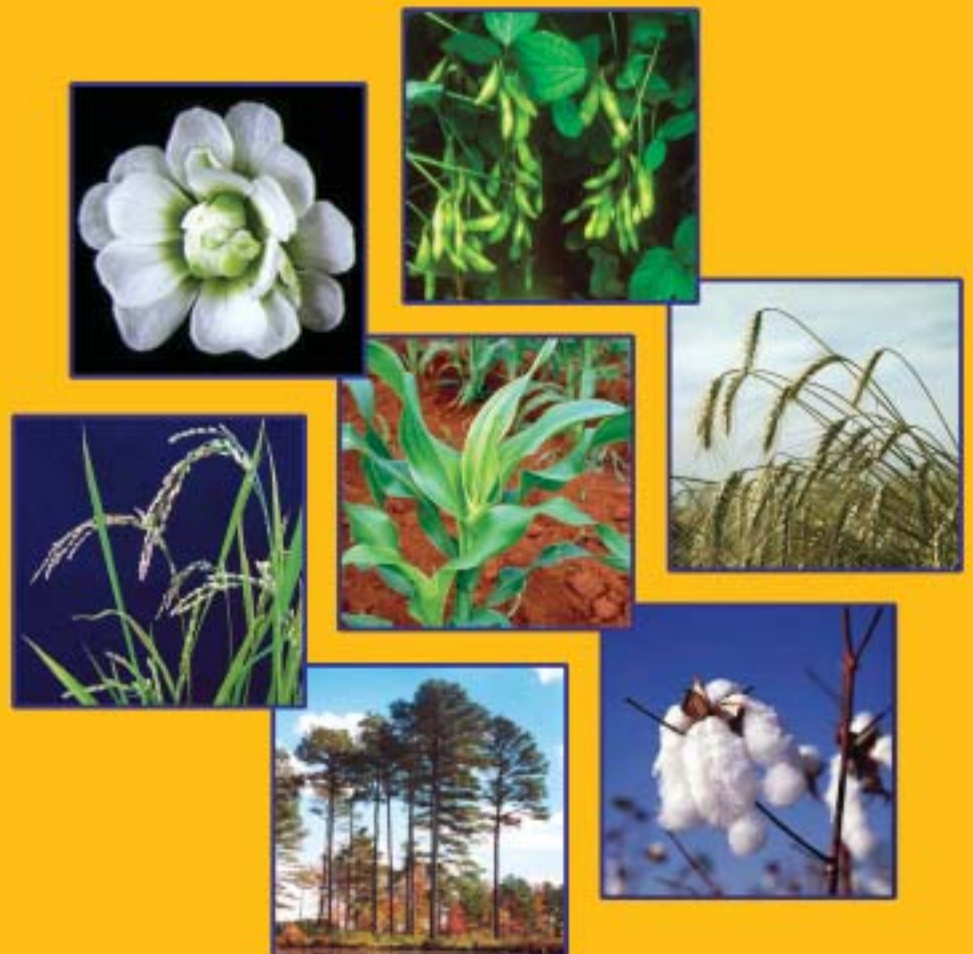




National Plant Genome Initiative: 2003 – 2008

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

January 2003



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 policies across the Federal Government. NSTC acts as a “virtual” agency for science 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. This council prepares research and development strategies that are coordinated across Federal agencies to form an investment package that is aimed at accomplishing multiple national goals.

To obtain additional information regarding the NSTC, contact the NSTC Executive Secretariat at (202) 456-6101.

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.

Cover photos (clockwise from top): soybean - courtesy of Dr. Henry Nguyen, University of Missouri; wheat - courtesy of the Agricultural Research Service, USDA; cotton - courtesy of the Agricultural Research Service, USDA; loblolly pine - courtesy of Chesapeake Bay Program; rice - courtesy of Dr. Susan McCouch, Cornell Genomics Initiative; *Arabidopsis thaliana* - courtesy of Dr. Jose Luis Reichmann and Dr. Elliot Meyerowitz, California Institute of Technology; (middle): corn - courtesy of Dr. M.G. Neuffer and MaizeDB, University of Missouri. Cover design by James Caras and Courtney Snyder of the National Science Foundation.

National Plant Genome Initiative: 2003 - 2008

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

January 2003



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

January 17, 2003

Dear Colleague:

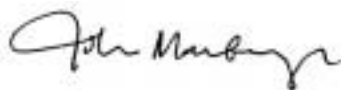
The enclosed report provides a five year (2003-2008) plan for the National Plant Genome Initiative (NPGI). Implementation of this plan will build on significant advances made in the first five years under the NPGI. The Initiative will advance the frontiers of plant science and keep the U.S. at the forefront of research in this field.

The National Science and Technology Council (NSTC) believes it is critical to continue and even accelerate research efforts in plant genomics in order to take advantage of exciting scientific opportunities that will lead to improved agriculture, energy and health, thus ensuring a high quality of life for future generations. The scientific community has established extensive interdisciplinary and international collaborations to achieve the goals of the NPGI. It is vital to maintain their scientific momentum.

The NSTC Committee on Science's Interagency Working Group on Plant Genomes received input from many sectors of the scientific community, including the National Academy of Sciences, industry, professional societies, and producer/grower organizations, in developing this report. I thank them for their contributions and look forward to the continued success of the NPGI.

