

# NATIONAL PLANT GENOME INITIATIVE

## Progress Report January 2006

National Science and Technology Council Committee on Science Interagency Working Group on Plant Genomes



The National Plant Genome Initiative (NPGI) was established in 1998 as a coordinated national plant genome research project. The completion of the eighth year marks the third year of the Initiative under the second NPGI Five-Year Plan entitled, "National Plant Genome Initiative: 2003-2008" (http://www.ostp.gov/NSTC/html/npg/2003/index.htm). The six major objectives are:

- Continued Elucidation of Genome Structure and Organization
- Functional Genomics
- Translational Plant Genomes
- Bioinformatics
- Education, Training and Outreach
- Consideration of Broader Impacts

Coordination of activities, guidance and oversight is provided by the Interagency Working Group on Plant Genomes (IWG) under the auspices of the Committee of Science of the National Science and Technology Council with the Office of Science and Technology Policy (OSTP). Participating agencies include the National Science Foundation (NSF), Department of Agriculture (USDA), Department of Energy (DOE), National Institutes of Health (NIH), National Aeronautic and Space Administration (NASA), Agency for International Development (USAID), OSTP and the Office of Management and Budget (OMB). Each agency participates in the NPGI in a manner consistent with its specific mission.

### EXECUTIVE OFFICE OF THE PRESIDENT NATIONAL SCIENCE AND TECHNOLOGY COUNCIL WASHINGTON D.C. 20502

## Dear Colleague:

This annual report of the National Plant Genome Initiative (NPGI) highlights the progress made in 2005 by the U.S. plant genomics research community. The report was prepared by the National Science and Technology Council's Interagency Working Group on Plant Genomes (IWG), which coordinates and provides oversight for the Federal investment in plant genome research through NPGI.

Most significant among NPGI's accomplishments in 2005 is the initiation of the maize genome sequencing project. Although sequencing the maize genome has been a goal of the NPGI for some time, this joint interagency effort to develop a high-resolution picture of the maize genome became technically feasible only in the last year. Whole genome sequence information will advance our understanding of maize biology, which will allow us to develop new and improved varieties as well as new uses for maize.

In its 8th year, the NPGI continues to support the goals established in the January 2003 long-range plan to accelerate advances in plant genomics. As is summarized in this report, NPGI has made a number of new genomics research resources and tools widely available for use in the study of plant function, resilience, vigor and development. In addition, NPGI has expanded the scope of Federally supported plant genome research activities to include applications in environmental biology and biodiversity. The cornerstone of NPGI's success continues to be the open, free and rapid access to all information and materials generated by Federally-supported plant genome research.

The IWG provides oversight for the NPGI to gamer interagency support and cooperation for plant genome sciences and ensure progress toward meeting the goals set forth in the long-term plan. Under the IWG's stewardship, the U.S. has become and continues to be the world leader in plant genomics research and its applications to agriculture, health, energy and the environment.

Sincerely,

Committee on Science Co-Chairs

Sharon Havs Chief of Staff Office of Science and Technology Policy

Arden L. Bement, Jr. Director National Science Foundation

Elias A. Zerhouni Director

National Institute of Health

Interagency Working Group on Plant Genomes **Committee on Science** National Science and Technology Council Co-Chairs

Mary E. Clutter\*, National Science Foundation (until September 2005) James Collins\*\*, National Science Foundation (from October 2005) Joseph Jen, U.S. Department of Agriculture

Members Robert B. Bertram, U.S. Agency for International Development Noah Engelberg, Office of Management and Budget Terri Lomax, National Aeronautics and Space Administration Anna Palmisano, U.S. Department of Agriculture Jane L. Peterson, National Human Genome Research Institute Judy St. John, U.S. Department of Agriculture Walter Stevens, U.S. Department of Energy Kurt A. Zuelke, Office of Science and Technology Policy

Executive Secretary Machi F. Dilworth, National Science Foundation



## THIS REPORT INCLUDES BRIEF HIGHLIGHTS OF ACTIVITIES IN THE PAST YEAR ILLUSTRATING THE RAPID ADVANCES IN PLANT GENOMICS RATHER THAN AN EXHAUSTIVE SUMMARY OF ONGOING PROJECTS BEING FUNDED, SUPPORTED AND COORDINATED BY THE NPGI MEMBER AGENCIES.

