



# NATIONAL PLANT GENOME INITIATIVE

## Progress Report January 2007

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



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

Dear Colleague:

This report is an annual update from the Interagency Working Group (IWG) on Plant Genomes on activities of the National Plant Genome Initiative (NPGI). It includes brief highlights that are illustrative of activities supported and coordinated by the NPGI member agencies.

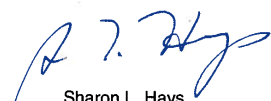
Plant genomics research in the U.S. continues to advance the frontier of plant biology, addressing fundamental biological questions using the tools, resources, and concepts of genomics. Exciting new discoveries over the past year include the possible role of small RNAs in improving hybrid vigor, the identification of the gene responsible for grain protein content, and the elucidation of the molecular mechanism underlying stem rust disease resistance. Publication of the complete sequence of the poplar genome and its analysis has opened up new research areas, including the application of tree genomics to developing new, non-food uses of plants. New discoveries are being harnessed to improve plant characteristics such as cold-tolerant field crops, and to develop new plant-based products such as biofuels. These activities will make critical contributions to building a sustainable, bio-based economy of the future.

U.S. scientists involved in plant genomics research continue to cooperate with their international colleagues, as evidenced by a number of internationally coordinated plant genome sequencing projects. Collaboration extends to developing countries where U.S. scientists work with partners abroad to accomplish common research goals and advance work on plants of particular interest to developing countries.

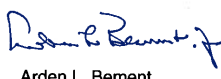
The Interagency Working Group on Plant Genomes is to be commended for its leadership and stewardship of the NPGI. The IWG will continue to coordinate the NPGI to ensure world-class research and the training of a new generation of scientists in plant biology.

Sincerely,

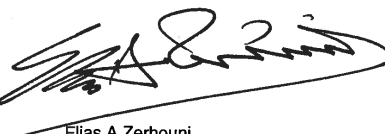
Committee on Science Co-Chairs



Sharon L. Hays  
Associate Director for Science  
Office of Science and Technology Policy



Arden L. Bement  
Director  
National Science Foundation



Elias A. Zerhouni  
Director  
National Institutes of Health

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

Co-Chairs

James P. Collins, National Science Foundation  
Judith St. John, US Department of Agriculture

Members

Robert B. Bertram, US Agency for International Development  
Noah Engelberg, Office of Management and Budget  
George S. Foster, US Forest Service  
Richard Greene, US Department of Energy  
Anna Palmisano, US Department of Agriculture  
Jane L. Peterson, National Human Genome Research Institute  
Jane Silverthorne, Office of Science and Technology Policy  
Sharlene Weatherwax, US Department of Energy

Executive Secretary

Machi F. Dilworth, National Science Foundation

## INTRODUCTION

Plants hold the key to sustainability on earth, as plants along with photosynthetic bacteria are the primary producers on which all life depends, directly or indirectly, for survival. Through photosynthesis using the sun's energy, terrestrial and aquatic plants convert carbon dioxide in the air and water into organic compounds, and store them in forms that other organisms can acquire and utilize through consumption. In addition to food and feed, plants provide oxygen we breathe, fibers for our clothes, materials to build our homes, and serve as sources of medicinal compounds as well as raw materials for industry. In fact, fossil fuels were formed from the remains of plants and animals that lived long ago.

In the US, crops such as corn and soybean, which are grown using highly efficient production methods, serve purposes ranging from food to industrial feedstock. World-wide, we need to increase food production for the growing world population utilizing methods that minimize the negative environmental impacts associated with some agricultural practices. There is also an increasing need for new crop plants with characteristics suitable for emerging applications such as biofuel production.

Work supported through the NPGI leverages scientific expertise and resources from multiple Federal agencies to expand our fundamental knowledge about the genetic makeup of plants, and to facilitate applications of plant biology to improve the environment, to meet our growing food and energy needs, to advance clean industrial processes, and to contribute to overall economic growth. This report includes brief highlights of activities in the past year, illustrating rapid and important advances across the spectrum of NPGI activities. It is not an exhaustive list of ongoing projects being funded, supported and coordinated by the NPGI member agencies.