Enclosure

Sincerely,



John H. Marburger, III
Director

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

Co-Chairs

Mary E. Clutter
Assistant Director
Biological Sciences
National Science Foundation

Joseph Jen
Under Secretary
Research, Education, and Economics
U.S. Department of Agriculture

Members

Gregory L. Dilworth
Section Chief
Office of Basic Energy Sciences
U.S. Department of Energy

Noah Engelberg
Resource Management Officer
Office of Management and Budget

Clifford Gabriel
Deputy to the Associate Director
Office of Science and Technology Policy

Jane L. Peterson
Associate Director
Division of Extramural Research
National Human Genome Research Institute
National Institutes of Health

Anne Vidaver
Chief Scientist
Cooperative State, Research,
Education and Extension Service
U.S. Department of Agriculture

Judy St. John
Associate Deputy Administrator
Agricultural Research Service
U.S. Department of Agriculture

Table of Contents

Executive Summary	1
Introduction	3
Major Accomplishments for 1998-2002	5
Plan for 2003-2008	6
Cost Estimates for Achieving Objectives	14
Appendix	15

Executive Summary

The National Plant Genome Initiative (NPGI) was established in 1998 as a coordinated national plant genome research program by the Interagency Working Group (IWG) on Plant Genomes with representatives from the Department of Agriculture (USDA), Department of Energy (DOE), National Institutes of Health (NIH), National Science Foundation (NSF), Office of Science and Technology Policy (OSTP), and the Office of Management and Budget (OMB). Since 1998, the field of plant genomics has made tremendous strides. It has changed the way research is conducted in plant biology; it has attracted a new generation of scientists into the field; and it has contributed new information and knowledge to science. Through development of plant genomic resources, the NPGI has built a foundation on which the scientific community can advance research, not only in plant genomics but also in diverse disciplines ranging from fundamental biological sciences to biotechnology.

In this report, the IWG describes the NPGI plan for next five years (2003 – 2008). The IWG solicited and received input from many sources, which were used as the basis for this plan.

NPGI Goal Statement

The ultimate goal of the NPGI is to understand the structure and function of all plant genes at levels from the molecular to the organismal and to interactions within ecosystems. The new knowledge and insights gained from plant genomics will lead to unexpected discoveries and conceptual advances in our understanding of the biology of plants. With a focus on plants of economic importance and plant processes of potential economic value, the NPGI will impact applied research related to agriculture, natural resources, the environment, health, and plant-based industries.

Guiding Principles for NPGI, 2003-2008

The NPGI will follow the same guiding principles as in the original five-year plan.

- The Initiative should be viewed as a long-term project, governed by a periodically updated plan based on scientific progress and with stakeholder input as a basis for setting goals and priorities.
- All research resources, including data, software, germplasm, and other biological materials and research tools should be openly accessible to all in a timely manner.
- The Federal portion of the Initiative should be coordinated by an interagency working group.
- All awards should be made based on scientific merit and rigorous, competitive peer review.
- Partnerships with the private sector and the international community are vital for success.

New Objectives for 2003-2008

Objectives for the next five years of the NPGI will build on recent scientific and technical advances to ensure continued advancement in plant genomics specifically and plant sciences in general. They include:

- Continued Elucidation of Genome Structure and Organization
- Functional Genomics – Understanding the Biological Role of Plant Genomic Sequence, Including Gene Sequences, Regulatory Sequences, and Repeated Sequences
- Translational Plant Genomics – Application of Genomics Tools
- Bioinformatics in Every Plant Scientist’s Research Tool Box
- Education, Training and Outreach
- Consideration of Broader Impacts

In order to capitalize fully on the investments made to date and to enable further advances in plant genome research, the IWG recommends continued and increased investment in the NPGI by all the participating agencies. The funds will continue to be expended on a competitive basis using rigorous peer review. As in the first five years of the NPGI, the IWG will monitor progress and report major accomplishments annually.

I. Introduction

The Interagency Working Group (IWG) for Plant Genomes was appointed in May 1997 by the Executive Office of the President in response to a request from the Senate VA, HUD and Independent Agencies Appropriations Subcommittee. The IWG consists of representatives from the Department of Agriculture (USDA), National Science Foundation (NSF), National Institutes of Health (NIH), Department of Energy (DOE), Office of Science and Technology Policy (OSTP), and the Office of Management and Budget (OMB). The IWG was charged with identifying science-based priorities for a national plant genome initiative and to plan and coordinate plant genome research activities for the Nation.

In January 1998, the IWG published the National Plant Genome Initiative (NPGI) five-year plan (<http://www.ostp.gov/NSTC/html/npgireport.html>). The plan recommended increased Federal investment focused on the following plant genome research objectives:

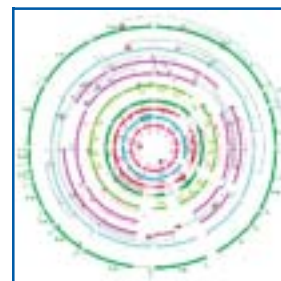
- Completing the sequencing of the model plant species, *Arabidopsis*;
- Participating in an international effort to sequence rice;
- Developing the biological tools to study complex plant genomes such as corn, wheat, soybean, and cotton;
- Increasing our knowledge of gene structure and function of important plant processes;
- Developing the appropriate data handling and analysis capabilities; and
- Ensuring this new information will be accessible to the broader community of plant biologists and maximizing the training opportunities that will arise from the Initiative.

In the original report, the IWG established the following guiding principles for implementing the NPGI five-year plan:

- The Initiative should be viewed as a long-term project, governed by a plan that will be updated periodically.
- All resources, including data, software, germplasm, and other biological materials should be openly accessible to all.
- The Federally funded components of the Initiative should be coordinated by an NSTC interagency working group with USDA, NSF, DOE, NIH, OSTP, and OMB representation.
- All awards should be made on a competitive basis with peer review.
- Partnerships with the private sector and with other nations are vital for success.

