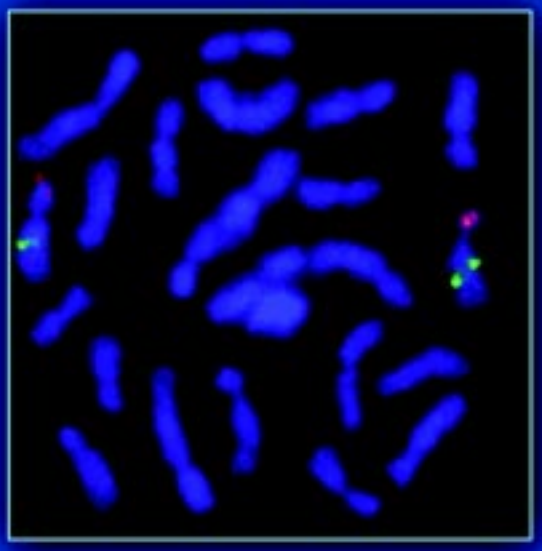
# National Plant Genome Initiative

## Progress Report

November 2000

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



# About the National Science and Technology Council

President Clinton established the National Science and Technology Council (NSTC) by Executive Order on November 23, 1993. This 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.

The NSTC is chaired by the President. Membership consists of the Vice President, Assistant to the President for Science and Technology, Cabinet Secretaries and Agency Heads with significant science and technology responsibilities, and other senior White House officials.

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 202-456-6100.

**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:** Late potato blight, caused by the fungus *Phytopthora infestans*, rapidly infects susceptible potato varieties, leading to almost complete destruction of the crop within two weeks. This disease was the cause of the Irish potato famine in the 19th century. Today, late potato blight still causes economic losses in the US and in developing countries such as Mexico. The NPGI supports a project designed to study genes found in wild potato species that confer natural resistance to late potato blight. The ultimate aim of the project is to develop improved potato varieties resistant to this deadly disease. [Center – Potato chromosomes marked with a fluorescent tag for a late blight resistance gene location: Left side, from top – Late blight infection on the stem of a cultivated potato plant; – A wild species of potato naturally resistant to late potato blight; Late blight disease spreading in a potato field; and potato diversity. Courtesy of the Potato Structural and Functional Genomics Project (https://www.fastlane.nsf.gov/servlet/showaward?award=9975866).]

# National Plant Genome Initiative

## Progress Report

November 2000

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

THE WHITE HOUSE

WASHINGTON

November 14, 2000

Dear Colleague:

This report provides an update from the National Science and Technology Council (NSTC) Interagency Working Group (IWG) on Plant Genomes on the progress of the National Plant Genome Initiative (NPGI). The NPGI was established in summer of 1997. The IWG published the NPGI's five-year plan in January 1998 and its first annual progress report in October 1999 (http://www.ostp.gov/html/genome/genome_1.html). This new progress report documents progress made since the publication of the first annual report and highlights selected advances and new scientific opportunities.

Impressive scientific progress has been made toward the goals set in the initial five-year plan for the NPGI. New discoveries and technological developments are pushing the frontiers of plant genomics in a manner that allows scientists to explore new ideas at an unprecedented rate. A large community of scientists participates in the NPGI, and their efforts are broadening educational and research opportunities for students and scientists across the Nation. I am confident that the IWG will continue to guide the NPGI so that it will be responsive to new scientific opportunities and in meeting new challenges these opportunities present.

Sincerely,

Neal Lane

Neal Lane
Assistant to the President
for Science and Technology

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

## Co-Chairs

Mary E. Clutter
Assistant Director for Biological Sciences
National Science Foundation

Eileen Kennedy
Deputy Under Secretary
Research, Education, and Economics
U.S. Department of Agriculture

## Members

Elke Jordan
Associate Director
National Human Genome Research Institute
National Institutes of Health

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

Sally Rockey
Deputy Administrator
Cooperative State, Research, Education
and Extension Service
U.S. Department of Agriculture

Gregory L. Dilworth
Division Director
Office of Basic Energy Sciences
U.S. Department of Energy

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

Noah Engelberg
Program Examiner
Office of Management and Budget

# Acknowledgement

The Interagency Working Group on Plant Genome acknowledges the assistance of Sharlene Weatherwax (Department of Energy), Jane Silverthorne (National Science Foundation), Ed Kaleikau (US Department of Agriculture), Leland Ellis (US Department of Agriculture), and Machi Dilworth (National Science Foundation) in preparation of this progress report.

