

## EDITORIAL

# Genomics and the tree physiologist

STAN D. WULLSCHLEGER,<sup>1,2</sup> G. A. TUSKAN<sup>1</sup> and S. P. DIFAZIO<sup>1</sup>

<sup>1</sup> Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6422, USA

<sup>2</sup> Author to whom correspondence should be addressed (wullschlegsd@ornl.gov)

Received March 27, 2002; accepted July 7, 2002; published online November 15, 2002

**Summary** During the last decade, a strong case has been made for viewing trees as model systems in plant biology. More recently, forest biologists have argued for the sequencing of the genome of a forest tree. Now, the United States Department of Energy has announced plans for sequencing the genome of the *Populus trichocarpa* clone, “Nisqually-1.” Thus, forest biology is poised to enter an exciting period of scientific discovery. Embracing new technology and new research paradigms, however, is never easy. It is, therefore, timely to ask how tree physiologists will take advantage of the *Populus* genome data. We contend that the most attractive opportunities will arise through genome-wide research designed to: (1) examine the differences between trees and herbaceous annuals; (2) explore questions relating to the temporal and spatial scales that characterize the life histories and growth of trees; and (3) investigate cause-and-effect relationships that are intractable to conventional research methodologies. To highlight the potential applications of genomics in tree physiology, we briefly discuss each of these approaches.

**Keywords:** *Arabidopsis*, biotechnology, functional genomics, molecular biology, *Populus*, whole-tree physiology.

## Introduction

The emergence of structural, functional, and comparative genomics has ushered in a new era in plant biology. The development of high-throughput genomics technologies has made possible the ongoing and detailed characterization of the genetic foundations of life. Out of this work, exciting ideas have emerged in areas as diverse as regulatory gene networks and evolutionary biology.

An important milestone in plant genomics was the sequencing of the complete genome of the higher plant, *Arabidopsis thaliana* L., which has been embraced as a model organism. Efforts have been initiated to sequence the genomes of other plants, including rice (*Oryza sativa* L.), a model cereal (Eckardt 2000), and *Medicago truncatula* Gaertn., a model legume (Frugoli and Harris 2001).

The forest genetics community has argued the case for viewing trees as model systems in plant biology (Bradshaw et al.

2000). This dialog has given rise to a convincing rationale for sequencing the genome of a forest tree, and the United States Department of Energy (DOE) has now announced plans to sequence the genome of a clone of the forest tree *Populus trichocarpa* Torr. & Gray (Tuskan et al. 2002). A member of the genus *Populus*, which includes cottonwoods and aspens, is well suited to serve as a model species because it has a relatively small genome—about 550 million base pairs, which is similar to rice, only 4× larger than *Arabidopsis*, and one fortieth the size of the *Pinus* genome. Other important attributes and resources include worldwide distribution of multiple species; high growth rates; ease of clonal propagation; availability of genetic maps and structured pedigrees; publicly accessible molecular markers, gene sequences, and bacterial artificial chromosome (BAC) libraries; and high-throughput plant transformation and regeneration capabilities. Finally, there exists a network of scientists who have, for many years, used *Populus* as a model to study tree biochemistry, physiology, morphology, genetics and molecular biology (Stettler et al. 1996).

The DOE Joint Genome Institute, with funding through the DOE Office of Biological and Environmental Research, will provide a 3× draft sequence of the female *Populus trichocarpa* clone “Nisqually-1” in late-2002 and a second 3× draft in mid- to late-2003. Draft sequence data are fragments of DNA whose approximate locations within the genome are known; however, the true order and orientation of these fragments within the genome are not. Thus, a draft sequence will be of a lower accuracy than a “finished sequence,” which might take several years to achieve. The project will generally employ a whole-genome shotgun approach. Genomic DNA will be fragmented into pieces of defined size, cloned into a plasmid vector and amplified, and sequence reads generated from a large number of sub-clones (cf. Green 2001). Sequence reads will be assembled into larger contiguous regions and these will be further assembled onto about 2000 BAC-end sequenced scaffolds. The final assembly will be accomplished using supplemental information from physical and genetic maps of the *Populus* genome. All sequence data will be freely available to the scientific community.

