

Poplar Genomics is Getting Popular: The Impact of the Poplar Genome Project on Tree Research

G. A. Tuskan¹, S. P. DiFazio¹, and T. Teichmann²

¹ Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6422, USA

² Forstbotanisches Institut, Georg-August-Universität Göttingen, Göttingen, Germany

Received: October 13, 2003; Accepted: October 30, 2003

Abstract: Trees, due to their long life-span, have characteristics that distinguish them from annual, herbaceous plants. It is likely that many of these properties are based on a tree-specific genetic foundation. The U.S. Department of Energy initiated a genome-sequencing project for *Populus*, a model perennial plant. Through international collaborations and input to the sequencing effort, the annotated whole genome sequence of *Populus trichocarpa* will be released to the public in early 2004. This genomic resource will, for the first time, allow comparison between a perennial and an annual plant on a whole genome basis and therefore provide clues for molecular research on tree-specific questions like dormancy, development of a secondary cambium, juvenile-mature phase change, or long-term host-pest interactions. The approximately 520 Mbp of annotated genomic sequence will complement and expand the knowledge provided so far by the 125 000 ESTs from poplar that are available in public databases. This article introduces the international poplar research programmes and points out the significance of the poplar genome project for plant research.

Key words: Poplar, whole genome sequence, model tree.

Introduction

Trees are unique among plant species in that they display distinctive developmental, physiological, and anatomical characteristics, such as phase change from juvenile to mature state, age-dependent adaptation to the environment, and a secondary cambium that produces wood and bark (Bradshaw et al., 2000; Teichmann, 2001; Wullschleger et al., 2002). Trees have a long life cycle and adapt their physiologic state to the seasons, alternating between dormancy and active growth. These properties are not found in non-perennial herbaceous plants, and it is likely that most of these characteristics are caused by a distinctive genetic foundation and/or tree-specific temporal and spatial gene expression patterns. Many of these tree-specific characteristics provide commercially important commodities (e.g. wood) that have a multitude of direct and indi-

rect domestic applications, ranging from furniture to renewable energy.

In order to build the scientific basis for the molecular biology of woody plants, poplar was chosen as a model tree, mainly due to its small genome size of ca. 520 Mbp and the availability of routine transformation systems (Tuskan et al., 2002). Additionally, saturated genetic maps, numerous structured pedigrees, and several BAC libraries are available. Large EST sequencing projects have been completed, and there are now approximately 125 000 ESTs in public databases and 75 000 more will be released in the near future. As a result, microarrays for gene profiling have been developed and deployed. Finally, the wide degree of variation among the various poplar species in their adaptation to the environment (e.g. tolerance to drought and salt) or in their habit (e.g. crown architecture) offers the chance to isolate genes from woody plants important for ecological and/or economic purposes.

In May 2002, the U.S. Department of Energy (DOE) initiated an international project to sequence the genome of a *Populus* tree within 18 months. A female *Populus trichocarpa* (Fig. 1), named Nisqually-1 (after its habitat along a Washington state river), was chosen for the sequencing project. The draft sequence is now publicly available (www.genome.jgi-psf.org/poplar0/poplar0.home.html) and the International *Populus* Genome Consortium (IPGC) has been founded to help guide post-sequencing activities in poplar (www.ornl.gov/ipgc/).

The Sequencing Project

The *Populus* genome was sequenced by DOE's Joint Genome Institute (JGI) to a depth of 6X using a whole genome shotgun approach. The raw sequence database consists of paired end-reads, averaging ca. 620 bp of high quality sequence from a 3-kb library (~1.9 million sequence reads) and an 8-kb library (~3.2 million sequence reads). Assembly of these sequences into oriented contigs was aided by end sequences from a 10X randomly sheared fosmid library (~350 000 sequences), and fingerprinting and end sequencing of a 10X BAC library (~90 000 sequences), made available by Genome Canada. A comprehensive genetic map, created at Oak Ridge National Laboratory (ORNL), and the physical map are being linked using a series of framework microsatellite (SSR) markers distributed across the 19 *Populus* chromosomes at 5 cM resolution. Thus, much of the genome sequence will be associated with



Fig. 1 Seedling of *Populus trichocarpa*, Nisqually 1.

large scaffolds representing the 19 *Populus* chromosomes. The worldwide collection of ESTs, brought together by the Umeå Plant Sciences Centre, is being used to generate a basal annotation of the assembled genome. Several hundred full-length cDNAs, sequenced by ORNL and Genome Canada, are being used to customize three autonomous *de novo* gene-calling algorithms, GRAIL at ORNL, EUGENE at the University of Ghent, and FgeneH at JGI. These algorithms will be used to refine the basal annotation. The draft genomic resource will be reviewed at an annotation jamboree in early 2004, after which the assembled and annotated database will be released to the public through a web-based portal hosted at JGI.