The accomplishments over the last five years far exceed the original goals set by the IWG in 1998 for the NPGI. In part, this is due to an increased investment in plant genome research, which has produced the necessary genomics research tools and biological resources for the research community to make rapid advances. In addition, as predicted in 1998 by the IWG, the cost of key technologies such as sequencing and high throughput experimental methods, as well as computer hardware/software and data storage, has decreased steadily. As a result, individual investigators and teams of investigators from a broad spectrum of institutional settings have been able to participate in plant genome research, thus contributing to accelerated advances and enabling the US to be at the forefront of plant genomics.

Success of the last five years has been due in part to strict adherence to principles governing the administration of the NPGI investment in research.



Projects were selected on a competitive basis with rigorous peer review. A requirement of funding was that all outcomes of NPGI-coordinated research were to be freely shared in a timely manner. Significant investments by NIH and other agencies in human genome research have also contributed significantly to the success of the NPGI. These investments have supported the development of new research technologies and tools in genomics, which are generally applicable to all genomics research.

In this report, the IWG documents significant NPGI achievements since 1998 and articulates a new plan for the next five years (2003 – 2008). The IWG solicited and received input from many sources in the development of this plan. Recommendations from the National Academy of Sciences and a variety of stakeholder workshops including scientists, growers, producers and the public were used to identify future scientific opportunities and challenges for the NPGI (See Appendix below). The new plan will be used by each of the participating agencies to establish priorities and initiate new activities consistent with the mission of each agency and with the overall goals of the NPGI.



II. Major Accomplishments for 1998-2002

Major accomplishments have been made in all areas of the original NPGI research objectives. Particularly notable examples are described below:

- Completion of the ***Arabidopsis* genome sequence** four years ahead of schedule. This is the most complete eukaryotic genome sequence finished to date and it has formed the basis for worldwide efforts to identify the function of all of the genes in *Arabidopsis* by the year 2010.
- Completion of a deep draft **rice genome sequence** six years ahead of schedule. The sequence information is being used by the research community to study the genomes of other cereals, including maize, wheat, barley and sorghum.
- New **fundamental science discoveries** including: (1) the structure and organization of centromeres in higher plants, which may lead to new vectors for genetic manipulation of large segments of plant genomes; (2) identification of most of the genes involved in the response of plants to environmental stresses; (3) the discovery of the first active transposon in rice and the first active miniature inverted-repeat transposable elements (MITEs) in eukaryotic organisms; (4) new insights into the mechanism of epigenetic gene silencing, which will impact methods for genetic engineering of plants; and (5) new findings about plant pathways and genes involved in detecting and resisting pathogens.
- Production of **plant genome research resources** including: (1) a large collection of plant ESTs (Expressed Sequence Tags) in GenBank, which number over 2,000,000 in 2002 compared to 50,000 in 1998, providing a valuable resource used widely by individual investigators to study gene structure and function; (2) Bacterial Artificial Chromosome (BAC) libraries for over 72 plant species available to the public, which are being used to identify and clone genes of interest; (3) a large, public collection of transposon-tagged lines for reverse genetics approaches to studying gene function in maize and *Arabidopsis*; (4) deep physical maps of maize, soybean, wheat and other plant species; and (5) various public plant genomic databases available for community use.
- Development of **plant genome research tools** such as: (1) gene expression profiling tools, including a whole-genome array for *Arabidopsis*; (2) informatics tools to access, analyze and synthesize all levels of plant genome data; (3) new optical mapping methods; (4) development of a publicly available efficient maize transformation system; (5) collection of public vectors for gene silencing, which are being used to study gene function in tomato, potato, and tobacco; and (6) a novel technology called "TILLING" for rapid selection of point mutations in any gene, which is being widely used by researchers in the US and Europe studying a range of plant species.
- **International collaborations** have been established to pursue coordinated international efforts to advance genomics of various plant species, including: (1) the Multinational Coordinated *Arabidopsis thaliana* Functional Genomics Project; (2) the International Rice Genome Sequencing Project; (3) the Cereal Genome Initiative; (4) the International Genome Research Organization for Wheat; (5) International Tomato Genome Sequencing Community; (6) the *Medicago truncatula* Genome Group; (7) the Poplar Functional Genomics Consortium; and (8) the Global Musa Genomic Consortium.



III. Plan for 2003 - 2008

NPGI Goal Statement

The ultimate goal of the NPGI is to understand the structure and function of plant genomic sequence information at all levels from the molecular to the organismal and to interactions within ecosystems. The new knowledge and insights gained from plant genomics will lead to unexpected discoveries and conceptual advances in our understanding of the biology of plants. With a focus on plants of economic importance and plant processes of potential economic value, the NPGI expects that research in plant genomics will impact applied research related to agriculture, natural resources, the environment, health, and other plant-based industries.

New Objectives for 2003-2008

The first five years of the NPGI focused heavily on building resources and tools that enable a broad community of scientists to participate in plant genome research. These tools have now opened up exciting opportunities to address many of the important unanswered questions in plant science in a way that was not possible previously. While progress has accelerated, new research needs have also been identified. It is essential that high throughput and cost-effective technologies continue to be developed. Additionally, the vast quantity of plant genome data must be managed effectively so that it is accessible to all researchers. Furthermore, new generations of scientists must be trained in the effective use and insightful interpretation of genomics approaches and information. The objectives for the next five years of the NPGI, described below, will build on recent scientific and technical advances to ensure continued advancement in plant genomics specifically and plant sciences in general.