CURRENT STATUS OF PLANT GENOMES SEQUENCES:					
Plant Species	Genome Size	Products Public Access to sequence	Project Status	Pls, Institutions, Countries	
Arabidopsis thaliana	130 Mb	Finished BAC-by-BAC whole genome sequence. Published in Nature (2000) http:// www.nature.com/nature/links/001214/001214- 1.html	Completed	Arabidopsis Genome Initiative	
Rice (Oryza sativa)	430 Mb	Finished BAC-by-BAC whole genome sequence. Published in Nature (2005) http:// www.nature.com/nature/journal/v436/n7052/ abs/nature03895.html	Completed	International Rice Genome Sequencing Project	
Poplar (Populus trichocarpa)	480 Mb	7.5x coverage whole genome shotgun sequence in GenBank http://genome.jgi-psf. org/Poptr1/Poptr1.home.html	In finishing	Joint Genome Institute	
Barrel Medic (Medicago truncatula)	550 Mb	Gene-rich BAC sequence in GenBank http:// www.medicago.org/, http://medicago.toulouse. inra.fr/	In progress	Medicago Genome Sequencing Project	
Tomato	950 Mb	Gene-rich region BAC sequence in GenBank http://www.sgn.cornell.edu/help/about/tomato_ sequencing.pl	In progress	International Tomato Genome Sequencing Project	
Sorghum	736 Mb	Whole genome shotgun sequence <i>http://www.jgi.doe.gov/sequencing/why/CSP2006/sorghum.html</i>	In progress	Joint Genome Institute	
Grape	500 Mb	Whole genome shotgun sequence http://www. vitaceae.org/molecular.html	In progress	INRA, France	
Maize	2,500 Mb	BAC and whole genome shotgun sequences will be deposited in GenBank	3 year projects Started in Nov 2005	Washington University, St. Louis; Cold Spring Harbor Laboratory; University of Arizona; Iowa State University; Joint Genome Institute; Stanford University	

## **DOMESTICATION OF CORN HAD A BIG IMPACT INVOLVING RELATIVELY FEW GENES** The corn cultivated across the US as a major

crop was domesticated from a wild grass relative, teosinte, about 6,000 years ago. Teosinte looks very different from corn. The major changes in plant architecture include a reduction in branching, a decrease in the number of ears and an increase in ear size. A group led by investigators at the University of Wisconsin, Madison recently reported that while these phenotypic changes are dramatic, only about 1,200 of the estimated 50,000 – 60,000 genes of corn appear to be involved in domestication. Some of these genes have been identified and characterized in detail. One gene, *teosinte branched 1 (tb1)*, is responsible for regulation of side branching while the *ramosa* genes regulate the branching of the tassels and ears.

GENOME-ENABLED MODIFICATION OF POPLAR ROOT **DEVELOPMENTFORINCREASED**CARBONSEQUESTRATION

Trees play a significant role in collecting carbon from the air and sequestering it in the soil. A group of investigators at Oak Ridge National Laboratory, Purdue University, and Michigan Technological University are using the poplar genome sequence to identify key genetic regulators that might enhance lateral root proliferation in poplar.

> The project will contribute to our basic understanding of the genomic basis for root architecture and may lead to the development of strategies to increase carbon storage in soils through roots with enhanced capacity to sequester carbon.

> > UNDERSTANDING LATE POTATO BLIGHT -THE **CAUSE OF THE IRISH POTATO FAMINE** Late Potato Blight is once

again becoming a problem in agriculture more than a century after the Irish Potato Famine. Outbreaks are still reported regularly around the world and throughout the US and new strains are emerging that no longer respond to traditional chemical controls. The causative agent is Phytophthora infestans, an oomycete or "water mold" that has features of fungi and algae. Researchers at Ohio State University recently made a breakthrough in understanding how this pathogen expresses certain avirulence genes that enable it to infect plants like potatoes.

Plants that have resistance to Late Potato Blight express genes that can ward off infection by detecting and counteracting these avirulence genes. These findings will be important for development of disease resistant potatoes as well as for dealing with related Phytophthora species causing Sudden Oak Death and Soybean Root Rot.