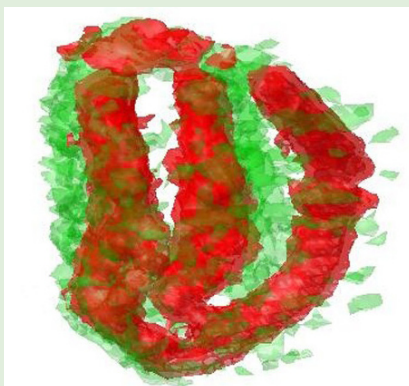
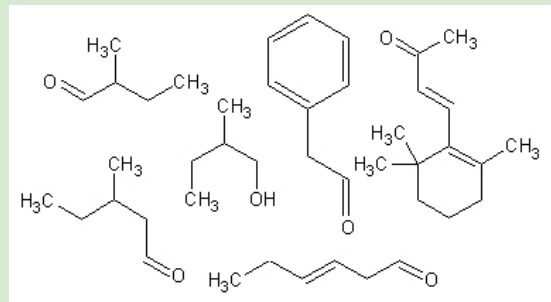
## SCIENTIFIC BREAKTHROUGHS

### RNA: MORE THAN JUST A RECIPE FOR PROTEINS

Plant systems continue to be important for studying processes regulated by small RNAs. The DNA in cells does not exist in long, extended strings but is wound around proteins into a coiled mass called chromatin. It is generally believed that in tightly wound regions of chromatin, the genes are silent, while in the unwound parts, genes can be expressed. However, researchers at Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, recently found that regulatory RNAs are expressed even in the tightly wound parts of chromatin. These RNAs serve to silence the genes from which they come by guiding silencing machinery to them. This mechanism is likely to play a role in regulating gene expression in polyploids, where instead of two copies, there may be as many as six copies of each gene. Polyploidy may play a role in hybrid vigor, and it is also observed in abnormal cells such as cancer cells. The outcomes of this project could have wide-reaching implications for understanding the impact of polyploidy in plants and animals.

### PLANT VOLATILES MAY PROVIDE CUES ABOUT NUTRITIONAL VALUE

Plants produce many volatile metabolites. The flavor volatiles of tomato fruits are derived from essential nutrients and could provide important indicators of the nutritional content of foods. A project led by researchers at the University of Florida, Gainesville, FL and their industrial collaborators resulted in a milestone paper in the past year describing approximately 35 different genes influencing tomato volatile production. This work was based on analysis of data from the growth of 75 plant lines over multiple seasons in multiple locations. These genes can now serve as markers to further study the complex metabolic pathways of flavor and nutrient compounds, and to improve the quality of tomato fruits.



**SEEDS STORE IRON IN MORE WAYS THAN ONE** Iron is an essential element for plants but it can also be toxic at high concentrations. Plants and animals store iron in a protein complex called ferritin as a way of keeping it available in a non-toxic form. However, it has been found that plants can also store iron in the vacuole, a membrane-bound compartment that can be used to store and sequester potentially toxic materials. Recent research at Dartmouth College, Hanover, NH, on the model plant *Arabidopsis* has shown that cells in seeds accumulate iron in their vacuoles using a transporter protein molecule. Visualization of the pattern of iron accumulation was possible because of a highly sensitive method that enabled the researchers to see where iron accumulates. Iron is an important component of the human diet and applying this knowledge to modifying crop plants could allow development of iron-rich seeds, including cereals and beans.