Objective 1. Continued Elucidation of Genome Structure and Organization

In the past five years, important guiding concepts have emerged from sequencing projects involving plants, animals and microbes. One concept has been the importance of focusing on “reference genomes”. The finished yeast, worm and *Arabidopsis* genomes represent references that serve as assembly templates for subsequent draft sequences of genomes from related organisms. Another important concept is the use of comparative sequence analysis. While a single genome sequence is useful, its utility increases dramatically as additional sequences become available. Comparative genomics uses these resources to extensively mine for all of the genes and regulatory sequences in a genome. For example, the sequencing of two mammalian genomes (human and mouse) has revealed new genes that were not found by sequencing one genome. Comparative genomics research will increase our understanding, at a sequence level, of the events that gave rise to new species or to the emergence of specific traits. The sequence resources

developed over the next five years can be used to describe the structure of individual genomes and also to clarify the dynamic processes that shape genomes.



Contribute to the international effort to finish the rice genome sequence

A deep and highly accurate draft rice sequence was completed in December 2002, representing the first publicly available and most complete rice sequence information to date. However, additional sequencing to close remaining gaps will be required to finish this genome and facilitate its use as a reference

genome for cereals. A complete, finished rice sequence will be an essential tool for the broader plant research community, both basic and applied.

Complete sequencing of the gene-rich regions of the maize genome

Like many of the cereals, maize has a large, complex genome, consisting of about 2,800 million base pairs (Mbp) of DNA, about the same size as the human genome and 21 times larger than the *Arabidopsis* genome. The maize genome organization is complex with more than 80% of the genomic DNA consisting of repetitive sequences and only about 15%, or about 300 Mbp, encoding genes. While it was not realistic to contemplate sequencing a genome of this size and complexity in 1998, it is now possible to sequence the gene-rich regions of the maize genome. Technical challenges are being surmounted by developing efficient methods to enrich for genes prior to sequencing and then assembling and mapping the sequences onto the existing master maize genome map. The technologies being developed for sequencing the maize genome can then be applied to sequence large and complex genomes, not just plant genomes. A complete sequence of the gene-rich regions of the maize genome would augment available genomic tools to address fundamental questions about gene function, evolution, development and physiology across all the cereals.

Detailed genome analysis of a few key plant species

At the present time, it is still prohibitively expensive to sequence all plant genomes since many are large and complex (Table 1). The current cost is approximately \$0.09 per base pair. At this price, a finished (99.99% accuracy) sequence for wheat would cost \$1.44 billion. The National Human Genome Research Institute intends to develop sequencing technology in the next decade that will produce complete genome sequences for \$1,000 each. Until then, the most efficient use of NPGI resources will be to develop a set of draft sequences for the gene-rich regions of key plant species, building on the concept derived from the first five years that reference genomes are essential genomics tools.

Criteria for selection of plant species for sequencing will minimally include the following considerations: (1) Experimental tractability; (2) Complexity of genome structure; (3) Potential for serving as a reference; and (4) Usefulness of the sequence information to advance plant science.

Genome analysis resources for a broad spectrum of plants of biological and economic importance

The majority of plants will not be candidates for detailed genome analysis in the next five years. In these cases, research needs can be met by the development of deep genetic and physical maps, Expressed Sequence Tags (ESTs) and Bacterial Artificial Chromosome (BAC) libraries. BAC libraries are relatively inexpensive to construct and are useful to many researchers who work on unique plant systems and to all researchers for comparative genomics research. A recently developed process called "Targeted Comparative Sequencing" uses BAC libraries as a promising tool to provide insight into genome evolution. ESTs prepared for unique cell types or plants grown under specific conditions are especially useful to identify networks of genes involved in specialized plant processes such as production of secondary metabolites or responses to specific stimuli.

Table 1
Size of sample plant genomes

Plant genome	Estimated size (M base pair)
Arabidopsis	130
Rice	430
Medicago	550
Poplar	550
Apple	770
Tomato	950
Sorghum	1,000
Soybean	1,000-2,000
Cotton	2,110
Maize	2,500-3,000
Barley	5,000
Wheat	16,000
Onion	18,000
Fern	160,000

Understanding the structural basis for plant genome organization

Plants are well suited for studying the structural basis of complex genome organization. Genome organization contains a record of the evolutionary history of the plant. Thus, comparison of select examples can reveal the processes that led to the current structure and organization of plant genomes. In the next few years, additional genome sequences, EST sequences, and other structural genomics resources will become available. These resources will make it possible to generate detailed, comparative maps for finding all genes and regulatory sequences, and studying genome evolution across a broad range of plants. Comparative studies will increase our understanding of the relationships between genome structure and organization and allow us to begin to ask major unanswered questions in plant sciences, such as:

Impact of domestication on genome structure and vice versa: Plant genomes, especially those of cultivated plants, are often radically different from other eukaryotic genomes, both in structure and in organization. It is likely that many of these differences reflect the strong selection applied during domestication over thousands of years. During domestication, whole genome duplication, segmental genome duplication or loss, and genome rearrangements have occurred in a number of crop plants. Understanding the basic biology of the domestication process will help researchers develop rational strategies for future crop improvement.

