The Plant Cell, Vol. 14, 2651, November 2002, www.plantcell.org © 2002 American Society of Plant Biologists

MEETING REPORT

Genomics and Forest Biology: *Populus* Emerges as the Perennial Favorite

Forest biologists have developed strong justifications for why trees should be viewed as model systems in plant biology, including the obvious challenges in extrapolating findings from annual, herbaceous plants to organisms that are distinguished by perennial growth, large size, complex crown architecture, extensive secondary xylem, dormancy, and juvenile–mature phase changes (Bradshaw et al., 2000; Taylor, 2002). Similar justification has been used to argue why the genome of a tree should be sequenced. The U.S. Department of Energy (DOE), Office of Science, announced earlier this year plans to sequence the first tree genome, that of the black cottonwood (*Populus trichocarpa*) (Figure 1).

In what has emerged as an exciting area in forest biology, two meetings recently convened at which advances in *Populus* genomics were featured: the International Poplar Symposium III, held August 26 to 29, 2002, in Uppsala, Sweden, and the *Populus* Functional Genomics Workshop, held August 30, 2002, in Umeå, Sweden. Together, the meetings attracted 200 scientists from 23 countries. Here, we summarize those portions of the symposium and workshop that highlight emerging tools, techniques, and research directions in *Populus* genomics.

SEQUENCING THE POPLAR GENOME

Although efforts to identify *Populus* as a model tree began long before sequencing a tree genome was a possibility, the choice of poplar was ideal in that the genome size is small, ~550 Mbp. This is similar in size to the rice genome, only 4 times larger than the genome of Arabidopsis, yet 40 to 50 times smaller than the genome of pine.

Not surprisingly, a presentation by Gerald Tuskan from Oak Ridge National Laboratory (ORNL; Oak Ridge, TN) drew a large audience interested in learning more about efforts to sequence the *Populus* genome. According to preliminary plans, the DOE Joint Genome Institute, with funds from the DOE Office of Biological and Environmental Research, will provide a 3× draft sequence of the female black cottonwood clone Nisqually-1 in late 2002 and a second 3× draft in late 2003. The first 3× draft will be generated by a random shotgun approach. The second 3× draft will be based on a minimum tiling path that has been established from 90,000 BAC end sequences provided by Genome Canada (Vancouver, British Columbia). Together, these approaches will provide ~6× coverage of the genome. Alignment of assembled

Figure 1. *Populus*: A Model System for Tree Genomics.

At left, 7-year-old hybrid poplars being harvested in western Oregon. Top right, expression of a poplar DEFICIENS homolog in female floral meristems of black cottonwood (Sheppard et al., 2000); bottom right, germinating pollen grains being tested for viability using a fluorescent stain.

MEETING REPORT

contigs will be based on information from physical and genetic maps. Annotation will take place using *Populus*-specific gene-finding models trained from full-length cDNA sequences provided by the Umeå Plant Science Center (UPSC; Umeå, Sweden), Genome Canada, and ORNL.

EST SEQUENCING

ESTs represent an informative tool for gene discovery. Stefan Jansson from UPSC reported on an extensive EST library being developed as part of the Swedish *Populus* Genome project, a joint collaboration between UPSC and the Genome Center at the Royal Institute of Technology in Stockholm. In the initial phase of this project, almost 5700 ESTs were developed for wood-forming tissues (Sterky et al., 1998). Today, this resource has grown to >95,000 ESTs sequenced from 20 different cDNA libraries and from a range of tissues and developmental stages. Analyses indicate that these ESTs derive from perhaps 15,000 to 20,000 genes, a significant fraction of the 40,000 to 50,000 genes believed to be coded by the *Populus* genome. Basic Local Alignment Search Tool (BLAST) searches against sequenced ESTs are possible through the project database (PopulusDB) or data deposited in GenBank.

Although a functional classification of all 95,000 ESTs has not been completed, several subsets of the data have been analyzed. Figure 2 shows the distribution of genes in various functional categories for young poplar leaves and leaves harvested before visible signs of senescence. According to Jansson, young leaves devote one-third (36%) of their transcript pool to "energy," whereas older leaves have a high abundance of transcripts in categories such as "cell death and aging" and "protein destination," which includes functions related to proteolytic degradation. Furthermore, the fraction of sequences with little similarity to known genes (BLAST score of <100) was almost twice as high in autumn leaves as it was in young leaves.