HOW DO PLANTS GROW? Like animals, plants produce hormones. One class of plant hormone, called "auxins" plays a role in many aspects of plant growth and development, including cell enlargement, formation of buds, roots, flowers, and fruit, and autumnal loss of leaves. While the roles of auxins in plant development have been described in detail, it has been very difficult to determine how auxins actually work. Researchers at Indiana University, Bloomington, used genomics tools to find a molecular receptor of auxin called TIR 1 in the model plant Arabidopsis. They further showed that binding of auxin to TIR1 forms a complex that allows de-repression of genes required for growth. This finding is a first step towards understanding how auxin regulates so many different developmental processes in plants and can be translated into economically important crop plants. Other research conducted at the University of Wisconsin, Madison recently uncovered the three dimensional structure of the plant photoreceptor phytochrome, revealing an unusual knot structure that enables it to act as a molecular switch regulating many aspects of plant growth and development.

NOVEL TOCOPHEROL (VITAMIN E) PROFILES IN SUNFLOWER Sunflower (Helianthus annuus L.) produces naturally high concentrations of a-tocopherol (Vitamin E) in seeds and leaves. Researchers at the University of Georgia are studying the synthesis of  $\alpha$ -tocopherol using induced mutations to greatly increase  $\beta$ -,  $\gamma$ -, or  $\delta$ -tocopherol levels. Their findings suggest that intragenic and intergenic effects can broaden the spectrum of novel tocopherols produced in sunflower and reveal strategies underlying manipulation of seed oils of other genera.

AGRICULTURAL

The

COORDINATED

*www.uark.edu/ua/ricecap/)* 

**PROJECTS (CAP)** RiceCAP (http://

University of Arkansas, Fayetteville

is leading a project to coordinate a multi-institutional research, education and extension effort to study sheath blight resistance and milling quality traits. The project will develop new technology for production rice educate and the generation next of scientists in use this technology of for plant breeding. Extension efforts will inform farmers and consumers of the value of using genomic information to improve agricultural crops. RiceCAP is utilizing the recently completed rice genome sequence, and other tools and resources to solve problems in rice production and improvement that cannot easily be solved by traditional plant breeding. WheatCAP The goal of this new project is to transfer new developments in genomics to wheat improvement by combining the expertise of genomics researchers, wheat breeders, educators, extension specialists and end-users. Led by the University of California, Davis, WheatCAP will establish a multi-institutional network of public wheat breeding programs and high throughput genotyping laboratories. This network will discover high value Quantitative Trait Loci (QTL) and genes and rapidly deploy them into adapted varieties. Gene mapping and QTL information will be incorporated into the GrainGenes, Gramene and Marker Assisted Selection (MAS) wheat databases. The wheat populations used for mapping will be deposited into the USDA Small Grain Collection, providing long-term public access of genetic resources for the wheat community.

**PHYTOREMEDIATION OF IONIC AND METHYLMERCURY POLLUTION** Scientists at the University of Georgia, Athens are applying genomic techniques to develop new strategies for using plants to remediate mercury contamination in the environment. Multiple genes are being introduced in concert into plants, to enhance mercury resistance or accumulation from soil, as well as to accelerate conversion of mercury into less toxic volatile forms. Data are beginning to emerge from field-testing of transgenic poplars for their ability to detoxify and extract mercury from polluted soil and water.



**MARKER ASSISTED SELECTION WORKSHOP FOR PLANT BREEDERS** A Marker Assisted Selection (MAS) workshop for plant breeders was held at the University of Arkansas Research and Extension Station and the USDA National Rice Research Center in Stuttgart, Arkansas. The goals were to better understand the reliability and versatility of molecular markers, as well as to increase precision and selection for the best trait combinations. Breeders were shown how the use of genomics tools such as molecular markers could provide improvements over existing breeding techniques, offering savings in time and labor. BIOINFORMATICS The massive amount of

data emerging from NPGI-supported projects has led to new challenges in data management and analysis. The full potential of the current resources has yet to be realized, in part, because of a need for new analysis tools as well as uniform data standards and ontologies to allow comparisons across multiple plants and multiple data types.