In the wake of this project, forest biology will enter an ex-

citing period of discovery. The outcome, however, depends on how well prepared forest scientists are to take advantage of tree genome data as they become available. Although genomics, especially functional genomics (Colebatch et al. 2002), offers tree physiologists many opportunities, we contend that the most significant benefits will be found in the search for answers to questions that are: (1) difficult to address in herbaceous model organisms; (2) specific to the temporal and spatial scales that characterize the life histories and growth of trees; and (3) otherwise intractable with conventional research methods. To highlight the potential application of genomics in tree physiology, we identify here a few of these questions.

### **Trees are not herbaceous annuals**

Trees can achieve great age and size. Bristlecone pine (*Pinus longaeva* DK Bailey), for example, can live to be more than 4700 years old, and coast redwood (*Sequoia sempervirens* (D. Don) Endl.) and giant sequoia (*Sequoiadendron giganteum* (Lindley) Buchholz) can exceed 100 m in height and 25 m in circumference, respectively. The great size and long life span of trees complicate almost every aspect of tree physiology studies: leaves are difficult to access for gas-exchange measurements; imposition of experimental treatments requires control of the conditions of large volumes of air or soil; significant inter- and intra-annual variations in growth due to changes in climate or pest populations limit the validity of conclusions from short-term studies; and the task of harvesting a mature tree for the purpose of quantifying biomass allocation to leaves, branches, stems and roots necessitates a Herculean effort. However, it is because of their great potential size and longevity that trees differ developmentally in so many respects from herbaceous plants, including patterns of apical dominance, perennial growth habit, secondary xylem (wood) formation, iterative development of complex crown form, phase change, dormancy cycles, seasonal nutrient reallocation and reproduction. Consequently, tree development is unlikely to be understood based solely on information from herbaceous systems. For example, secondary xylem formation in *Arabidopsis* can be a model for wood formation (Chaffey et al. 2002), but it is essential that results be placed in the context provided by parallel studies with trees. Chaffey and colleagues recognize this, and emphasize that the identification of genes involved in secondary xylem formation would be greatly accelerated if comparative studies between *Arabidopsis* and a model tree such as *Populus* were conducted.

### **Trees present unique spatial and temporal challenges**

Because of their longevity, trees are likely to experience relatively extreme climatic events. However, trees possess various means of adapting to environmental change. Respiration, for example, is known to acclimate to temperature; photosynthesis is thought to acclimate to atmospheric carbon dioxide concentration [CO<sub>2</sub>]; the vertical distribution of roots is responsive to the distribution of water and nutrients; and as trees

increase in size, they may respond to changes in light climate and water availability with changes in net CO<sub>2</sub> assimilation rate, leaf anatomy and canopy structure (Bond and Franklin 2002). Yet the metabolic pathways involved in these responses are not well understood. The same can be said for characteristics that distinguish northern and southern populations within the geographic range of a species, and the role that genetic adaptation to drought or temperature plays in determining species distribution. Such adjustments and adaptations could be probed with genomic tools.

Because of the size of the organism and the volume that it occupies, it is logistically difficult to manipulate trees and their environment experimentally. Therefore, we continually search for approaches that allow scaling up from small studies and extrapolation beyond experimental conditions. The tools of genomics are rarely considered in relation to the problems of scaling or extrapolation, but they may nonetheless prove useful in this context. Genomics is perceived as a reductionist approach to scientific investigation because it involves intricate dissection of the genetic code to the finest scale. However, because genomic data can be so comprehensive, they can be applied to problems at the scale of the whole organism, the plant community or entire ecosystems. As a result, genomics may provide a mechanism for translating understanding of cellular processes to scales that allow one to address questions in forestry and ecology. Although the potential of such an approach will likely take years to fully realize, the possibilities seem worth exploring now.

### **Overcoming the constraints of existing methods**

Traditionally, plant biologists have faced methodological tradeoffs between generality and precision, and researchers have either been forced, or have chosen, to work near the extremes of a continuum. At one extreme, the reductionist biologist has studied the action of one or a few genes on a specific biochemical process. At the other extreme, the generalist has conducted studies at the whole-plant level, integrating the effects of thousands of genes and thousands of gene products. Both approaches have their strengths. Plant biologists have, however, experienced considerable frustration in using fine-scale knowledge to predict phenomena at larger scales. Extrapolation (i.e., scaling) continues to be one of the greatest challenges in plant biology.