Most of the ESTs sequenced to date have been for hybrid aspen (*P. tremula* × *P. tremuloides*), European aspen (*P. tremula*), and *P. trichocarpa*. Chung-Jui Tsai from Michigan Technological University (Houghton) reported that >12,000 ESTs also have been sequenced from various tissues of quaking aspen (*P. tremuloides*). Based on initial comparisons, gene sequences among the different species

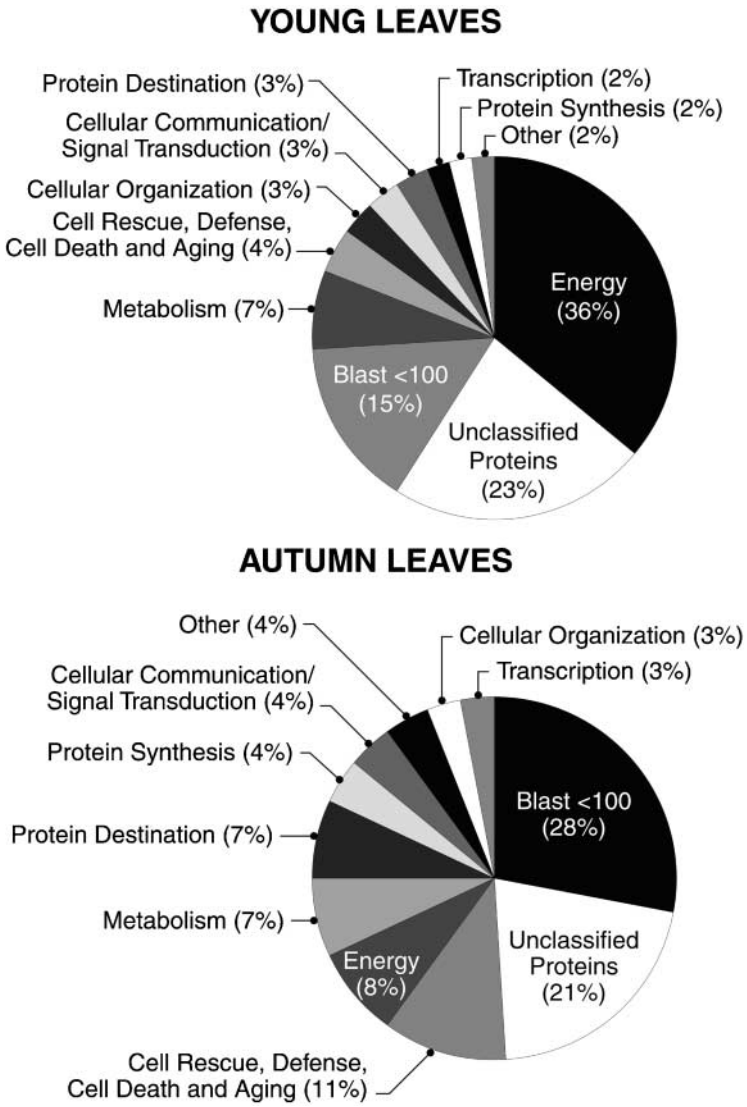

Figure 2. Functional Distribution of Genes According to a Modified MIPS (Munich Information Center for Protein Sequences) Classification Scheme of 4842 ESTs from Young *Populus* Leaves and 5128 ESTs from Leaves Collected in Autumn.

Unclassified proteins show similarity to a gene of unknown function, typically an Arabidopsis open reading frame. (*Data courtesy of S.J.*)

show high (>95%) similarities. Sequence similarities also exist between poplar ESTs and ESTs from other tree genera, including *Salix* (i.e., willow). Rishikesh Bhalerao from UPSC described a *Salix* EST sequencing program in which a cDNA library has been prepared from stem tissues. To date, ~2000 *Salix* ESTs have been sequenced and compared with ESTs from Arabidopsis and poplar. Analyses indicate a moderate degree of similarity with Arabidopsis but a high degree of sequence conservation between *Populus* and *Salix*. Such conservation suggests that resources being developed for *Populus* (e.g., microarrays) also can be used in related genera.

GENE EXPRESSION MICROARRAYS

Transcript profiling can reveal patterns of gene expression during developmentally regulated events such as wood formation and leaf expansion and maturation. Göran Sandberg from UPSC described how early poplar microarrays were used to study developmental processes involved in wood formation (Hertzberg et al., 2001). In this case, microarrays helped characterize how gene expression varied throughout cell division, expansion, secondary wall formation, lignification, and cell death. These studies used a microarray spotted with 2995 ESTs unique to hybrid aspen.

