the whole organism (sometimes referred to as the 'phenome'). The way in which each protein contributes to the phenotype depends on a variety of factors such as regulation of expression, interaction with other proteins or nucleic acids, response to small molecules, subcellular localization, and so on. Technologies for genome-wide analysis of gene expression such as microarray hybridization are now commonly used and genome-wide analyses of protein-protein or protein-DNA interactions (the 'interactome') are emerging. The complexity of higher eukaryotic genomes makes analysis difficult, however, particularly for interactomes. This was reflected in a meeting on the functional analysis of plant genomes held last December in Cold Spring Harbor, where most of the results presented on interactome analysis were in fact carried out on the less complex genome of yeast.

A paradigm for this approach is a yeast synthetic genetic array (SGA) analysis described by Charles Boone (University of Toronto, Canada). Out of the 6,000 yeast genes, 5,000 have been shown to be non-essential in a genome-wide single-gene-knockout project, but double mutants of these non-essential genes often have lethal phenotypes (synthetic lethal phenotypes). SGA analysis allows the identification of genetic interactions, because if a double mutant has a synthetic lethal phenotype the two corresponding wild-type genes often have a functional relationship. By testing 132 genes as double mutants with each of the other 5,000 non-essential genes, Boone and his colleagues determined that each gene has an average of 30 synthetic genetic interactions and that there may be 100,000 such interactions in the yeast genetic network. Furthermore, they observed that physical (protein-protein) interactions and genetic interactions do not overlap because redundant protein complexes are present. Thus, one mutant protein in each complex yields a lethal phenotype whereas two mutant proteins in the same complex will be buffered by the other complex, resulting in a viable phenotype. They also showed that, using cluster analysis of SGA results, the function of an unknown gene could be predicted on the basis of the genes with which it is connected in the SGA network. SGA analysis paints a much more complex picture of the yeast interactome than previously reported protein-protein interaction networks alone. As SGA and protein-protein interaction networks are not complicated enough for Boone, his team is now moving towards SGA analysis of essential genes using inducible gene constructs. Although plants have several times more genes than yeast, large collections of knock-out mutants and high-throughput protein expression resources give us hope that this type of study will soon be feasible in plants.

Progress is being made towards understanding complex plant systems through the development of new technologies. A new *in vivo* analysis method for computational modeling of shoot apical meristems (SAMs) has been developed by Elliot Meyerowitz (California Institute of Technology,

Pasadena, USA). He described how the method takes advantage of *in vivo* confocal laser-scanning microscopy of *Arabidopsis* meristems. First, all cells are visualized using yellow fluorescent protein (YFP) fused to a plasma membrane protein, which enables cell divisions to be followed over time. In a second step, fluorescent protein fusions to gene products that are localized to the three different meristematic zones (central zone, peripheral zone and rib meristem) enable the identity of each cell in the meristem to be determined. As an example of how this technology can be used to dissect meristematic functions, Meyerowitz described how inducible overexpression of WUSCHEL (normally expressed in the rib meristem and a repressor of CLAVATA3, which is expressed in the central zone) resulted in an expansion of the central zone, as revealed by the presence of a fluorescent version of the CLAVATA3 protein outside its normal boundaries. These results, combined with the results of the cell-division timing experiment, enabled his team to determine that WUSCHEL respecifies peripheral meristem cells as central zone cells rather than increasing cell division in the central zone. The challenge now is to automate the data-acquisition process for large-scale analyses. For this purpose, a fluorescent histone fusion protein is used as a nuclear marker and software is being developed to automatically locate and track the nuclei as cells move and divide. Data of this type should enable computational modeling to identify all the meristem cells and eventually follow their lineages as they become part of the three different meristematic zones.

Small RNAs are emerging as important regulatory molecules, and high-throughput discovery of small RNAs can provide a comprehensive view of their function. Pamela Green (University of Delaware, Newark, USA) described how she and her collaborators have developed a sequencing method to identify and quantify these RNA molecules by modifying the massively parallel signature sequencing (MPSS) technology. There are two kinds of small RNAs: microRNAs (miRNAs) and small interfering RNAs (siRNAs). The former are derived from hairpin-containing precursors and the latter originate from double-stranded RNAs through the action of an RNA-dependent RNA polymerase (RDR). Green reported the first truly genome-wide analysis of small RNAs, which showed that small RNAs are widespread in the *Arabidopsis* genome and that differential silencing occurs between different tissues. Furthermore, siRNAs can be distinguished from miRNAs by northern-blot hybridization of small RNAs against wild-type and RDR mutants.

As more and more genome sequences are completed, comparative analyses become more effective in gene discovery and even in determining gene function. In a compelling example of such a study, Susan Dutcher (Washington University, St. Louis, USA) reported the use of cross-kingdom genomic comparisons to identify a gene responsible for a rare human disease. A comparison between the proteomes of the alga *Chlamydomonas reinhardtii* and humans (both of which have flagellate cells, despite being otherwise very different) resulted in around 4,000 proteins shared by both species. When proteins present in *Arabidopsis* (a non-flagellate organism) were subtracted, 688 proteins remained. This set contained most flagellum-related proteins, including a human protein that shows similarity to a *Chlamydomonas* flagellar protein and is encoded in a region of the genome containing one of the genetic loci (*BBS5*) known to be responsible for Bardet-Biedl syndrome, a complex disease that is believed to be caused by defects in flagellar function. This correspondence enabled identification of the *BBS5* gene as encoding a flagellar protein; there is a correlation between mutations in *BBS5* and Bardet-Biedl syndrome and further analysis of the *BBS5* gene confirmed its function and involvement in the disease.

Intra-kingdom genome comparisons are also useful in the understanding of plant biology and evolution, but the plant species for which genome sequences are available span only 200 million years of land plant evolution. Fortunately, as announced by Jody Banks (Purdue University, West Lafayette, USA) in her presentation, the genome of the ancient seedless plant *Selaginella moellendorffii* will soon be sequenced, adding another 200 million years of evolutionary history to comparative plant genomics. It is estimated that its compact, gene-rich genome is less than 100 megabase-pairs (Mb) long and contains homologs of most known and putative plant genes as well as genes not present in angiosperms.

For small genomes, a high level of refinement can be achieved by intra-species comparisons. Mark Johnston (Washington University School of Medicine, St. Louis, USA) reported on the identification of functional features in the non-coding sequence of yeast. By sequencing six closely related and divergent yeast strains, putative target sequences of transcriptional regulators could be identified. Intra-specific comparative analysis of the *Arabidopsis* genome is also becoming a reality. A high-resolution genotyping study of 20 *Arabidopsis* accessions using oligonucleotide microarrays was presented by Justin Borevitz (University of Chicago, USA). Among the polymorphisms investigated, disease resistance-like genes and genes for receptor-like proteins, for example, show higher levels of variation than genes for basic helix-loop-helix DNA-binding proteins.

The new tools, technologies and genomes available for plant biology will sooner or later allow plant phenome research to catch up with the rapidly growing yeast field. We hope to be listening to many more exciting plant proteomics, phenomics, and interactomics talks at the next Cold Spring Harbor Plant Genome meeting in 2007.