# Table of Contents

# I. Executive Summary

The Interagency Working Group (IWG) for Plant Genomes was appointed in May, 1997, by the Office of Science and Technology Policy (OSTP), in response to a request from the Senate VA, HUD and Independent Agencies Appropriations Subcommittee. The charge was to identify science-based priorities for a national plant genome initiative and to plan for a collaborative interagency approach to address these priorities. The IWG recommended establishment of the National Plant Genome Initiative (NPGI) and developed a five-year plan for the NPGI, which outlined six specific goals. The NPGI was implemented in FY1998 with an appropriation for plant genome research at NSF.

The NPGI is envisioned as a long-term project that requires constant monitoring and a periodic reassessment in order to stay in step with scientific and technical advances in the field. The IWG issued the first year's progress report in October 1999, describing the status of the NPGI activities since the NPGI's inception. This report documents progress in the second year of the initiative, emphasizing selected project highlights and new scientific opportunities, rather than reviewing the overall status of the NPGI.

## Scientific Progress:

- One of the initial goals—accelerated sequencing of the *Arabidopsis* genome has been met successfully. The plant research community now has access, through a public database (GenBank), to the complete DNA sequence of the entire genome of a reference plant.

- The international rice genome-sequencing consortium has so far deposited 26 million basepairs (Mb) into GenBank, of which the US sequencing groups contributed 9.4 Mb. An unexpected development in the rice genome-sequencing project is Monsanto's offer to share with the international consortium its database of rough draft sequence of the entire rice genome.

- Researchers supported by the NPGI made an important fundamental discovery that the centromeric region of plant chromosomes contains functional genes. The centromere is the region of a chromosome that mediates chromosome movement and segregation during cell division. This work represents the first sequence level analysis of centromeres in any higher eukaryote, including human, fruitfly and nematode, and has directly challenged the established paradigm that the eukaryotic centromeres do not contain any active genes .

- Another contribution to fundamental biology from a NPGI project is the first successful cloning and functional confirmation of a plant quantitative trait locus (QTL) in tomato. Most of the economically important traits (e.g., yield, size and flavor of tomato) are controlled by QTL that result from complex interactions of multiple genes. This work successfully isolated the specific QTL responsible for increased fruit size.

## Genome Research Technology Development:

- Plant genome researchers have been exploring the use of microarray technology as a promising means to identify all genes involved in a certain process, such as responses to elevated $CO_2$ or cotton fiber formation. The NPGI researchers have now established this technology as a viable and important tool. Several community microarray resources are now available through the NPGI funded sites for individual researchers interested in this technology.

- The NPGI supports a number of novel genome research technologies. A technique called "chromatin charting" enables real-time visualization of movement of chromosomes in living cells within a plant. The information generated by this method has broad application potential for understanding the fundamental biology of plants and for optimizing transgenic strategies.

## Development of Community Research Tools and Resources:

- The number of Expressed sequence tags (ESTs) for various plant genomes continues to increase. In January of 1998, only *Arabidopsis* and rice had more than 20,000 ESTs entered into GenBank. As of August of 2000, 10 additional plants including maize, sorghum, wheat, barley, tomato, soybean, cotton and pine have more than 20,000 ESTs in GenBank. ESTs are considered a powerful tool for genome mapping and for gene discovery.

- In maize, another resource besides ESTs has been developed to aid gene discovery—maize tagged lines. They are a collection of maize plants that contain insertions of small pieces of DNA of known sequence (tag) distributed randomly throughout the genome. The research community now has access to at least three complementary sets of maize tagged lines.

- No large-scale genome program will succeed without robust informatics tools to store, access, and analyze all the massive amounts of disparate datasets resulting from their research activities. For the NPGI program, highly sophisticated informatics tools have been developed and constantly upgraded for the analysis of EST data, and the display and analysis of microarray data.