## ROLE OF SUGAR MOLECULES IN SIGNALING

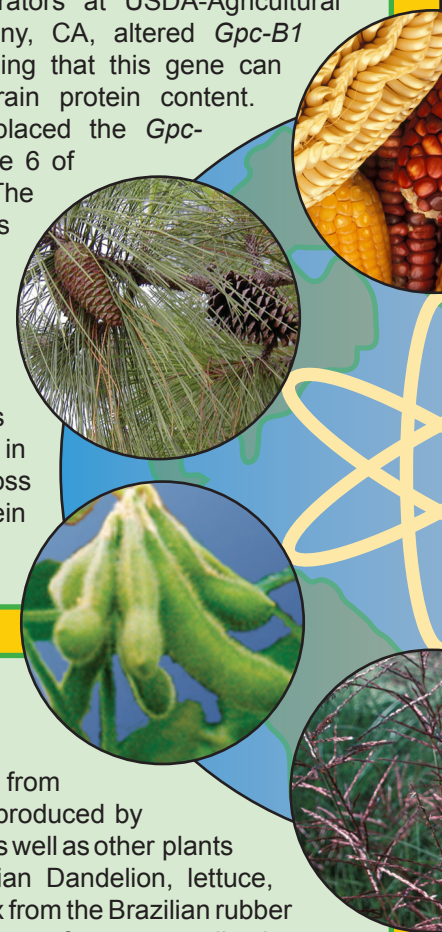
The development of ears and tassels in corn is regulated by a complex network of genes, including the *Ramosa* genes. These genes are involved in determining ear and tassel architecture, affecting the arrangement of kernels in the ear as well as the number of branches on tassels. Researchers at Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, made the surprising discovery that one of these genes, *Ra3*, works by modifying a simple sugar called trehalose. Trehalose appears to act as a signal in plant cells and in this specific case, as a localized short-range signal to regulate the identity of cells that give rise to the branches of ears and tassels. The gene networks in corn have been shown to be similar to those in other cereals and grasses, enabling this discovery to be applied across a wide range of plants.



**STEM RUST RESISTANCE** Stem rust, caused by *Puccinia graminis* f. sp. tritici, has historically been an important disease in barley and wheat in North America. Durable resistance in barley has been achieved for over 60 years by the widespread use of a single resistance gene, *Rpg1*. While the *Rpg1* gene has been cloned, the molecular mechanisms regulating the *Rpg1*-mediated disease resistance remain to be clarified. Researchers at Washington State University, Pullman, WA, have had a breakthrough and identified a gene, *Rpr1*, required for *Rpg1*-dependent resistance to stem rust in barley. The results of this research will facilitate understanding why the *Rpg1* gene has been so durable and allow us to learn how to engineer durable resistance in other disease-resistance genes in barley, wheat, and other crops.



**GRAIN PROTEIN CONTENT** Grain protein content is important for human nutrition as well as the quality of food products derived from cereal grains. Collaboration between researchers at the University of California, Davis, CA and the University of Haifa in Israel has led to the identification of a gene from wild wheat germplasm, called *Gpc-B1*. Collaborators at USDA-Agricultural Research Service, Albany, CA, altered *Gpc-B1* levels in wheat, confirming that this gene can significantly increase grain protein content. Genetic mapping has placed the *Gpc-B1* gene on chromosome 6 of the wheat genome. The DNA sequence from this region has been used to develop a molecular marker that can identify the presence or absence of the *Gpc-B1* gene in different wheat lines. This marker is now being used in breeding programs across the US to improve protein content in wheat.