Role of subgenomes in allopolyploids: Many plants are hybrids of two or more progenitor plants, called “allopolyploids”. For example, bread wheat (*Triticum aestivum*) contains three ancestral genomes termed A, B and D. The D genome, is derived from *Aegilops tauschii*, and contains genes for bread quality. Having sequence information of these genomes would provide scientists with the tools to understand how diverse genomes combine to generate new plant species.

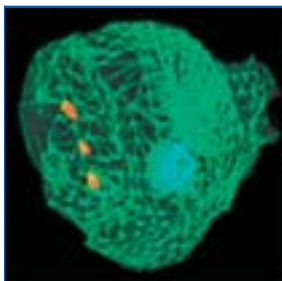
Objective 2. Functional Genomics – Understanding the Biological Role of Genomic Sequences

With the completion of sequencing of the *Arabidopsis* and rice genomes, focus has shifted from the acquisition of primary sequence information in those plants towards understanding the biological role of genomic sequence, including coding, regulatory and repeated sequences. This is particularly important across species boundaries, where genes of similar sequence may serve unique or different functions. With a large percentage of a sequenced genome devoted to hypothetical or unknown gene function, new tools will be required to develop additional strategies for linking sequences to the biological functions that may well be unique to plants.

Complete the Arabidopsis functional genomics project as a Rosetta stone for all plants

The allocation of resources to assign a function to the more than 26,000 predicted *Arabidopsis* genes by the year 2010 is an important first step towards developing a reference for other plants for which there is limited genomic information. This multinational coordinated project implements a whole-system approach to genomics.

Develop a rice functional genomics project as a reference species for cereals



A multinational coordinated functional genomics project should be initiated for rice following the successful model established for *Arabidopsis*. Rice was selected as a reference species for monocots due to its small genome size (Table 1). With the cooperation of government, academic, and industrial scientists, the accurate draft sequence was completed in December 2002. The stage is now set to exploit the functional genomics tools for rice that were developed for *Arabidopsis*. Preliminary comparative analyses suggest the rice genome contains almost twice the number of genes found in *Arabidopsis*, many with

unassigned function. Consequently, novel technologies should be developed in order to assign function to the unique genes in rice that have no apparent biological counterpart in *Arabidopsis*.

Establish repositories for plant genome research resources

Biological research resources such as germplasm, DNA libraries and clones, are major products of functional genomics research. To be truly useful, these resources must be accessible to the research community through public repositories, which adhere to a common set of standards for curation, maintenance, distribution and quality control. Such standards still need to be developed.

Objective 3. Translational Plant Genomics – Application of Genomics Tools

As functions are assigned to genes in a few key plant species, the next, and most significant, opportunity will be to use this information to address questions of basic biology and to develop technologies that will produce plants with enhanced economic value and expanded utilities.

Application of genomics tools to understand traits of biological and economic importance

Humans have long exploited the unique biochemical, structural and physiological diversity of the plant world for agriculture and other benefits. Genomics has greatly increased the potential for understanding the basis for this diversity. Some of the many examples of specialized traits of both fundamental biological interest and economic importance are:

Tuber and bulb development: Many useful plant traits are not found in reference species, such as *Arabidopsis* or rice. For example, potato tubers are underground stems that are uniquely modified as starch-storing organs. Sweet potatoes are roots modified as starch-storing organs; and onions are leaves modified to form bulbs as nutrient-storing organs. Understanding the genetic networks that guide the development of these unique structures could yield important information for improving or further modifying these crop plants.

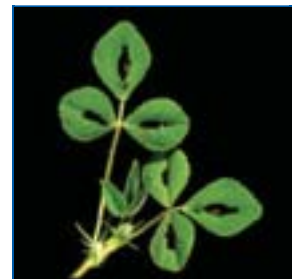
Wood formation: Production of softwood and hardwood is unique to woody plants. The composition and final stages of wood production are biologically and chemically complex, and expected to require the coordinated expression of many genes. Therefore, an understanding of the genetic networks for wood formation will contribute to future improvements in silviculture and natural resource management.

Fruit development: Edible fruits have been developed from wild species for differences in shape, size, texture, color, flavor, and chemical composition. The physiological and developmental processes yielding fruits such as peaches, apples, berries, oranges, papaya, bananas, peanut and others cannot be studied directly in reference species. However, as genome resources increase for a few key species, the biological processes for diverse species can be better understood.

Phosphate and nitrogen nutrition: Some plants are more capable of utilizing nutrients from the soil than others, especially nitrogen and phosphate, the main ingredients of fertilizers. Legumes form specialized association with microorganisms to utilize nitrogen from the air. Many plants including legumes and trees associate with a special group of fungi to utilize phosphate in the soil efficiently. Understanding how these plants accomplish efficient uptake and utilization of nitrogen and phosphate would help in the design of plants that require less chemical input for optimal performance in the field.

Expansion of genomics approaches to biodiversity, ecology and ecosystem studies

Understanding biodiversity is very likely to lead us to discover new mechanisms and higher order principles of biology. What are the key switches that have allowed plants to develop into such a vast array of morphological types



and occupy such a wide range of environmental niches? How and why do plants produce the widest array of chemical compounds in the natural world, and how is it that the spectrum of chemicals in plants can adjust so rapidly to environmental changes? How have gene functions changed as plant species have evolved to meet the challenges of new environments and new ecosystems? The NPGI project portfolio should include significant and broad efforts to understand plant interactions with pathogenic, mutualistic, and symbiotic organisms. New alliances among basic and applied biologists, plant and animal breeders, and computational experts will allow us to understand plants within the context of their biotic and abiotic environment.