There is no question that the NPGI has changed the plant science research community in a profound way. Scientists are collaborating more across disciplinary, institutional and geographical boundaries. The research community at large is taking notice of the exciting scientific opportunities opened up in plant genomics and the field is attracting bright young scientists as well as established researchers with non-plant science training. The NPGI has had impact in increasing international research collaborations at the scientist's level as well as at the institutional level. There have been several positive interactions between industry and academia as evidenced by the Monsanto offer of their rice genome sequence data and Cereon's release of its data on single nucleotide polymorphisms in *Arabidopsis*.

## New Opportunities and Future Challenges:

- Plant genomics research tools and resources being generated by the NPGI projects are opening up new opportunities for the entire community of plant biologists to participate in the plant genomics revolution.

- New strategies to sequence large genomes, so-called rough draft sequencing, is gaining prominence. In plants, new strategies to condense gene rich regions of large genomes such as maize are being developed. This will surely be followed by an opportunity to begin large-scale sequencing of those regions, as an alternative to whole genome sequencing.

- The NPGI research community as well as the genomics research community at large, is struggling with the issue of data management including the need for new informatics tools. A lack of trained personnel in informatics is a serious issue for the field. The NPGI has an opportunity here to play a key role.

- As the public investment in plant genomics research increases, opportunities to collaborate with industry are on the rise. At the same time, the increasing impact of intellectual property issues on sharing of materials and data between private sector researchers and publicly funded researchers has become a big issue.

- Recent public debates over genetically modified organisms and plant biotechnology have been a wakeup call to the plant genome research community. The NPGI researchers are now aware that they must be involved in communicating to the public the importance of plant genome research and its societal impact.

# II. Introduction

The Interagency Working Group (IWG) for Plant Genomes was appointed in May, 1997, by the Office of Science and Technology Policy (OSTP), in response to a request from the Senate VA, HUD and Independent Agencies Appropriations Subcommittee. The IWG was comprised of representatives from the National Science Foundation (NSF), the Department of Agriculture (USDA), the Department of Energy (DOE), the National Institutes of Health (NIH), the Office of Management and Budget (OMB) and OSTP. The charge to the IWG was to identify science-based priorities for a national plant genome initiative and to plan for a collaborative interagency approach to address these priorities. The IWG recommended establishment of the National Plant Genome Initiative (NPGI) and developed a five-year plan for the NPGI, which outlined six specific goals (http://www.whitehouse.gov/WH/EOP/OSTP/NSTC/html/npgireport.html).

The NPGI began supporting new research activities in FY1998 as a result of a significant increase in funding at NSF. In FY1999, DOE, USDA and NSF provided additional research support for the NPGI when they began jointly supporting the US rice genome-sequencing project as part of the international effort. In FY2000, increased funding for the NPGI was provided through the authorization of the Initiative for Future Agriculture and Food Systems (IFAFS). It gave priority to projects integrating a research, extension, and education approach to achieve a multidisciplinary research environment that spans the entire spectrum from basic research to practical applications. The $32 million IFAFS agricultural genomics program specifically targeted the development of genomic resources for agriculturally significant plant, microbial and animal species, especially those for which a critical threshold of genetic markers and physically mapped traits must be achieved. The IFAFS awards favorably complement ongoing NPGI efforts and broaden the availability and utilization of plant genome resources in addressing a wide audience of current and future producers, researchers, and extension and education specialists.

In the short time since the establishment of the NPGI in 1998, the plant genome research community has made rapid progress. This is due in part to the fact that many of the fundamental, conceptual and technical advances fostered by the Human Genome Project laid the foundation from which the plant genome research community could leap forward. Also, the management of the NPGI by the participating agencies has benefited from the experience and expertise of NIH and DOE. The benefit is, in fact, mutual. It is known that certain processes such as the genomics of polyploidy (the number of chromosome sets in an organism) can be studied more easily in plants than in animals. Sometimes, new conceptual discoveries are made in plants first, such as the existence of genes in the centromere of eukaryotic chromosomes (see “Scientific and Technical Progress” below) or the phenomenon of gene silencing in transgenic organisms. Further investment into the NPGI will not only lead to new discoveries and breakthroughs in plant genomics but also enhance synergy and coevolution of the field of genomics in general.