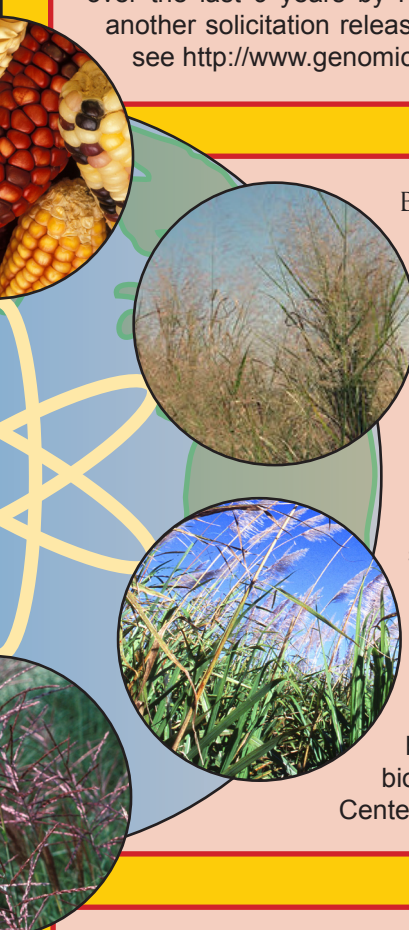
## NATURAL RUBBER – A PLANT PRODUCT

Natural rubber is made from latex, a milky white sap produced by the Brazilian rubber tree as well as other plants such as Guayule, Russian Dandelion, lettuce, and sunflower. While latex from the Brazilian rubber tree is the only suitable source for some applications, such as airplane tires, its use in other applications such as latex gloves can cause severe allergic reactions in some people. Although rubber biosynthesis has been studied for more than a century, biochemical approaches have been unsuccessful in identifying the enzymes involved. A project led by researchers at the University of Nevada, Reno, NV, is using a combination of proteomic and genomic approaches to identify candidate genes for rubber biosynthetic enzymes and then test their functions genetically. The project is currently testing the role of twenty candidate genes in rubber biosynthesis using a transgenic approach in Russian dandelion. This work could allow development of novel rubber-producing crop plants as well the ability to produce large amounts of hypoallergenic latex.



## NEW GENOME-ENABLED EFFORTS IN PLANT-BASED PRODUCT DEVELOPMENT

**A JOINT PROGRAM IN PLANT FEEDSTOCK GENOMICS** In 2006, DOE's Office of Biological and Environmental Research (BER) and the USDA Cooperative State Research, Education, and Extension Service (CSREES) National Research Initiative began a joint competitive Plant Feedstock Genomics grants program. The goal of the joint fundamental research program is to facilitate the use of plant tissues, specifically lignocellulosic materials, for bioenergy or biofuels. In August 2006, the agencies awarded nine research grants supporting genomic research projects on poplar, alfalfa, sorghum, wheat, and other grasses. All awards leveraged information and research tools developed and made available over the last 9 years by NPGI projects. The program continues with another solicitation released in October 2006. (For more information, see <http://www.genomicsgtl.energy.gov/research/DOEUSDA/>.)



**BIOENERGY RESEARCH CENTERS** The DOE BER program issued a Funding Opportunity Announcement to establish and operate two new Bioenergy Research Centers to accelerate basic research on the development of cellulosic ethanol and other biofuels (<http://genomicsgtl.energy.gov/centers>). The purpose of the new Bioenergy Research Centers is to give greater impetus and focus to developing usable knowledge to advance biotechnology-based strategies for biofuel production and could ultimately lead to technologies deployable in the nation's energy economy. The Centers will likely include interdisciplinary research in such areas as plant and microbial genomics, microbial and plant biology, biochemistry, structural and computational biology, bioinformatics, and engineering. The Centers will be selected in the summer of 2007.

**BIOMASS GENOMICS** Two recent NSF awards complement the DOE and USDA efforts in this area. One project, led by researchers at the University of California, Davis, CA, in collaboration with the US Forest Service, will develop genomic markers for members of the *Pinaceae* (conifers). Several thousand Single Nucleotide Polymorphisms (SNPs) will be developed for each of eight different conifer trees to enable forest tree researchers to conduct breeding and gene resource programs at a depth and precision not previously possible. This work will provide new information on the conifer genome in general, and genes involved in wood qualities in particular. The work is likely to have a broad impact on development of new wood-based materials. Another project, led by researchers at the University of Delaware, Newark, DE, will focus on sequencing all of the small RNAs in a range of plant species, including poplar, switchgrass, and pine. As these RNAs are known to regulate key developmental pathways, this basic research will likely lead to identification of the gene networks governing biomass formation and accumulation in the plant.

**WORKSHOP FOR FIFTH- TO TWELFTH-GRADE TEACHERS OF SCIENCE AND AGRICULTURE** Advances in biotechnology are rapidly changing the way plants are developed for use in modern agriculture. A workshop was held in June 2006 at the Arkansas Agricultural Extension Office and Laboratory in Lonoke, AR that provided teachers with the knowledge and technical tools they need to pass on an understanding of technology to students. Discussion and demonstrations connected to science education national standards addressed the following topics: the role of DNA as the genetic code, DNA isolation using common household items, analysis of DNA via gel electrophoresis, the power of mutation as a tool for plant breeders, computer-based activities for DNA analysis, and plant development and biotechnology.