Expansion of genomics approaches to renewable resources and novel biomaterials

Renewable resources hold the promise for stabilizing the losses incurred from our accelerated use of nonrenewable resources. Ultimately, whole genome analysis will allow us to develop rational strategies for optimizing renewable resources and for facilitating environmental remediation. For example, a small increase in the efficiency of photosynthetic energy capture and conversion, stimulated by translation of genomic information to an applied context, would have a major impact on plant production. Similarly, the discovery of new plant-based materials or the introduction of new genetic processes could revitalize plant improvement strategies in renewable resources.

Objective 4. Bioinformatics in Every Plant Scientist's Research Tool Box

NPGI has produced, and will continue to produce, enormous amounts of plant genome data, which need to be made accessible to a broad community of scientists in a useable form. A measure of the success of the NPGI will be in the direct usefulness of genome information. Significant and broad efforts should be directed toward programs that enable individuals or groups to access, analyze and compare data. The engineering of information systems, the development of data-mining tools, and the creation of computation based predictive models for functional analysis will continue to advance the goals of the NPGI.

Develop informatics tools to access and use plant genome databases

High throughput genomics technologies have led to a flood of sequence information, gene expression array data, and map data. All databases must have capabilities that will allow the broadest access by the community. Open access will lead to the widest utilization of the data and the development of innovative and more sophisticated tools, which, in turn, will enable individuals or groups to access and query all the available, current and future resources in the most imaginative ways possible.

The plant community should try to utilize existing tools for informatics to the largest extent possible. For example, the Generic Model Organism Project (GMOD) is a joint project between the National Human Genome Research Institute and the Agricultural Research Service, and aims to develop generic software modules for common function of a model organism genome data. Networking with these and other similar efforts will be an efficient and cost-effective way to leverage investments already made.

Build community databases with standards for interoperability

Databases should be developed that incorporate a common set of standards and interfaces in order for individuals located anywhere in the world to make full use of all publicly available resources. One way to make databases become interoperable is through the development of controlled vocabularies. The plant community (e.g., *Arabidopsis* and maize) is already actively participating in the Gene Ontology (GO) consortium that is developing controlled vocabularies for all model genomes. This effort should be encouraged for any new community and datasets.



Institute an internationally coordinated data repository mechanism

Data repository mechanisms and standard operating principles for the reference species *Arabidopsis* and rice as well as species-specific databases are critical and should be coordinated internationally.

Develop new algorithms to analyze plant genomics data

Computational resources and analytical tools that can mine genomic data and lead to hypothesis testing, validation and application, are essential and should be made freely available. Especially needed at this time is the development of algorithms for comparative genomics and population genetics. These tools will lead to increased knowledge of fundamental plant processes such as photosynthesis, respiration, carbon and nitrogen metabolism, nitrogen fixation, primary and secondary metabolism, polyploidy and domestication, and plant-microbe associations.

Objective 5. Education, Training and Outreach

The ultimate success of the NPGI will be gauged by the impact it has in advancing the plant sciences in the U.S. The impressive array of plant genomics resources and tools that have been developed in the last five years will be most relevant when a broad community of scientists are fully engaged in the plant genomics revolution. Equally important will be clear communication of the importance of plant genomics research to the general public. In order for the NPGI to have maximal impact, it must educate and train the next generation of scientists in plant genomics. More specifically, several activities are considered important for the NPGI to undertake in the next five years.

Traineeships for undergraduates, graduate students and postdoctoral researchers in plant genomics research

Currently there is no training grant program specifically targeted to train the next generation of scientists in plant genomics. Plant genomics research provides an ideal opportunity for exposing students to the excitement of scientific discovery in a field at the cutting edge of biology. Such a training grant program would include recruiting and retention of underrepresented groups in plant science research.

Informatics training for both established and young investigators

In genomics research, bioinformatics is an essential research tool that must become a basic tool for all levels of researchers. Bioinformatics should be as seamless a part of everyday research activity as are other biological tools such as molecular biology, cell biology or biochemistry. Opportunities should be made available for established and new investigators to receive training in this area.

Mid-career training program on plant genomics for university and college faculty and plant science professionals

Well-trained plant science professionals of all disciplines, from basic science faculty to applied plant breeders, are needed in order to fully realize the impact of genomics research on the future of plant sciences. Mid-career training awards are needed to fill this need

Workshops to inform the broader research community about accessing and using the NPGI research resources

Enormous amounts of data and research resources and tools are being produced by NPGI supported projects. Workshops by major databases such as GenBank, TAIR (the *Arabidopsis* Information Resources) and Gramene (A Comparative Mapping Resource for Grains) should be held at major plant biology conferences to demonstrate: how to deposit data into these public databases; and how to access and utilize the content for research. Similar information workshops should be offered by the biological resource centers such as the Maize Genetics Cooperation Center and the *Arabidopsis* Biological Resource Center.



Outreach to the K-12 community

As with all scientific fields, the future of plant genomics research depends on recruitment of motivated students. Reaching out to K-12 schoolteachers would be an efficient way to reach a large number of potential future undergraduate students. Plant genome researchers and students can contribute to this effort by serving as resources for science education in K-12 schools, and inviting teachers and their students to participate in plant genome research. Expected outcomes include improved communication and teaching skills for graduate and undergraduate students, enriched learning by K-12 students, professional development opportunities for K-12 teachers, and strengthened partnerships between institutions of higher education and local school districts.