The IWG coordinates the NPGI by documenting progress and achievements of the program, as well as by reassessing the need for additional resources. The first year’s progress report, issued in October 1999, highlighted the progress made since the NPGI’s inception and described future plans for implementation of each of these goals (http://www.ostp.gov/html/genome/index.html). This report documents progress in the second year of the initiative, and emphasizes specific project highlights and new scientific opportunities.

# III. Progress to Date

## A. Scientific and Technical Progress

### 1. Sequencing of model plant species

One of the goals of the NPGI is to support the high-resolution full genome sequencing of the model plant species *Arabidopsis thaliana* (mustard) and *Oryza sativa* (rice). Recent developments in both these sequencing projects are presented.

***Arabidopsis:***
The international consortium of scientists involved in an effort to sequence the entire genome of *Arabidopsis thaliana* has been making steady progress toward the goal of completing this activity by the end of 2000. In fact, a landmark was reached in August 2000 when the number of *Arabidopsis* sequence entries in GenBank, the public genome sequence database, surpassed 130 million basepairs—the estimated size of the *Arabidopsis* genome. Publication, in a major scientific journal, of the *Arabidopsis* genome sequences and their analysis is expected before the end of 2000. There is no question that the project will conclude successfully, ahead of schedule and within budget.

A highlight of this project during the last year was the publication of two papers in the journal Nature in December 1999, reporting the first complete DNA sequences of plant chromosomes, specifically *Arabidopsis* chromosomes 2 and 4. The reports provide new information about chromosome structure and organization, with evidence of duplication within the same genome and transfer of information between organellar and nuclear genomes. This latter finding was unexpected, and has spurred genomics researchers of other organisms to reexamine their data to determine if this is a universal feature of all eukaryotic genomes. In both chromosomes, approximately 50% of the genes identified by sequence have either no known function or only hypothetical function. The genes of known or predictable function encode for various major life processes in plants including metabolism, energy, cell division, transcription, protein synthesis, protein transport, membrane transport, cellular biogenesis, signal transduction, and defense mechanisms. In addition, completely unanticipated functions have been found, including genes with a high degree of similarity to those implicated in human diseases ranging from cancer to premature aging. The determination of how these human disease-related genes function in a plant is sure to yield new discoveries not only for *Arabidopsis* but all plant biology. This poses stimulating challenges to the plant biology research community as well as exciting opportunities for major discoveries.

***Rice:***
The International Rice Genome Sequencing Project (IRGSP) Working Group, consisting of members from Canada, China, the European Union, France, India, Japan, Korea, Singapore, Taiwan, Thailand, and the United States, has been implementing a coordinated effort to determine the sequence of the rice genome by the year 2008. The plan draws upon resources developed at many different institutions in different countries, seeking to capitalize on advances in genomic technology, coordinate development and use of genetic tools and expertise, and avoid duplication of infrastructure and human resources. The benefits are obvious to both academic and industrial researchers, with implications for the development of new strains of transgenic rice that display improved nutritional qualities, tolerate greater temperature extremes, or require less intensive arable land for cultivation.

In FY1999, the United States initiated a competitive program to support US participation in the international rice genome sequencing project led by the Japanese. USDA, NSF, and DOE jointly awarded $12.3 million over three years to two U.S. groups, the CCW (Clemson University, Cold Spring Harbor Laboratory, and Washington University) consortium and TIGR (the Institute for Genomic Research),

to sequence and annotate Chromosomes 3 and 10 of the Nipponbare variety of rice. These two groups have coordinated their research effort with other groups who contributed to the rice genome sequencing project with private funding. A central repository of the resulting data has been established (at http://www.usricegenome.org) and is available for the entire community of rice researchers. As of September 2000, the federally-funded groups have generated 9.4 million of the 26 million bases of IRGSP rice sequence deposited in GenBank. The latest estimates from the sequencing groups are that the completed sequence of rice chromosome 10, comprising approximately 5000 genes, will be available by spring of 2001. This project extends beyond domestic borders, with substantial input from the Rice Genome Program (RGP) of Japan, which developed and contributed an extensive set of physical markers, facilitating assignments of genome fragments for minimal redundancy and efficient sequencing. All international sequencing groups continue to coordinate their efforts and assess current technology and strategies, through a series of regularly scheduled workshops and meetings.