Today, genomics is providing tools that allow examination of whole-plant responses at single-nucleotide resolution, thus providing a bridge between cell biology and whole-plant physiology. For example, we have detailed information about the sequences and functions of many genes involved in photosynthesis and carbon and nitrogen metabolism. However, plants respond to chemical and climatic perturbations such as elevated [CO<sub>2</sub>] or nitrogen availability in ways that could not be predicted based solely on the genes specific to a single process. Armed with the tools of genomics, a complete genome and gene expression microarrays (Lockhart and Winzler 2000), gene activity for all relevant processes can be monitored under contrasting conditions. This will provide an inte-

grated view of plant responses at the finest scale. As an illustration of how this technology might be used, consider that physiologists have, for years, struggled with the interrelationships among carbon allocation, source–sink balance, and feedbacks to photosynthesis. Although some progress has been made in showing apparent down-regulation of the enzyme Rubisco under conditions of high carbohydrate status—a response indicative of source–sink feedback—a predictive model for the biochemical, physiological, and whole-plant processes that link source activity, sink strength, and photosynthetic rate has yet to emerge. The tools of genomics would free tree physiologists from the need to focus only on single genes or enzymes, and permit them to examine the simultaneous expression of hundreds or thousands of genes.

Expression arrays have already been used to assess qualitative differences between plants exposed to ozone and drought (Matsuyama et al. 2002, Seki et al. 2001), and it is only a matter of time before microarrays are used in a variety of physiologically relevant studies. Genomics tools will, of course, continue to be used to study the action of individual genes, and the availability of an assembled and annotated genome will further enrich such investigations. Nonetheless, the hope is that gene expression arrays will enable the actions of individual genes to be studied and interpreted in a broader context. For example, it should be possible to couple physiological measurements with gene expression profiles, thus illuminating the function of genes, biochemical pathways and cellular processes that are affected in a coordinated manner by a given set of conditions. Studies such as these would lay the groundwork for mapping regulatory networks and depicting linkages among gene products, biochemistry and whole-plant physiology. This could, in turn, facilitate predictive modeling of cellular processes and, with time, genome-based modeling of plant growth and physiology in response to a range of cultural, edaphic and climatic perturbations.

## Conclusions

Understandably, the world of genomics seems foreign to many forest scientists. The language is unfamiliar and there is seldom reference to anything resembling a whole plant. Fortunately, the true utility of genomics for the majority of researchers in forestry lies well beyond nucleotide sequences and encoded genes. Instead, its value resides in understanding how genes and gene products (i.e., proteomics) determine how cells function, how cells form organisms, and how those organisms respond to their environment. This is the realm of the tree physiologist, and as a discipline we should welcome, and when appropriate apply, all technologies that serve to move our science forward. Kramer (1948) recognized the value of embracing new frontiers when he challenged the post-World War II forestry community by writing: “It should be remembered that the more fundamental the research, the wider its possible practical applications. All that is needed is enough imagination to see the potential applications and to plan effective research that will produce results which can be applied.”

Although Kramer was not referring specifically to molecular biology or genomics, he nonetheless believed that a knowledge of mechanisms, described at any scale of organization, is essential to the optimization of forestry practice.

Writing from a similar perspective, we have argued that genome-wide research can lead to advances in the basic and applied goals of the tree physiologist. Within each of three broad areas, we have outlined how genomics could shed light on interesting, yet poorly understood processes. These examples are presented solely for the purpose of illustration and to stimulate discussion. However, each of us will have our own ideas about how genomics can be applied in tree physiology. As a result, the list of potential applications will be long and its contents will likely reflect our varied interests in processes that regulate structural development: vegetative and reproductive growth; photosynthesis and respiration; carbon and nitrogen metabolism; translocation; seed physiology; growth regulator synthesis and mode of action; and environmental and cultural factors that affect growth.