At the request of the Interagency Working Group on Plant Genomes, the curators of the major plant databases recently undertook a needs assessment for plant databases. The resulting white paper (http://www.gramene.org/), describes, as a most pressing need, the research community's need for a system of curated data repositories where information is actively acquired, organized, maintained and distributed. The group further points out that in order to meet this need, a trained cadre of skilled knowledge workers who are able to curate complex biological data, as well as a system of stable funding that enables such repositories to be established and maintained for extended periods of time. The white paper detailed a number of recommendations for the NPGI to consider, including the establishment of standardized data formats and user interfaces, and the development and deployment of new technologies explicitly aimed at integrating across speciesspecific databases and clade-oriented databases.

THE UNIVERSITY OF CALIFORNIA, DAVIS PARTNERSHIP FOR PLANT GENOMICS EDUCATION (**PPGE**) PPGE is a consortium of investigators dedicated to integrating plant genomics and biotechnology into secondary school curriculum. In addition to creating and providing educational software and on-line materials, associated hands-on activities, and equipment loan programs to local schools (http:// ppge.ucdavis.edu/), PPGE hosted two workshops in genomics and biotechnology last summer on the University of California, "Genomics: from Mendel to Microchips" Davis campus. included teacher participants from eight states, many of whom serve underrepresented minority students. "Biotechnology: Techniques, Applications, and Implications" was conducted in collaboration with the Southern AgBiotech Consortium for Underserved Communities, a group of 11 Historically Black Colleges and Universities (HBCUs) dedicated to promoting agricultural biotechnology outreach and strengthening pre-college science education. In this effort, PPGE provided the instructors, training facilities, materials and on-going support for the teacher participants from schools serving students of low-income or rural isolation. In FY 2005, PPGE trained and/or distributed information to over 800 teachers, impacting approximately 8,000 students across the US.

## NEW INTEGRATIVE GRADUATE EDUCATION AND RESEARCH TRAINEESHIP PROGRAM (IGERT) PROJECTS IN FY 2005

The NSF IGERT program supports graduate education programs that integrate training and research expertise from across a wide range of disciplines not typically aligned in more traditional graduate training programs. In FY 2005, two new IGERTs were initiated in the area of plant genomics. An IGERT award to the University of California, Riverside will establish ChemGen, a doctoral training program in advanced chemical genomics. The goal of the program is to prepare a cohort of scientists with depth and breadth of understanding across the disciplines of cell biology, chemistry, computational sciences and engineering, with competence in collaborative research aimed at solving complex problems in biology. An IGERT award to the University of California, San Diego will train students in plant systems biology at the interface of computational genomics, systems modeling and plant sciences. This program should provide a unique training environment for graduate students and will position them at the frontier of systems biology to address major challenges facing plant scientists and agricultural biotechnology.

## DEVELOPING COUNTRY COLLABORATIONS IN PLANT GENOME RESEARCH (DCC-PGR) The DCC-

PGR program was started in FY 2004 by NSF to accelerate the application of new tools and resources to agriculture in developing countries. In FY 2005, DCC-PGR awards were made to 13 different projects involving collaborators in eight different countries. Some of the first projects are already bearing fruit. For example, a collaboration between the University of Missouri, Columbia and the Research Laboratory for Agricultural Biotechnology and Biochemistry (RLABB) in Katmandu, Nepal, is focused on extending the project's study of the proteomics of oil seed filling (deposition of stores of proteins and oils) in Brassica species to varieties local to Nepal. The goal is to develop new plants with improved oil and reduced erucic acid content amenable to processing for food and feed. A research visit of RLABB personnel to the University of Missouri led to the development and publication of a simple, cost-effective procedure for detecting phosphoprotein profiles in developing seeds. To the surprise and delight of the project participants, many of the phosphoproteins identified are involved in intermediary metabolism and fatty acid synthesis and are thus likely to serve as potential targets for developing improved Nepalese Brassica varieties.