A significant development in the rice genome sequencing project was the contribution of industrially funded sequencing data by the Monsanto Corporation. This data represents an incomplete draft sequence of approximately 85% of the rice genome, and will greatly facilitate the international sequencing group's ability to map contiguous sequence locations on the rice genetic map. A dynamic dialogue has ensued, as all sequencing partners assess the best strategy with which to maximize the benefit of this rough draft sequence and to merge the information with a high-resolution physical map of the rice genome. The scope of this partnership between private and public plant international research sectors is unprecedented, and will likely result in an accelerated time frame for completion of the rice genome, now estimated to conclude around the year 2004.

Even before the completion of sequencing of the entire *Arabidopsis* genome, the primary sequence information has already been used by the scientific community at large. Scientists studying economically important plants such as corn and soybean can use *Arabidopsis* gene sequences as probes to identify genes of interest and to study their functions in their organism of choice. An example of the synteny or gene order conservation between *Arabidopsis* and soybean chromosomes was recently reported by D. Grant, P. Cregan and R. Shoemaker at Iowa State University (*Proc Natl Acad Sci* USA 97:4168-73. 2000). The same group has shown further colinearity between the *Arabidopsis* genome and the genomes of common bean and mung bean.

Rice genome sequences are also being utilized as soon as they are placed in the public databases. An example of how the rice genome sequence data might be utilized to understand another cereal species can be illustrated below in Figure 1, which shows the comparison of gene order information in rice to two other cereals, pearl millet and fox millet.

### 2. Techniques and methodology development for plant genome research

The NPGI also set a goal to develop methodologies that will enhance and

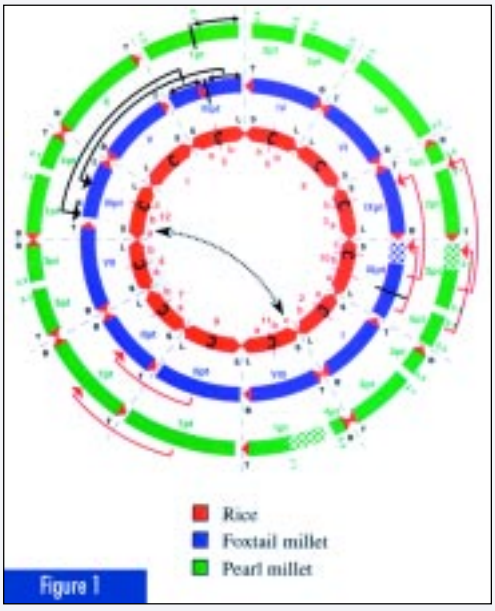



***Figure 1.*** *Synteny between rice and millet genomes. Reprinted from Devos, K. and Gale, M. Plant Cell 12: 637-646. Copyright © 2000 by the American Society of Plant Physiologists.*

facilitate the usage of plant genome information. Exciting developments in two emerging techniques are highlighted below.

***Microarrays:***
Microarrays are high-density displays of DNA fixed onto a solid support. They can be used to obtain genome-wide expression profiles under varying developmental and environmental conditions in wild type or mutant plants, as well as to map expressed sequences onto genomic fragments or to use as hybridization probes to obtain the corresponding full-length genes. The *Arabidopsis* Functional Genomics Consortium (AFGC), in collaboration with the Stanford Microarray Database (SMD), completed the first phase of an NPGI-funded project to develop, optimize, and utilize technologies for functional analysis of the *Arabidopsis* genome. The project exploits two powerful and complementary technologies: global gene expression analysis by microarrays and large-scale gene inactivation by T-DNA insertions. The results from over 95 arrays, performed on plants under a variety of developmental, environmental, and metabolic conditions, are now freely available to the research public (at http://afgc.stanford.edu). Through this user-friendly web interface, plant researchers worldwide can simply click on a particular spot on an array and directly view the primary data and obtain statistical information or experimental details.

The first results obtained from experiments performed using the AFGC microarrays reveal a powerful overall picture of changes in gene expression in plants on a genome-wide scale. Figure 2 shows part of a microarray showing differential gene expression in response to iron starvation.