**TEACHER TRAINING** It has long been recognized that a new breed of plant scientist is required to manage and analyze the massive data sets being generated by plant genomics efforts. Research training of K-12 teachers and their students is preparing them to meet this challenge. For example, the Teacher Internships in Plant Genomics program at the University of Arizona, Tucson, AZ, recruits teachers from school districts that serve primarily Hispanic and/or Native American students in Arizona to participate in cutting edge genomics research in a six- to eight-week summer program. The University of Missouri Plant Genomics Research Experiences for Teachers (MUPGRET) program provides research training in plant biology with a special emphasis on plant genomics. In the past four years, MUPGRET has trained over 70 teachers, impacting an estimated 30,000 students from both metropolitan and rural areas in the state of Missouri.

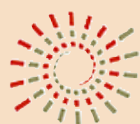


**INTERNATIONAL TRAINING PARTNERSHIP IN RICE BREEDING** A new project led by researchers at Cornell University, Ithaca, NY builds on the observation that rice plant populations have genetically distinct subgroups. The project's goal is to understand if the use of different combinations of genetic subgroups in breeding leads to "hybrid vigor" in such key traits as grain yield and grain quality. As part of the project, the investigators have teamed up with the International Rice Research Institute (IRRI) in Los Banos, the Philippines and developed a training program at IRRI. Beginning in May 2007, the three-week program will be offered for 25 students from around the world each year. Activities will combine hands-on experience in paddy preparation and rice cultivation with classroom training in rice biology, genomics and informatics. The long-term goal is for US universities to develop partnerships with IRRI and other Consultative Group on International Agricultural Research (CGIAR) Centers to allow two-way research and training activities.



**BROADENING PARTICIPATION** Iowa State University, Ames, IA and the USDA-Agricultural Research Service's North Central Regional Plant Introduction Station in Ames, IA, held the first offering of the "Plant Genome Research Outreach Program for Native Americans/American Indians", a research training program for Native American undergraduate students. The focus of this eight-week summer internship program is to investigate the diversity of Native American plant species of cultural, historical, and present-day importance in which the students gain experience in plant genetics, molecular biology, and bioinformatics. Another program led by an interdisciplinary team of scientists from the University of Wyoming, Laramie, WY, the Institute for Genomic Research (TIGR), Rockville, MD, and several other collaborating institutions is designed to enhance educational and research training opportunities in biology for both faculty members and students at Little Big Horn College in Crow Agency, MT. Activities include a combination of scientific exchange visits involving project leaders, tribal college faculty and students with courses and hands-on workshops held at the participating institutions.

INTERNATIONAL COLLABORATION



**eusol**

THE SOL GENOMICS NETWORK GROWS

The *Solanaceae* includes such important crop plants as tomato, potato, eggplant and pepper. The International Solanaceae Genomics Project (SOL) was established in 1993 by researchers from the US, Europe, Australia, Asia, and South America as a 10-year program to develop a comparative research framework for studying diversity and adaptation across the family. An international effort involving scientists from ten countries to sequence the tomato genome is coordinated through the SOL network. In addition, the EU-SOL project has now been funded through the European Commission 6th Framework Programme to develop high quality tomato and potato varieties by mining the rich biodiversity present in the *Solanaceae* for new breeding efforts in the public and private sectors.



**COMPLETION OF THE POPLAR GENOME SEQUENCE** Poplar has emerged as a leading model for tree genomics because of its compact genome and rapid growth. The whole genome sequence of poplar, published in September 2006, was the result of a four-year effort led by the DOE Joint Genome Institute, Walnut Creek, CA, involving 34 institutions around the world. More than 45,000 genes were identified in the 480 Mb sequence. Ninety-three (93) of them are predicted to be associated with the production of cellulose, hemicellulose, and lignin, which are major components of plant cell walls. The sequence can be accessed at [http://genome.jgi-psf.org/Poptr1\\_1/Poptr1\\_1.home.html](http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html). Poplar is the first tree genome to be sequenced and the sequence could contribute to development of trees ideally suited for feedstock for a new generation of biofuels such as cellulosic ethanol.