Advantages of a Whole Genome Approach

The sequencing of the complete genome of *Arabidopsis* illustrates the impact genome projects have on the study of a model organism. The *Arabidopsis* genome-sequencing project galvanized an active and burgeoning community, catalyzing a cascade of synergistic collaborations. This is exemplified by the CATMA project, where gene-specific primers have been used to amplify fragments of all *Arabidopsis* genes with the goal of obtaining a whole genome microarray; and the 2010 project, an initiative aimed at functional characterization of all *Arabidopsis* genes by the end of the decade. Similar communal efforts are being spawned by the *Populus* genome-sequencing project, as exemplified by the science plan authored by the International *Populus* Genome Consortium (see below).

The availability of the *Populus* genome sequence will also allow researchers to compare the genome of a herbaceous plant with the genome of a perennial plant. From the comparison of the *Populus* genome to the *Arabidopsis* genome, it will be immediately obvious if a gene of interest belongs to a known gene family or represents a class of genes that is unique to trees. These unique genes will provide insights into the developmental and evolutionary processes that distinguish perennial, woody plants. In addition, the analysis of EST databases of poplars with special characteristics (e.g. salt tolerant *P. euphratica*, or vegetative root propagation in *P. tremuloides*), in comparison to the whole genome sequence of *P. trichocarpa*, will point to genes that may be involved in the adaptation of a tree to extreme environments, such as deserts or habitats with frequent fires. The orthologs of genes that have been identified in *Arabidopsis* and which have commercial applications in poplar can be isolated using the sequence information of the poplar genome project. Finally, determination of the whole genome sequence will initiate the post-genomic era of forestry research, including the analysis of the full complement of proteins and metabolites of *Populus*.

International Poplar Research

Populus genomic research has long had an international flavour, and these collaborations are strengthening in the post-genomic era. Several international *Populus* research consortia have contributed to the poplar genome project, and will benefit from the availability of the whole genome sequence of poplar.

Current poplar databases and resources

Genome Canada initiated a project that involves wood formation and pest resistance of poplar, and has included some genetic mapping, as well as fingerprinting and end sequencing of BACs (as noted above). The Umeå Plant Science Centre, UPSC, and Royal Institute of Technology in Stockholm have produced the majority of poplar ESTs and have pioneered *Populus* microarray work. Additional poplar ESTs come from various institutions: 15 000 ESTs of *P. euphratica* from leaves and roots exposed to a variety of stresses have been sequenced within the framework of the EU project ESTABLISH at the Institute of Biotechnology, University of Helsinki; 12 000 ESTs from *P. tremuloides* at the Michigan Technological University in Houghton, and 7000 ESTs from roots of *P. × canescens* at INRA, Nancy. The combined data from the EST databases and the poplar genome will facilitate the production of a poplar whole genome array, containing all the predicted genes of poplar. In addition, a database of all mapped *Populus* SSRs is available at: www.ornl.gov/ipgc/Links.htm.

EU projects

During the Fifth Framework Programme of the European Union, the projects EUROFACE (www.lukac.plus.com/euroface/) POPYOMICS (www.soton.ac.uk/~popyomic/), ESTABLISH (www.user.gwdg.de/~establis/), ROST, POOPWOOD (www.isv.cnrs-gif.fr/cr/popwood/) and COPOL (www2.rhnc.ac.uk/~uhbc006/CopolWEB/copol.htm) were initiated. EUROFACE deals with the physiological and molecular responses of several poplar species to an increase in atmospheric CO₂. POPYOMICS employs molecular genetic mapping and genomics to identify genes that determine yield and disease resistance in *Populus*, with the aim of using this species as a sustainable

source of bioenergy and timber. Objectives of ESTABLISH and ROST are to improve stress tolerance of poplar. For that purpose ESTABLISH employs the salt-tolerant *Populus euphratica*, while ROST uses *Arabidopsis* as a model system and aims at transferring this knowledge into poplar. POPWOOD and COPOL focus on the xylem differentiation and secondary growth of poplar to improve wood properties for commercial use. Germany has focused its research on the molecular biology of poplar in a joint effort involving seven groups (PRG – Poplar Research Group, Germany). The main objective of these groups is fundamental research on the interaction between nutrition and adaptation to environmental stresses using microarrays, RNAi, and laser micro-dissection techniques (www.pappelgruppe.de).