A number of NPGI projects have refined the original microarray technology, resulting in an almost 10-fold reduction in sample production costs and significant increases in the quality of the array elements. The technology is still evolving and the NPGI researchers using microarray technologies have formed an electronic forum (plantarrays@genome.stanford.edu) to exchange ideas and to coordinate their activities. Over 400 scientists subscribe to the electronic forum.

***Chromatin charting:***
Although tremendous advances in sequencing complete plant and animal nuclear genomes have occurred in the past few years, the way this vast amount of genetic information is stored and physically manipulated within the living cell is still poorly understood. The recent

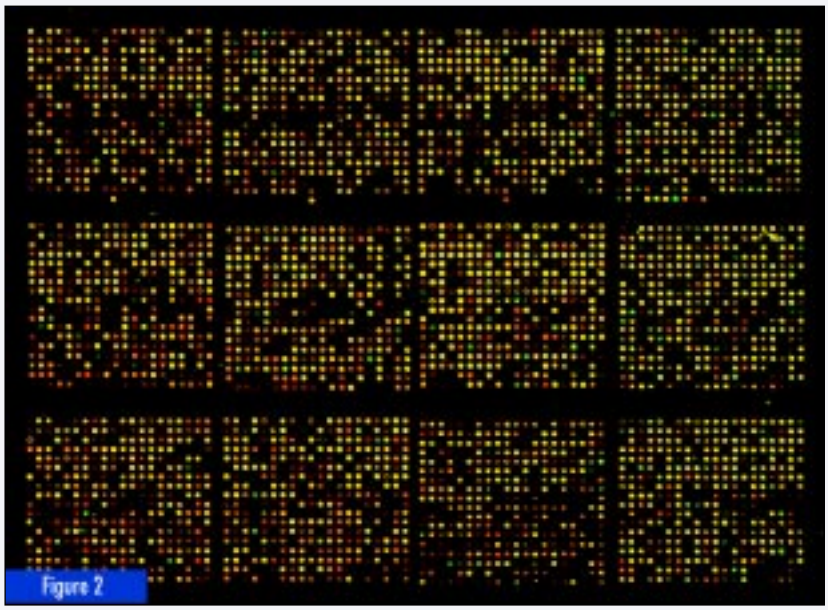
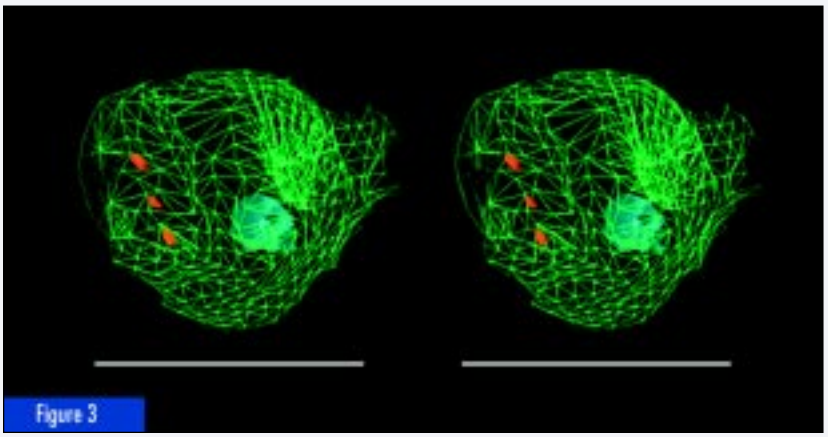



***Figure 2.*** *Differences in Arabidopsis gene expression in response to iron starvation. Red colored spots correspond to increased expression while green colored spots indicate decreased gene activity (data courtesy of M. L. Guerinot, Dartmouth College).*




application of the Green Fluorescent Protein (GFP) as a chromosomal tag has enabled scientists to use fluorescent microscopy to visualize chromosomes in living cells within a complex multicellular organism. A NPGI award has shown that it is possible to generate a three-dimensional video chronicle of chromosome movements within the nucleus by taking snapshots of the illuminated chromosomes over various time intervals. The success of this work opens up the exciting possibility of being able to visually tag chromosomes at defined locations and investigate the specific effect that a gene's physical position on a chromosome may have on its expression, transposition, or recombination. This may also lead to the development of spatial, temporal, and functional topology maps of the genome within a plant nucleus. The information generated will be relevant to our understanding of such epigenetic phenomena in plants as well as the optimization of transgenic strategies for expression of foreign genes.