Preparation of a higher density *Populus* microarray based on the 95,000-EST library described above has been initiated. As reported by Peter Nilsson from the Royal Institute of Technology in Stockholm, the Swedish *Populus* Genome project has produced a spotted EST microarray containing >13,000 clones. This array is based on the unigene set extracted after sequencing 35,000 clones. Swedish researchers are using these microarrays to study many processes in *Populus*, and access to the arrays will be provided on a collaborative basis. One such collaboration was described by Gail Taylor (Southampton University, UK), who presented data on global gene expression after long-term exposure to increased CO_2 concentrations of poplar trees grown as part of the PopFACE (free air carbon dioxide enrichment) experiment in central Italy. Approximately 1500 array ESTs showed enhanced expression, whereas 1000 ESTs showed reduced expression, at increased CO_2 concentrations. Taylor emphasized that these studies eventually should provide much-needed insights into the long-term adaptation of trees to future climate conditions. The challenge, of course, will be to determine how changes in gene expression are related to altered biochemical and physiological function and, ultimately, to tree growth.

TRANSGENIC LINES AND FUNCTIONAL GENOMICS

The creation of transgenic lines with enhanced or reduced levels of gene expression is the most straightforward way to determine gene function. Rishikesh Bhalerao described an ambitious program in which a commercial partner, SweTree Genomics, will use RNA interference to knock out 2000 genes involved in wood formation. A high-throughput recombinational cloning system will allow the necessary constructs to be generated within 2 years and the production of regenerated transgenic plants within 48 months. The group has devised a tiered phenotyping program that, as reported by Björn Sundberg from UPSC, will identify transformed lines with altered physical and chemical properties, thus highlighting economically and biologically important processes involved in wood formation.

Richard Meilan (Oregon State University, Corvallis) described activation tagging as an approach to the creation of dominant mutations in trees. Activation tagging is a method whereby an enhancer element is inserted randomly into the genome, enhancing the expression of a nearby gene. Meilan and Oregon State University colleagues Amy Brunner, Victor Busov, and Steve Strauss have used this technique to produce a large population of transformed individuals. Nine of these plants showed a visibly altered phenotype. One mutant exhibiting stunted growth, thick stems, and dark leaves was identified as having an open reading frame with high nucleotide similarity to a gibberellic acid 2-oxidase gene. The Oregon State University group also is working with Andrew Groover (U.S. Forest Service, Davis, CA) to identify genes based on expression patterns using enhancer-trap and gene-trap insertion lines. In this project, T-DNAs carrying the β-glucuronidase reporter gene are introduced randomly into the genome, and insertion results in a pattern of β-glucuronidase expression that mimics the normal expression patterns of interrupted genes. These techniques are being used to identify genes with preferential expression in leaf vascular tissues, wood-forming tissues, roots, and adventitious root primordia.

In addition to T-DNA–based methods, an alternative transposable tagging system is being explored by Sandeep Kumar and Matthias Fladung from the Institute for Forest Genetics and Forest Tree Breeding (Grosshansdorf, Germany) for use in generating mosaics of insertional and activation mutants within a single plant. The system combines the dominant *rolC* marker gene from *Agrobacterium rhizogenes* and the *Ac* transposon from maize. Excision of the *Ac* element is indicated by pale green leaf sectors caused by *rolC* expression, thus visually marking cells in which the *Ac* element has excised native genes upon reinsertion. Plants then can be regenerated from the affected cells in tissue culture. A comparison of *Ac*/T-DNA insertional data indicated that rates of *Ac* insertion in coding regions were twofold greater than for T-DNA–based systems.

It is anticipated that the number of transgenic lines produced by the methods described above will eventually exceed the capacity for their maintenance and characterization. Chung-Jui Tsai and colleagues are investigating the feasibility of using a cryonics-based preservation system for the long-term storage of transgenic germplasm.

MEETING REPORT

METABOLIC CHARACTERIZATION OF PHENOTYPES

Because visual differences among transformed individuals are expected in <1% of a transgenic population, methods to characterize the metabolic phenotypes of trees possessing gain- or loss-of-function mutations must be developed. Consistent with the terminology of Fiehn (2002), metabolic profiling is being explored in *Populus* by Chung-Jui Tsai and Tim Tschaplinski from ORNL, and metabolomics is being explored by Thomas Moritz from UPSC. Each group is using similar sample preparation and derivatization protocols, followed by either quadrupole gas chromatography–mass spectrometry or fast gas chromatography–time-of-flight mass spectrometry. Few results were presented, although everyone agreed that the large amount of data generated from these analyses would dictate compromises between speed of analysis and number of metabolites detected and quantified. Multiple online fractionation techniques are being considered, and additional analyses involving liquid chromatography–mass spectrometry are being developed to detect metabolites that are difficult to analyze by gas chromatography–mass spectrometry.

Given that many of these techniques are just being developed, all speakers agreed that various analytical and computational challenges would be encountered in this area of investigation. Thomas Moritz illustrated the complexities that will have to be overcome in *Populus* metabolomics by discussing how fragmentation patterns for hundreds of compounds might be detected and analyzed statistically using principal-component analysis and partial least-squares projections to latent structures. Such analyses would eventually link data obtained with expression arrays to consequences in biochemical pathways, and thus to function. Tim Tschaplinski pointed out that there are many unique low molecular mass compounds in *Populus* and suggested that the establishment of standardized mass spectrometry libraries would aid greatly in the development of subsequent deconvolution strategies.

