

National Science Foundation-Sponsored Workshop Report: "The 2010 Project"

Functional Genomics and the Virtual Plant. A Blueprint for Understanding How Plants Are Built and How to Improve Them

Joanne Chory¹, Joseph R. Ecker¹, Steve Briggs, Michel Caboche, Gloria M. Coruzzi, Doug Cook, Jeffrey Dangel, Sarah Grant, Mary Lou Guerinot, Steven Henikoff, Rob Martienssen, Kiyotaka Okada, Natasha V. Raikhel, Chris R. Somerville, and Detlef Weigel

Howard Hughes Medical Institute/The Salk Institute for Biological Studies, Plant Biology Laboratory, San Diego, California 92186-5800 (J.C.); University of Pennsylvania, Department of Biology, Philadelphia, Pennsylvania 19104-6018 (J.R.E.); Novartis Agricultural Discovery Institute, La Jolla, California 92037 (S.B.); Unité de Recherche en Génomique Végétale Institut National de la Recherche Agronomique, 78026 Versailles cedex, France (M.C.); New York University, New York, New York 10003 (G.M.C.); Texas A&M University, College Station, Texas 77843 (D.C.); University of North Carolina, Chapel Hill, North Carolina 27599 (J.D., S.G.); Dartmouth College, Hanover, New Hampshire 03755-3576 (M.L.G.); Fred Hutchinson Cancer Research Center, Seattle, Washington 98109 (S.H.); Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724-2212 (R.M.); Kyoto University, Kyoto 606-8502, Japan (K.O.); Plant Research Laboratory, Michigan State University, East Lansing, Michigan 48824-1312 (N.V.R.); Carnegie Institute of Plant Biology, Stanford University, Stanford, California 94305-4101 (C.R.S.); and The Salk Institute for Biological Studies, La Jolla, California 92037 (D.W.)

Although future developments in a rapidly emerging field can be difficult to predict, some projections and evaluations are helpful in planning for the development of community resources and other initiatives that may facilitate progress. To make these projections and evaluations, a group of scientists met at The Salk Institute for Biological Studies, La Jolla, CA, on January 13 and 14, 2000. The workshop, co-chaired by Joanne Chory and Joseph Ecker, included the following group of scientists: Steven Briggs, Michel Caboche, Gloria Coruzzi, Doug Cook, Jeff Dangel, Sarah Grant, Mary Lou Guerinot, Steven Henikoff, Rob Martienssen, Kiyotaka Okada, Natasha Raikhel, Chris Somerville, and Detlef Weigel. To give the report from this group the widest possible distribution, a summary is printed below. The complete text of this document and an accompanying report from a meeting held on November 23 and 24, 1998, at the Carnegie Institution of Washington (Stanford, CA) are posted on the The Arabidopsis Information Resource Web site (<http://www.Arabidopsis.org>). The Carnegie-Stanford workshop was co-chaired by Chris Somerville and Joe Ecker and included the following group of scientists: Roger Beachy, Charles Gasser, Steve Kay, Kenneth Keegstra, Rob Martienssen, Susan McCouch,

Elliot Meyerowitz, Eric Ward, and Susan Wessler. The idea of "The 2010 Project" is a synthesis of the two workshops (Carnegie-Stanford and San Diego-Salk) and a result of the work of a group of people with diverse scientific backgrounds. The ultimate goal of The 2010 Project is to broaden the scope to consider how genomics could be applied to understand all plants. Since *Arabidopsis* is far ahead of all other plant systems, many initiatives are focused on *Arabidopsis*. However, an implicit goal is to begin asking how and what we need to do to bring other plants forward as efficiently as possible, taking advantage of the power of a model system. We invite the community of plant biologists at large to read both reports and to express your views and offer your suggestions; these will be posted on the The Arabidopsis Information Resource Web site. Your suggestions can also be sent via e-mail to the following address (project2010@aspp.org) and they will be posted on the *Plant Physiology* Web site at the American Society of Plant Physiologists (<http://www.aspp.org>). Additional community input will be sought at the Arabidopsis meeting in Madison, Wisconsin, in June 2000. We will take into consideration all comments and suggestions in preparing the final workshop report to be submitted to the National Science Foundation this summer. Your opinions are very important and we are looking forward to a discussion of the important issues raised by this report.

¹J.C. and J.R.E. were co-chairpersons of The 2010 Project Workshop held at The Salk Institute.