### 3. Developing the biological tools and resources to study complex plant genomes

***Expressed sequence tags (ESTs):*** Expressed sequence tags (ESTs) are small sequenced pieces of mRNAs, the expressed portion of the genome. They represent a powerful tool for anchoring genetic markers to a physical genomic map, an essential feature of whole genome sequencing, as well as a platform from which to isolate the corresponding gene or its homolog in a different species. ESTs are currently the only practical means for gene discovery in plants with large genomes such as barley (5 Billion basepairs) and wheat (16 Billion basepairs) that do not lend themselves readily for complete genome sequencing. A major impact of the NPGI over the past two years has been the dramatic increase in the number of plant/crop ESTs deposited in the National Center for Biotechnology Information's (NCBI) dbEST database (see Figure 4).

All researchers benefit from the EST data. Their usefulness can be attested by the increasing number of research proposals received by the IWG agencies that use the EST data and clones as a starting point for identifying and cloning or determining the function of a gene of interest. While most of the EST projects funded by the NPGI have focused on building the EST databases as public resources, some projects have used the EST data for gene discoveries. For example, the plant stress genomics project at the University of Arizona (the consortium includes Purdue University, University of Nevada at Reno, and Oklahoma State University) has identified ESTs that are involved in stress sensing and stress signal transduction in plants. In another example, a set of 600 barley ESTs, selected on the basis of their priority and interest to the scientific community and the public, is being anchored to the linkage map and the developing physical map of the barley genome. NPGI support has allowed an added benefit of bridging this barley EST information to the wheat unigene mapping project, thus maximizing



***Figure 3.*** *A snap shot stereo view of a plant cell nucleus (outlined in green) using chromatin charting. Three GFP tagged sites on the chromosomes are visible as fluorescent red spots. The nucleolus containing ribosomal RNA is shown in blue (data courtesy of E. Lam, Rutgers University).*

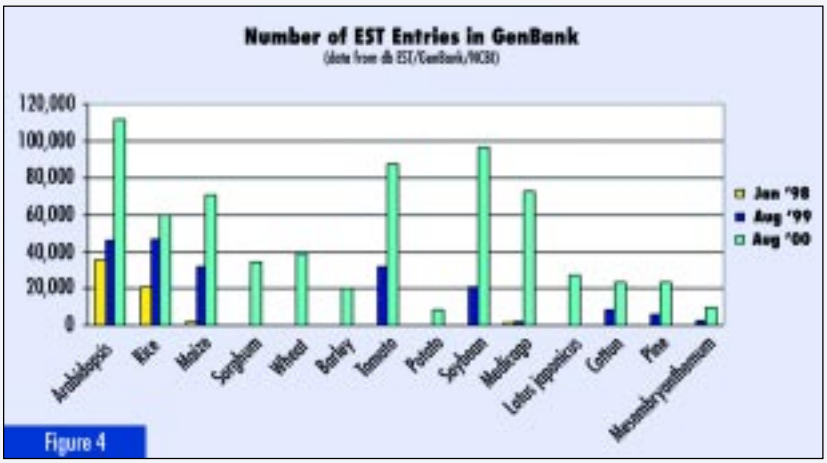


Figure 4

the opportunity to obtain syntenic information between these two important crops.

***Maize tagged lines for gene identification and function:***
The size of the maize genome is estimated to be 2.5 billion basepairs, which is 20 times the size of the *Arabidopsis* genome and 6 times the size of the rice genome. Also, it is estimated that genes contribute less than 3% of the entire maize genome, and are often buried within large stretches of repetitive and methylated DNA. Because of these characteristics, it is not practical at this time to contemplate a whole genome sequencing approach to identify genes in maize. Developing alternative efficient approaches to identify and characterize the relatively scarce but gene-rich portions of the maize genome has been a priority for maize researchers, and complementary approaches have been supported by the NPGI. One approach is building an EST database for the maize genome (see Figure 4 above). Another approach is to generate collections of plants that contain insertions of transposable elements, small mobile pieces of DNA that are distributed randomly throughout the genome. When a transposable element inserts in a gene, it results in an easily identifiable phenotype or trait, at the same time tagging the gene with a known DNA sequence and allowing for easy recovery of the disrupted sequence. Various transposon-tagged collections have been generated using Robertson's Mutator (Mu) and Activator (Ac), both naturally occurring maize transposable elements. The Mu-tagged collections (available at http://mtm.cshl.org/) allow investigators to either screen for insertions into known genes of interest, or to screen for desired growth characteristics resulting from unknown sites of transposon insertion. The Ac-tagged collection contains preferential transposon disruption of regions of the genome most likely to contain expressed genes; the nature of the Ac element allows the generation of clustered sets of tagged mutants containing multiple insertions into a single gene or region of the genome.