Objective 6. Consideration of Broader Impacts

Public-private partnerships

The NPGI continues to encourage public – private industry interactions in the conduct of plant genomics research. While the goals of a public research program such as the NPGI are different from the goals of industrial research programs, it should be possible to find a common ground where both parties can collaborate in a synergistic manner. A good example is the human genome SNP (Single Nucleotide Polymorphism) consortium that pooled the public and private resources to complete the human SNP database. The collaboration resulted in availability of a key human genome research resource quicker and with a higher quality than would have been possible with each working independently. Private sector participation in the international rice genome-sequencing project is another example where company resources (draft sequence of the rice genome) were pooled with the sequences produced with public funds, and as a result, the public release of rice genome sequence was accelerated.

The NPGI, as a public project, will not directly support research in private industry. The key to successful public-private collaboration is participation of both sides as equal partners by bringing their own resources to the project and sharing the results openly and quickly with the rest of the community. Under these conditions, the private sector investigators should receive proper credit for their contributions.

International research coordination and cooperation

All successful genome projects to date, including the *Arabidopsis* genome research project and the rice genome sequencing project, were carried out by cooperating groups of international scientists coordinating their activities in order to maximize efficiency and to minimize duplication of efforts. The resulting synergism advanced the science more rapidly and at a larger scale than was possible with each country working independently. The NPGI has utilized the principle that each international partner is supported by its own national sources, and that coordination is carried out by participating scientists based on the science. As more and more countries are organizing national plant genomics programs, the NPGI expects that there will be increased international collaborative research in this area. The NPGI will encourage NPGI-supported research projects to take full advantage of opportunities for international collaboration.

Communicating the outcomes of plant genome research to the public

The general public should be given accurate information to make informed decisions on issues such as genetically modified crops and the role of biotechnology in society. It is the responsibility of every scientist involved in the NPGI projects to contribute accurate information about the application of plant genomics research to daily life. Research is needed to identify methods for more effective communication with the general public.



Guiding Principles for NPGI, 2003-2008

The NPGI will follow the guiding principles as in the original plan. They have served well for advancing the plant genomics for the nation and are expected to continue to serve the community well for the next five years. To reiterate, those guiding principles are:

- The Initiative should be viewed as a long-term project, governed by a periodically updated plan based on scientific progress and with stakeholder input as a basis for setting goals and priorities.
- All research resources, including data, software, germplasm, and other biological materials and research tools should be openly accessible to all in a timely manner.
- The Federal portion of the Initiative should be coordinated by an NSTC interagency working group with USDA, NSF, DOE, NIH, OSTP, and OMB representation.
- All awards should be made based on scientific merit and rigorous, competitive peer review.
- Partnerships with the private sector and the international community are vital for success.



IV. Cost Estimates for Achieving Objectives

In the initial planning document, it was estimated that approximately \$400 million would be needed to accomplish the goals set by the IWG in 1998 for the NPGI. In the past five years, NSF, DOE and USDA have expended approximately \$350 million. These investments have opened new and exciting research opportunities. Expanded opportunities are now available to a large segment of the plant sciences research community interested in genomics of diverse plants beyond a few key model plant species. In order to fully capitalize on the investments made to date and to enable further advances in plant genome research, the IWG recommends continued investment in the NPGI by all participating agencies. The IWG currently estimates that achievement of the objectives outlined in this five-year plan could be accomplished through an investment, over the next five years, of approximately \$1.3 billion total as follows. These estimates, which are subject to future change and refinement, do not represent a commitment to specific or cumulative resource levels by the Federal government or any other partners in annual budget proposals.

- \$400M for generating sequences and sequence resources for genome structure and organizational studies will result in the production of: (1) a completely finished rice genome sequence; (2) completely finished and mapped sequences of gene-rich regions of the maize genome; (3) highly accurate draft sequences of gene-rich regions of several key plant species; and (4) a variety of genome analysis tools to study structure and organization of a large number of plant species of economic importance.
- \$200M for functional genomics studies will allow US scientists to participate in international projects to determine the function of all of the genes in *Arabidopsis* and rice. The resulting functional genomics research resources will be shared freely and quickly, building a foundation for functional genomics research for all plant species.
- \$300M for translational genomics studies will enable a broad community of scientists to begin applying the knowledge, resources and tools of genomics to understand the fundamental biology of plants and the underlying mechanisms for economically important plant processes.
- \$250M for data management and informatics tools development will enable a broad community of both basic and applied scientists to utilize the outcomes of NPGI research activities. \$250M is a conservative estimate since all plant genome research activities described above will include informatics as an integral component, and thus the actual expenditure for data management and informatics will be considerably higher.
- \$125M for training, education and outreach will allow establishment of a NPGI training grant program and incorporation of training activities in all NPGI research activities.

The above budget estimates and distribution are based on the current state of science and technology in plant genomics. Continued advances in technology are expected to decrease costs in some areas, while new technologies and unexpected opportunities will necessitate increased investments in other areas. The Interagency Working Group will continue to monitor and report annually the constantly changing opportunities and needs to maintain effective and efficient utilization of resources in this coordinated national research program.



V. Appendix

In developing this document, the Interagency Working Group on Plant Genomes (IWG) has solicited and collected input from many sources. The following are some of the workshops and reports that the IWG considered in developing the NPGI plan for 2003-2008.