## PLANS FOR THE NEXT YEAR

All agencies participating in the NPGI will continue support of plant genome research based on the NPGI plan as appropriate for each agency's mission. Below are brief descriptions of areas of special emphasis identified by each participating agency for the coming year:

**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. In addition, a new competition will be held in the coming year to establish a Plant Science Cyberinfrastructure Collaborative. The goal of this program is to create an organization that will enable new conceptual advances through integrative, computational thinking. The Collaborative will provide tools for the broad community of scientists to collect, analyze, integrate, and synthesize all kinds of data that are stored and being generated world-wide, and convert the data to knowledge.

**DOE** will continue to support fundamental plant research in genomics and biotechnology that support its missions in bioenergy, environmental remediation, and carbon cycling. The DOE Joint Genome Institute (JGI) is evolving to reflect a greater emphasis on fundamental genome research involving plants. Some of this emphasis is on providing discovery-enabling resources, such as genome sequencing at the JGI. Recent bioenergy-relevant plants selected for whole genome sequencing include soybean, sorghum and *Brachypodium* (a model plant for perennials such as switchgrass); additional smaller pilot sequencing studies will focus on cotton, switchgrass, and cassava. The Genomics:GTL (Genomes to Life) program will complement this available and emerging genomic information with a solicitation requesting applications for new analytical and imaging technologies for lignocellulosic material degradation, and for multiplexed screening for plant phenotypes.

**USDA** will continue to support research to increase fundamental knowledge of plant genome structure, organization, and function, and to develop tools, genetic resources and bioinformatics capacity for genome-wide analysis and discovery. One focus area will be plant feedstock genomics relevant to bioenergy, through development of more efficient and effective methods for mining plant germplasm collections for valuable traits, and advancing translational genomics research to integrate education and extension opportunities with molecular breeding technologies and traditional breeding practices for crop and forestry improvement. Genomic analytical approaches will be extended further to specialty crops, many of which are horticultural, to help accelerate their improvement.

**US Forest Service Research and Development (FS R&D)** will continue to support genomics-based research that will lead to improved use of biomass and plant feedstocks for producing ethanol, renewable chemical feedstocks, and fossil fuel substitutes; high-value wood products; restoration feedstocks; and for enhancing forest health and productivity. Genomics research will be coupled with plant breeding to optimize progress.

**NIH:** It should be noted that while NIH does not support plant genomics research directly, the plant genomics research community has continued to benefit from NIH's contributions to advances in genomics technologies.

As in the past years, the above activities will be coordinated by the IWG. In the coming year, the community input will be sought to assess the impact of NPGI on plant science, research infrastructure, and training/education. The IWG plans to host a symposium to celebrate the tenth anniversary of the NPGI in 2007.



## ABOUT THE NATIONAL PLANT GENOME INITIATIVE

The National Plant Genome Initiative (NPGI) was formally established in 1998 as a coordinated approach to advance plant science and its applications to address issues of national interest. During its nine year history, the NPGI has developed two five-year plans; this report highlights progress during the third year of work under the second five-year plan ("National Plant Genome Initiative: 2003-2008" (<http://www.ostp.gov/NSTC/html/npgi2003/index.htm>)). The program is coordinated by the Interagency Working Group on Plant Genomes (IWG) which is convened under the auspices of the National Science and Technology Council's Committee on Science. Currently participating agencies include the National Science Foundation (NSF), Department of Agriculture (USDA), Department of Energy (DOE), National Institutes of Health (NIH), Agency for International Development (USAID), Forest Service (USFS), the Office of Science and Technology Policy (OSTP), and the Office of Management and Budget (OMB). Each agency participates in the NPGI in a manner consistent with its specific mission.