FORMATION OF AN INTERNATIONAL GENOME CONSORTIUM

One concern that was raised during the International Poplar Symposium III, and again at the *Populus* Functional Genomics Workshop, was the fact that the genomics community in forestry is small compared with other plant science communities (e.g., rice and Arabidopsis). As such, participants emphasized that groups conducting research in *Populus* should strive to organize their scientific agendas to complement, not compete with, one another. This coordination would enable groups to avoid unnecessary duplication and allow limited resources to be focused on specific problems that broadly challenge the community.

Returning again to the podium, Gerald Tuskan proposed the formation of an International *Populus* Genome Consortium, the purpose of which would be to develop and guide postsequencing activities in poplar. Foremost among the goals highlighted for the consortium would be the development of a *Populus* science plan. Teams would be formed to (1) examine genetic and genomic resources currently available to *Populus* researchers, (2) identify areas in which tools, techniques, and additional resources must be developed, and (3) assess applications and opportunities for future research associated with the completion of the *Populus* genome sequence. Applications would emphasize tree growth and development in the context of general poplar culture, basic science investigations, bio-based products and energy, carbon sequestration, and forest responses to changes in physical and chemical climates. The formation of an international consortium and the development of a science plan were endorsed strongly by symposium and workshop participants. A consortium World Wide Web site has been established (www.ornl.gov/ipgc).

FUTURE EVENTS

Two future meetings will focus on the molecular genetics and functional genomics of trees:

- International Plant and Animal Genome XI Conference, January 11 to 15, 2003, in San Diego, CA;
- Tree Biotechnology 2003, June 7 to 12, 2003, in Umeå, Sweden.

Although neither is specific to *Populus*, the sequencing of the poplar genome will be a topic of discussion at both meetings, as will the formation and ongoing functions of an International *Populus* Genome Consortium and the writing of the science plan.

CONCLUSIONS

As we have seen for rice and Arabidopsis, the sequencing of the *Populus* genome promises to enrich the study of forest biology. Used in conjunction with microarrays, metabolomics, high-efficiency transformation technologies, and high-throughput phenotyping, sequence data will enable researchers, in time, to attain a truly mechanistic understanding of tree function. Furthermore, access to the poplar genome will allow fundamental questions to be asked not only in tree biology but also in ecology, because this genus is distributed widely across the northern hemisphere and in Asia. However, targeted investments will be required to spur a genomics revolution in tree biology, forestry, and the ecological sciences.

ACKNOWLEDGMENTS

Special thanks to Stephen P. DiFazio for providing comments on an early version of this report. The International Poplar Symposium III was organized by the International Union of Forest Research Organizations and the Swedish University of Agricultural Sciences. The *Populus* Functional Genomics Workshop was organized

MEETING REPORT

by UPSC. ORNL is managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract DE-AC05-00OR22725.

Stan D. Wullschleger
Environmental Sciences Division
Oak Ridge National Laboratory
Oak Ridge, TN 37831-6422
wullschlegsd@ornl.gov

Stefan Jansson
Umeå Plant Science Center
Department of Plant Physiology
Umeå University, SE-901 87
Umeå, Sweden

Gail Taylor
School of Biological Sciences
University of Southampton
Bassett Crescent East
Southampton S016 7PX, UK

REFERENCES

Bradshaw, H.D., Ceulemans, R., Davis, J., and Stettler, R. (2000). Emerging model systems in plant biology: Poplar (*Populus*) as a model forest tree. J. Plant Growth Regul. **19,** 306–313.

Fiehn, O. (2002). Metabolomics: The link between genotypes and phenotypes. Plant Mol. Biol. **48,** 155–171.

Hertzberg, M., et al. (2001). A transcriptional roadmap to wood formation. Proc. Natl. Acad. Sci. USA **98,** 14732–14737.

Sheppard, L.A., Brunner, A.M., Krutovskii, K.V., Rottmann, W.H., Skinner, J.S., Vollmer, S.S., and Strauss, S.H. (2000). A DEFICIENS homolog from the dioecious tree black cottonwood is expressed in both female and male floral meristems of the two-whorled, unisexual flowers. Plant Physiol. **124,** 627–639.

Sterky, F., et al. (1998). Gene discovery in the wood forming tissues of poplar: Analysis of 5,692 expressed sequence tags. Proc. Natl. Acad. Sci. USA **95,** 13330–13335.

Taylor, G. (2002). *Populus*: Arabidopsis for forestry. Do we need a model tree? Ann. Bot. **90,** 677–687.