- USDA/ASPB CSREES plant genome stakeholders workshop, November 2002 (<http://www.aspb.org/publicaffairs/stakeholders/>)
- Meeting of International Rice Functional Genomics Working Group at the International Rice Congress, Beijing, September 2002
- White Paper on *Medicago truncatula* genome initiative, September 2002
- The National Academy of Sciences Workshop report “*The National Plant Genome Initiative: Objectives for 2003-2008*,” August 2002 (<http://www.nap.edu/books/0309085217/html>)
- Multinational *Arabidopsis* Steering Committee’s long range plan on functional genomics of *Arabidopsis*, June 2002 (<http://www.nsf.gov/pubs/2002/bio0202/start.htm>)
- Managing Plant Genomics Resources Workshop, Asilomar, June 2002
- The International Genome Research on Wheat (IGROW) workshop, June 2002
- The IWG workshop discussion at the Plant and Animal Genome Conference, San Diego, January 2002
- The Extension Vision for the 21st Century Committee/ NASULGC report on “The Extension System. A Vision for the 21st Century,” February 2002
- Meeting of the international tomato genomics group at San Diego, January 2002
- Task Force on Building a Science Roadmap for Agriculture / NASULGC report, “A Science Roadmap for the Future,” November 2001
- Plant Genome Databases and Bioinformatics Workshop, September 2001
- White paper by the US Legume Crops Genomics group: A workshop report, July 2001 (http://129.186.26.94/Legume_Initiative/LegGenomicsPaper10Oct01.html)
- Maize Executive Committee Workshop on Maize Genome Sequencing, July 2001. Report published in *Plant Physiology*, Vol 127, pp 1572-1578
- Global Promusa Consortium Workshop, July 2001 (<http://www.promusa.org/genomics/genomics-web.pdf>)
- Meeting of the Cereals Genomics Initiative Group, Mexico, April 2001
- Soybean Genomics White Paper, January 2000
- 2010 Project workshop report, January 2000 (<http://www.arabidopsis.org/workshop1.html>)
- The National Plant Breeding Study series I, II, III & IV (1996-2000), USDA/CSREES

In addition, a number of workshop reports and position papers on various genomics subjects have been published or electronically posted, including microarray data handling (December 2001 and September 2002), next generation technology for genome sequencing (July 2002), international genome sequencing consortium (January 2002), and broadening participation in genome research (April 2002). Many of these reports were part of the planning activities of the National Human Genome Research Institute and can be found at their website, <http://www.genome.gov/Page.cfm?pageID=10001312>.



Acknowledgement

The Interagency Working Group on Plant Genomes acknowledges the assistance of Sharlene Weatherwax (Department of Energy), Anne Sylvester (National Science Foundation), Jane Silverthorne (National Science Foundation), Ed Kaleikau (U.S. Department of Agriculture), and Machi Dilworth (National Science Foundation) in preparation of this report. Courtney Snyder (National Science Foundation) provided assistance in production of the report.

Image Credits

Page 3 - Syntenic map of various cereals; Dr. Michael Gale, John Innes Centre, Norwich, United Kingdom.

Page 4 - Wheat; Dr. Calvin O. Qualset, Agriculture and Natural Resources, University of California.

Page 5 - Wild-type *Arabidopsis* flower; Dr. Detlef Weigel, Salk Institute and Max Planck Institute.

Page 6 - Rice; Agricultural Research Service, USDA.

Page 8 - Plant cell nucleus; Dr. E. Lam, Rutgers University.

Page 9 - *Medicago truncatula*; Agricultural Research Service, USDA.

Page 10 - Corn; Dr. Peggy Lemaux, University of California, Berkeley.

Page 11 - Soybean; Dr. Mark Tucker, Agricultural Research Service, USDA.

Page 12 - Loblolly pine; Dr. David Byres, Florida Community College Jacksonville.

Page 13 - Tomato; Agricultural Research Service, USDA.

Page 14 - Cotton; Agricultural Research Service, USDA.

Page 15 - Potato flower; Dr. Barbara Baker, University of California, Berkeley.

Abstract

The National Plant Genome Initiative (NPGI) was established in 1998 as a coordinated national plant genome research program by the Interagency Working Group (IWG) on Plant Genomes with representatives from the Department of Agriculture (USDA), Department of Energy (DOE), National Institutes of Health (NIH), National Science Foundation (NSF), Office of Science and Technology Policy (OSTP), and the Office of Management and Budget (OMB). Since 1998, the field of plant genomics has made tremendous strides under the NPGI. It has changed the way research is conducted in plant biology; it has attracted a new generation of scientists into the field; and it has contributed new information and knowledge to science. Through development of plant genomic resources, the NPGI has built a foundation on which the scientific community can advance research, not only in plant genomics but also in diverse disciplines ranging from fundamental biological sciences to biotechnology.

In this report, the IWG describes the NPGI plan for the next five years (2003 – 2008). The IWG solicited and received input from many sources, which formed the basis for this plan. Objectives for the next five years of the NPGI will build on recent scientific and technical advances to ensure continued advancement in plant genomics specifically and plant sciences in general. In order to fully capitalize on the investments made to date and to enable further advances in plant genome research, the IWG recommends continued and increased investment in the NPGI by all the participating agencies. The funds will continue to be expended on a competitive basis using rigorous peer review. As in the first five years of the NPGI, the IWG will monitor progress and report major accomplishments annually.

For further information, contact:

**National Science and Technology Council Executive Secretariat
at (202)456-6101 (voice) or (202)456-6027 (fax).**

Also available on the NSTC Home Page at http://ostp.gov/NSTC/html/NSTC_Home.html



EXECUTIVE OFFICE OF THE PRESIDENT
Office of Science and Technology Policy
Washington, D.C. 20502