## STATUS OF ON-GOING PLANT GENOME SEQUENCING PROJECTS

Plant Species	Genome Size	Products Public Access to Sequence	Project Status	PIs, Institutions, Countries
Barrel Medic ( <i>Medicago truncatula</i> )	550 Mb	Gene-rich BAC sequence in GenBank	In progress, anticipated completion in 2008	Medicago Genome Sequencing Project
		<a href="http://www.medicago.org">http://www.medicago.org</a>		
		<a href="http://medicago.toulouse.inra.fr">http://medicago.toulouse.inra.fr</a>		
Tomato ( <i>Solanum lycopersicum</i> )	950 Mb	Gene-rich BAC sequence in GenBank	In progress	International Tomato Genome Sequencing Project
		<a href="http://www.sgn.cornell.edu/help/about/tomato_sequencing.pl">http://www.sgn.cornell.edu/help/about/tomato_sequencing.pl</a>		
<i>Sorghum bicolor</i>	736 Mb	Whole genome shotgun sequence	In progress	DOE Joint Genome Institute
		<a href="http://www.jgi.doe.gov/sequencing/why/CSP2006/sorghum.html">http://www.jgi.doe.gov/sequencing/why/CSP2006/sorghum.html</a>		
Grape ( <i>Vitis vinifera</i> )	500 Mb	Whole genome shotgun sequence	In progress	INRA (France) IASMA (Italy)
		<a href="http://www.vitaceae.org/molecular.html">http://www.vitaceae.org/molecular.html</a>		
Maize ( <i>Zea mays</i> )	2600 Mb	BAC and whole genome shotgun sequences deposited in GenBank	In progress, anticipated completion in 2009	Maize Genome Sequencing Project
		<a href="http://www.gramene.org/Zea_mays/index.html">http://www.gramene.org/Zea_mays/index.html</a>		
Soybean ( <i>Glycine max</i> )	1115 Mb	BAC and whole genome shotgun sequences deposited in GenBank	In progress	DOE Joint Genome Institute
Potato ( <i>Solanum tuberosum</i> )	840 Mb	Gene-rich BAC sequence in GenBank	In progress	The Potato Genome Sequencing Consortium
		<a href="http://www.potatogenome.net/">http://www.potatogenome.net/</a>		
<i>Brachypodium distachyon</i>	355 Mb	Whole genome shotgun sequence	In progress	DOE Joint Genome Institute
		<a href="http://www.jgi.doe.gov/sequencing/why/CSP2007/brachypodium.html">http://www.jgi.doe.gov/sequencing/why/CSP2007/brachypodium.html</a>		

**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, in turn, coordinate the diverse parts of the Federal research and development enterprise. An important part 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 (<http://www.ostp.gov/nstc/>).

**Image Credits** The following individuals and organizations provided images used in this report: *Cover*, plants of on-going genome sequencing projects (from left to right): *Brachypodium distachyon*, Dr. Michael Bevan; Potatoes, USDA; Soybean, Dr. Henry Nguyen; Sorghum, Wikipedia; Maize, USDA; Tomato, Dr. James Giovannoni; *Medicago truncatula*, Dr. Douglas Cook; Grapes, USDA. *Inside* (from left to right): Iron/manganese storage in seeds, Dartmouth College; The ramosa3 mutant of maize, Dr. David Jackson; Stem rust, USDA; Wheat grains, USDA; Teacher Internship in Plant Genomics, University of Arizona; Poplar, USDA; Plant Genome Outreach to Native Americans student participants, Dr. Carolyn Lawrence. *Inside* (center wheel, clockwise from top): Maize, USDA; Switchgrass, NREL; Sugarcane, USDA; Miscanthus, John Mosesso, NBI; Soybean, Dr. Mark Tucker, USDA; Loblolly pine, Dr. David Byers.

**Acknowledgements:** The Interagency Working Group on Plant Genomes acknowledges the assistance of Sharlene Weatherwax (Department of Energy), Jane Silverthorne (National Science Foundation), and Ed Kaleikau (US Department of Agriculture) in preparation of this report. Lauren Kitchen (National Science Foundation) provided assistance in the production of this report.