Commercial and Conservation Applications

With respect to public acceptance, tree biotechnology is at a turning point. Tree biotechnology does not carry the stigma of novel engineered food and, as a result, potential products from genetically modified trees may be successful in the market. In Europe, the 1998 moratorium on the release of transgenic plants has been suspended. The new regulations are very strict, but in principle make the commercial cultivation of transgenic plants possible within the European Union. The availability of the poplar genome, combined with gene expression data from microarray studies, offers the opportunity to study and modify genes related to commercially important traits. Companies like SweTree Technologies in Sweden and ArborGen in the U.S. are already aiming at improving wood for pulp and timber applications using genetic engineering of trees.

Access to the *Populus* genome will also provide an opportunity to better understand the distribution of genes across the landscape, facilitating the conservation of unique populations, gene complexes, and/or genes themselves. The genomic database will provide a wealth of information on genome structure and evolution, along with enhanced understanding of the genetic architecture of adaptation in natural populations. Coupled with insights on gene flow and other population processes, these data will provide unprecedented insights into community diversity and evolution. Preservation, containment and isolation measures can then be undertaken, based on an understanding of the whole genome and not just on sampled measures of the phenotype.

International *Populus* Genome Consortium

The IPGC, in its first 6 months, attracted ca. 220 members from 32 countries, and the consortium website (www.ornl.gov/ipgc/) is receiving 400 hits per day (September 2003). The purpose of the IPGC is to help develop and guide post-sequence activities in poplar. Foremost among the list of goals will be the development of a comprehensive *Populus* science plan. The long-term goals of the science plan are to examine the suite of genetic resources in *Populus* that are currently available to the scientific community, integrate genomics with physiology and ecology in an effort to understand and influence tree growth, development, and function, and develop the ability to attain predictive understanding of tree growth, development and complex function. A subset of consortium members are currently completing the draft science plan that will be reviewed, modified, and approved by the entire consortium by early 2004. The science plan includes panels on: bio-

informatics, genetic resources, metabolomics, microarrays, phenotyping, physical mapping, proteomics, and tissue culture and transformation.

Perspective

With the imminent creation of the *Populus* Genome Resource comes the promise and possibility of uncovering and understanding mechanisms uniquely associated with perennial woody plant growth, development, and ecology. Despite the fact that woody perennial plants contain over two-thirds of the terrestrial biomass, produce the majority of our structural building materials, and supply an indispensable portion of the world's energy needs, woody perennial plants are relatively poorly understood. With a fully developed set of genomics tools and resources we will now be able to address the issues related to inter-annual cycling of nutrients, water movement up dozens of meters in height, decade-long host-pest interactions, perennial crown development, wood formation and maturation at a foundational level. Characterizing these and other perennial plant functions will provide insights that should lead us from simple descriptive understanding of woody plant growth to a more predictive understanding of perennial plant biology.

Acknowledgements

Financial support from the "Deutsche Forschungsgemeinschaft" (DFG-Forschergruppe FOR 496/1) and the European Union (ESTABLISH QLK5-CT-2000-01377) is gratefully acknowledged. Partial support for writing this manuscript was provided by Oak Ridge National Laboratory. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S. Department of Energy, under contract DE-AC05-00OR22725. The *Populus* Genome Sequencing effort is supported by the U.S. Department of Energy, Office of Biological and Environmental Research, through funding to the Joint Genome Institute, Walnut, CA, USA.

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.
- Teichmann, T. (2001) The biology of wood formation: Scientific challenges and biotechnological perspectives. In: *Recent Res. Devel. Plant Physiol.* 2, Trivandrum, India: Research Signpost, 269–284.
- Tuskan, G. A., Wulschleger, S. D., Bradshaw, H. D., and Dahlman, R. C. (2002) Sequencing the *Populus* Genome: Applications to the Energy-Related Missions of DOE. San Diego, CA: Plant, Animal and Microbial Genomes X, p. 40.
- Wulschleger, S. D., Tuskan, G. A., and DiFazio, S. P. (2002) Genomics and the tree physiologist. *Tree Physiol.* 22, 1273–1276.

T. Teichmann

Forstbotanisches Institut
Georg-August-Universität Göttingen
Büsgenweg 2
37077 Göttingen
Germany

E-mail: tteichm@gwdg.de

Section Editor: H. Rennenberg