A MICROARRAY WORKSHOP FOR FACULTY AT HISTOR-ICALLY BLACK COLLEGES **AND UNIVERSITIES** (HBCUs) The Genome Consortium for Active Teaching (GCAT) was established in 1999 to bring functional genomic methods into undergraduate curriculum primarily through student research, particularly at predominantly undergraduate institutions and minority-serving institutions (http:// www.bio.davidson.edu/projects/GCAT/ gcat.html). GCAT held two best practices workshops for faculty members who conduct research with undergraduate students and/or teach undergraduate courses at minority serving institutions on the use of microarrays. The workshops were held in Summer 2005 at Morehouse College, an HBCU. Each workshop trained about

to incorporate their experiences into teaching and research activities at their own institutions.

**GENOMICS FOR BIOFUELS** A December 2005 DOE workshop entitled "Biomass to Biofuels" was jointly sponsored by the Office of Energy Efficiency and Renewable Energy/ Office of Biomass Program and the Office of Science/ Office of Biological and Environmental Research. Researchers from academic, federal, and industrial sectors identified science and technology challenges ranging from feedstock genomics to bioconversion of cellulose that must be met in order to realize the potential for biological conversion of biomass to biofuels.

**RICE SNPS AND THE IRFGC** (*http://www.iris.irri.org/IRFGC/*) Researchers at Colorado State University Ft. Collins, The Institute for Genomic Research, North Carolina State University Raleigh, Perlegen Sciences, and the International Rice Research Institute are participating in an effort led by the International Rice Functional Genomics Consortium (IRFGC) to identify a large fraction of the Single Nucleotide Polymorphisms (SNPs) in cultivated rice. This will provide rice researchers access to most of the sequenced Nipponbare "reference" genome and a public database will be developed to house the SNP and related data. The phenotypes of all rice varieties used in this project will be evaluated and catalogued to facilitate association of traits with the mapped SNP variation. These experiments represent an unprecedented collaborative effort to apply sequence information to discover gene function in a model crop plant, an approach that has been successfully demonstrated in understanding human diseases.

## PLANS FOR THE NEXT YEAR

All agencies participating in the NPGI plan to continue support of plant genome research based on the NPGI plan as appropriate for each agency's mission. Specifically:

NSF will continue to support basic research in plant genomics to accelerate the acquisition and utilization of new knowledge and innovative approaches to elucidating fundamental biological processes in plants. The focus will be on plants of economic importance and plant processes of potential economic value.

The USDA will continue to support projects to increase fundamental knowledge of the structure, function and organization of plant genomes. The focus will be on feedstock genomics, translational genomics to integrate molecular breeding technologies and traditional breeding practice for crop and forestry improvement. In addition, educational and extension activities will focus on informing students, scientists, farmers, processors and consumers in the use of genomics technology for gene discovery and application.

USDA CSREES National Research Initiative (NRI) and DOE Office of Biological and Environmental Research have issued a joint request in FY 2006 for grant proposals focused on the genomics-based research that will lead to the improved use of biomass and plant feedstocks for the production of fuels such as ethanol or renewable chemical feedstocks.

DOE will continue to support genomics-based research that will lead to improved use of biomass and plant feedstocks for the production of fuels such as ethanol or renewable chemical feedstocks. The goal is to leverage information from genome sequencing in combination with analytical technologies and systems biology approaches in order to confidently predict and manipulate the biological function of plants for bioenergy resources.

USAID will continue to support genomic and marker assisted selection projects as part of Harvest Plus to enhance the Vitamin A, iron and zinc content of major crops. In addition, USAID will support training in the use of genomics and marker-assisted selection for scientists in key African crops.

The IWG will continue to coordinate and provide oversight to the NPGI. In 2006 the IWG will begin the long process of gathering community input to begin developing the next 5-year plan.

**About the National Science and Technology Council:** The National Science and Technology Council (NSTC), a cabinet-level Council, is the principal means for the President to coordinate science, space, and technology to coordinate the diverse parts of the Federal research and development enterprise. An important objective of the NSTC is the establishment of clear national goals for Federal science and technology investments in areas ranging from information technologies and health research, to improving transportation systems and strengthening fundamental research. The Council prepares research and development strategies that are coordinated across Federal agencies to form an investment package aimed at accomplishing multiple national goals. Additional information regarding the NSTC can be obtained from the NSTC website, <a href="http://www.ostp.gov/NSTC/html/NSTC\_Home.html">http://www.ostp.gov/NSTC/html/NSTC\_Home.html</a>.