About 10 years ago, plant scientists adopted the widespread use of an easily manipulated model called *Arabidopsis*, and established an international research effort called The Multinational Coordinated *Arabidopsis* Genome Research Project. As an outcome of this effort, the entire DNA sequence of this plant will be completed by mid-year in 2000. For the first time, we will know the sequence of 25,000 genes that are necessary for a plant to function as a flowering plant. As a follow-up to the *Arabidopsis* genome sequencing efforts, an important and revolutionary new initiative—to exploit the revolution in plant genomics by understanding the function of all genes of a reference species within their cellular, organismal, and evolutionary context by the year 2010 (The 2010 Project)—has been proposed by the community of plant biologists. Implicit in this mission statement is an endorsement of the allocation of resources to attempts to assign function to genes that have no known phenotype, which represents a significant departure from the common practice of defining and justifying a scientific goal based on the biological phenomena. The rationale for endorsing this radical change is that for the first time it is feasible for plant biologists to envision a whole-system approach to the study of plant form and function. The committee discussed the feasibility of commencing a publicly funded program to determine the function of all *Arabidopsis* genes during the next decade, using a systems approach. The committee has identified specific project goals that include: the development of expanded genetic toolkits as a service to the research community; a whole-systems approach to identification of gene function from the molecular to evolutionary levels; an expanding of the role for bioinformatics; the development of human resources; and international collaboration.

MISSION STATEMENT

To exploit the revolution in plant genomics by understanding the function of all genes of a reference species within their cellular, organismal, and evolutionary context.

LONG-TERM GOAL

In order to most efficiently and safely manipulate plants to meet growing societal needs, we must create a wiring diagram of a plant through its entire life cycle: from germinating seed to production of the next generation of seeds in mature flowers. These processes are guided by genes and the proteins they encode. They are directed by both intrinsic developmental cues and environmental signals. The long-term goal for plant biology following complete sequencing of the *Arabidopsis* genome is to understand every molecular interaction in every cell

throughout a plant lifecycle. In essence, to understand the function of every gene by the year 2010.

The ultimate expression of our goal is nothing short of a virtual plant which one could observe growing on a computer screen, stopping this process at any point in that development, and with the click of a computer mouse, accessing all the genetic information expressed in any organ or cell under a variety of environmental conditions. Completion of the *Arabidopsis* genome sequence provides significant leverage for future plant genome projects. The reference genome is a platform from which useful comparisons are simplified. We will ultimately be able to predict the evolution of new gene function by comparative genomics with other key plant species.

WHERE WILL THE 2010 PROJECT LEAD US IN THE FUTURE?

Whole-systems-based knowledge of the entire biology of a reference species confers predictive power that will enable the following:

1. Predictable outcomes to directed experimental genetic changes.
2. Directed genetic changes that accelerate domestication of wild species.
3. Facile genetic manipulation that ensures maintenance of, and expansion of germplasm bases.
4. A description of the underlying mechanisms of heterosis, and the ability to use this phenomenon more effectively.
5. Enhanced understanding of the genetic basis of phenotypic plasticity, which will have a profound impact not just in plants, but also in animals, including humans.
6. Knowledge of the minimum gene set required for plant life.
7. Understanding of the genetic basis of plant evolution, which will enrich our understanding of the diversity of life on earth.
8. An understanding of interactions between plants and other organisms in their environment, up to the level of ecosystems.

SCIENTIFIC OBJECTIVES

New experimental tools that investigate gene function at the subcellular, cellular, organ, organismal, and ecosystem level need to be developed. New bioinformatics tools to analyze and extract meaning from increasingly systems-based datasets will need to be developed. These will require, in part, creation of entirely new tools. An important and revolutionary aspect of The 2010 Project is that it implicitly endorses the allocation of resources to attempts to assign function to genes that have no known function. This represents a significant departure from the common practice of defining and justifying a sci-

tific goal based on the biological phenomena. The rationale for endorsing this radical change is that for the first time it is feasible to envision a whole-systems approach to gene and protein function. This whole-systems approach promises to be orders of magnitude more efficient than the conventional approach.

Exploitation of the Arabidopsis genome sequence to achieve the goals outlined above will require significant new investment in dedicated, focused, genome technology centers. These centers should be both dispersed and single site. They will serve the research community at large by providing services and by producing new tools using economies of scale. The centers will be dedicated to the creation of genome-wide tools, rather than to the application of genome-wide tools to solving specific research problems. They will thus be explicitly public service oriented, as was the Arabidopsis genome sequencing project.

In creating genome-wide tools, the centers must complement and significantly enable investigators throughout the U.S. and the world. Individual investigators will be at once the main clientele for the

centers and the dispersed creators of knowledge. The value of this project therefore depends on the availability of significant support for individual research laboratories throughout the plant biology research community to leverage investment in both the Arabidopsis genome sequencing project and the proposed centers to solve a wide range of specific biological problems.

The committee has discussed the impact that this project would have on the progress of basic plant research, as well as strategic interests of the U.S. as they relate to agriculture, energy, the environment, and human health. Over the last 10 years, The Multinational Arabidopsis Steering Committee and the North American Steering Committee have been important liaisons between the community at large and policy makers in North America, Europe, and the Pacific Rim. To date, the Arabidopsis community is a model of how these structures can minimize duplication of efforts and enhance research resources worldwide. The committee recommends that The 2010 Project be implemented as soon as possible and we solicit your comments (see above).