In conclusion, we believe that the real power of genomics lies in the fact that all known organisms share the same genetic code. Furthermore, sequences for many genes have been largely conserved through evolutionary time, so that information gained about the functioning of genes in one organism can be applied to distantly related organisms. For example, the functions of more than half the genes in *Arabidopsis* were inferred from experimental data gathered in animal, yeast and microbial model species (Somerville and Somerville 1999). Similarly, we can expect that, through comparative genomics, many of the rapid advances in understanding structural and regulatory genes in other model plants and organisms will translate directly to poplar and other tree species. And the reverse will also be true; research with *Populus* will aid in understanding the genetic regulation of metabolic processes in other plants and organisms. It should be recognized, however, that the course we chart based on our knowledge today is uncertain and it will be necessary to change direction as we learn more about genomics and the basic mechanisms that underlie the biochemistry, physiology and growth of trees. Nonetheless, we should begin to lay a strong scientific foundation for opportunities that are just now becoming visible on the horizon.

## Acknowledgments

We thank Drs. R.J. Norby, T.J. Tschaplinski, H.D. Bradshaw and S.H. Strauss for comments provided on an earlier draft of this paper, and Dr. R.C. Dahlman, United States Department of Energy, Office of Biological and Environmental Research, who drew our attention to the possibility of sequencing the *Populus* genome. Collaborations with the University of Washington, Genome British Columbia, and the Swedish Centre for Tree Functional Genomics are acknowledged. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract No. DE-AC05-00OR22725.

**References**

- Bradshaw, H.D., R. Ceulemans, J. Davis and R. Stettler. 2000. Emerging model systems in plant biology: Poplar (*Populus*) as a model forest tree. *J. Plant Growth Regul.* 19:306–313.
- Bond, B.J. and J.F. Franklin. 2002. Aging in Pacific Northwest forests: a selection of recent research. *Tree Physiol.* 22:73–76.
- Chaffey, N., E. Cholewa, S. Regan and B. Sundberg. 2002. Secondary xylem development in *Arabidopsis*: a model for wood formation. *Physiol. Plant.* 114:594–607.
- Colebatch, G., B. Trevaskis and M. Udvardi. 2002. Functional genomics: tools of the trade. *New Phytol.* 153:27–36.
- Eckardt, N.A. 2000. Sequencing the rice genome. *Plant Cell* 12: 2011–2017.
- Frugoli, J. and J. Harris. 2001. *Medicago truncatula* on the move! *Plant Cell* 13:458–463.
- Green, E.D. 2001. Strategies for the systematic sequencing of complex genomes. *Nature Rev. Genetics* 2:573–583.
- Kramer, P.J. 1948. Plant physiology in forest research. *J. For.* 46: 918–921.
- Lockhart, D.J. and E.A. Winzeler. 2000. Genomics, gene expression and DNA arrays. *Nature* 405:827–836.
- Matsuyama, T., M. Tamaoki, N. Nakajima, M. Aono, A. Kubo, S. Moriya, T. Ichihara, O. Suzuki and H. Saji. 2002. cDNA microarray assessment for ozone-stressed *Arabidopsis thaliana*. *Environ. Pollut.* 117:191–194.
- Seki, M., M. Narusaka, H. Abe, M. Kasuga, K. Yamaguchi-Shinozaki, P. Carninci, Y. Hayashizaki and K. Shinozaki. 2001. Monitoring the expression pattern of 1300 *Arabidopsis* genes under drought and cold stresses by using full-length cDNA microarray. *Plant Cell* 13:61–72.
- Somerville, C. and S. Somerville. 1999. Plant functional genomics. *Science* 285:380–383.
- Stettler, R.F., H.D. Bradshaw, Jr., P.E. Heilman and T.M. Hinckley. 1996. *Biology of Populus and its implications for management and conservation*. NRC Research Press, Ottawa, Canada, 539 p.
- The *Arabidopsis* Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815.
- Tuskan, G.A., S.D. Wullschleger, H.D. Bradshaw and R.C. Dahlman. 2002. Sequencing the *Populus* genome: Applications to the energy-related missions of DOE. *Plant, Animal and Microbial Genomes X*, San Diego, CA, 40 p.