The following individuals and organizations provided images used in this report:

*Cover* (from left to right): Cotton flower, Charles T. Bryson\*, IPM Images; Soybean flower, Glen Hartman, National Soybean Research Laboratory; Corn ear, ISU Photo Service, Iowa State University of Science and Technology, © 1992; *Arabidopsis thaliana* flower, Elliot Meyerowitz, California Institute of Technology; Wheat ear (antithesis stage), Wikipedia; Rice flower, Nori Kurata, National Institute of Genetics (Japan), Oryzabase; Poplar flower, Virginia Tech Forestry Department © 2003 *Inside*: (page 2) Wheat spiklet, Wikipedia; (middle spread, left to right) Poplar flower, Bill Cook, Forestry Images; *Arabidopsis thaliana* flower, Peggy Greb\*; Sunflowers, Edward McCain\*; Foxtail Barley flower, Klen Gray, Oregon State University photo collection; Potato flower, Keith Weller\*; Flowering alfalfa, Keith Weller\* \*USDA/ARS

**Note:** This document does not represent the final determination in an overall Administration budget decision-making process. The programs presented in this report will have to compete for resources against many other high-priority Federal programs. If these programs compete successfully, they will be reflected in future Administration budgets.

URLS FOR MAJOR RESOURCES					
Resource		Location	URL		
Biological Resources	Arabidopsis Biological Resource Center (ABRC)	Ohio State University <i>http://www.biosci.ohio-state.edu</i> ~ <i>plantbio/Facilities/abrc/abrchore</i>			
	Maize Genetic Stock Cooperation	University of Illinois, Urbana- Champaign	http://w3.aces.uiuc.edu/maize-coop/		
	Genetic Stocks – Oryza Collection (GSOR)	Dale Bumpers National Rice Research Center, Stuttgart, AK	http://ars.usda.gov/Main/docs. htm?docid=8318		
DNA Resources	BAC libraries, cDNA libraries	University of Arizona Genomics Institute	http://www.genome.arizona.edu/		
	BAC libraries, cDNA libraries	Clemson University Genomics Institute	http://www.genome.clemson.edu/		
Microarrays/ Chips	Maize oligonucleotide arrays	University of Arizona and the Institute for Genomic Research	http://www.maizearray.org/		
	Rice oligonucleotide arrays	The Institute for Genomic Research	http://www.ricearray.org/		
	Potato oligonucleotide arrays	The Institute for Genomic Research	http://www.tigr.org/tdb/potato/ microarray_comp.shtml		
	Barley, rice, maize, Medicago, soybean, sugarcane, tomato, grape, wheat GeneChips	Affymetrix	http://www.affymetrix.com/products/ application/quantitation.affx		
Information Resources	The Arabidopsis Information Resource (TAIR)	Carnegie Institution at Stanford	http://www.arabidopsis.org/		
	Gramene	Cold Spring Harbor Laboratory	http://www.gramene.org		
	MaizeGDB	Iowa State University	http://www.maizegdb.org		
	International Sequencing Consortium	NIH/NHGRI	http://www.intlgenome.org/		
	Legume Information System (LIS)	National Center for Genome Resources (NCGR)	http://www.comparative-legumes.org/		
	Consensus Legume Database	University of Minnesota	http://www.legumes.org		
	Genome Database for the Rosaceae (GDR)	Clemson University	http://www.mainlab.clemson.edu/gdr/		
	PlantGDB	Iowa State University	http://www.plantgdb.org/		
	Populus Genome Portal	DOE Joint Genome Institute	http://genome.jgi-psf.org/Poptr1/ Poptr1.home.html		
	Solanaceous Genomes Network (SGN)	Cornell University	http://soldb.cit.cornell.edu/		
	Plant Genome Research Outreach Portal (PGROP)	Iowa State University	http://www.plantgdb.org/PGROP/ pgrop.php		
	BarleyBase	Iowa State University	http://www.barleybase.org/		

Acknowledgement The Interagency Working Group on Plant Genomes acknowledges the assistance of Sharlene Weatherwax (Department of Energy), Jane Silverthorne (National Science Foundation), and Ed Kaleikau (U.S. Department of Agriculture) in prepa-ration of this report. Jessica Corman (National Science Foundation) provided assistance in production of the report.