All of these maize tagged lines are freely available to the scientific community.

## 4. Increasing our knowledge of gene structure and function of important plant processes

The NPGI seeks to characterize important structural features of complex genomes and to understand not only the initial role they play in regulating gene expression but also their relationship to important processes in plant growth and physiology.

***Centromere structure and function:***
The centromere is the region of a chromosome that mediates chromosome movement and segregation during cell division. Its function in higher eukaryotes was presumed to be largely architectural, merely providing a physical scaffold for chromosome sorting. This view was



***Figure 4.*** *The number of expressed sequence tags (ESTs) for plant species that are available to the research community has dramatically increased in the 2 years of NPGI implementation.*

radically challenged by the results from an NPGI supported research project to define the structure and organization of the centromeres for all five chromosomes in *Arabidopsis.* These results presented the astounding finding that *Arabidopsis* centromeres contain a relatively high density of genes, some of which are actively expressed and involved with important functions such as chloroplast development. This work provided an unparalleled view of centromere composition, enabling a comprehensive analysis of the sequence motifs, DNA modifications, and structural features that contribute to centromere function in higher plants. More significantly, it represents the first sequence level identification of centromeres in any higher eukaryote, because to date the human, mouse, fruitfly, and nematode genome sequencing efforts have not included centromeric regions. By stimulating genome researchers to study other organisms' chromosomal centromeres in search of important new genes that would otherwise have been overlooked, this work has significance beyond simply elucidating the general mechanisms that govern centromeric function.

***Quantitative trait loci (QTL):*** Most of the plant traits important to agriculture, such as increased size, yield, flavor, and drought tolerance, result from a number of different genes interacting with each other and with the environment. These genes act together to provide a quantitative difference, and they are referred to as quantitative trait loci, or QTLs. An NPGI project reported the first successful cloning and functional confirmation of a plant QTL. Using a molecular genetic map of tomato, researchers applied a novel technique, referred to as Advanced Backcross QTL Analysis, to focus in on a subset of the QTLs discovered in wild germplasm. Since the QTLs of modern commercial tomato varieties have been extensively modified by a long breeding history, a comparison between wild and domesticated tomato QTLs enabled scientists to zero in on the specific QTL responsible for the desired trait of increased fruit size (see Figure 5). The development of this novel strategy might facilitate identification and transfer of useful QTLs from wild species for adaptation to economically important plant species.

### 5. Developing the appropriate data handling and analysis capabilities

Because the NPGI makes it a condition of an award that results generated with NPGI support be placed in the public domain, the amount of data being released into various publicly accessible databases is increasing daily at an exponential rate. The kind of data available is extremely diverse, ranging from DNA sequences such as genome and EST sequences, to detailed mapping data for all sorts of plant species, to expression profiles of various genes under all sorts of growing conditions, to description of biological resources such as DNA clones and tagged lines, to tutorials. In order to manage and make efficient use of these massive amounts and disparate types of data stored in different formats, development of informatics has been a top priority for the NPGI.


***Query and analysis of EST data:*** One of the most comprehensive resources available at present for the query and analysis of EST data is TIGR's Gene Indices and Expressed Gene Anatomy Database. This database provides a non-redundant view of the ESTs and genes for each species, as well as "data on their expression patterns, cellular roles, functions, and evolutionary relationships". A major anticipated use of ESTs is for gene expression studies using microarrays.



***Figure 5.*** *Scientists demonstrate that a single quantitative trait locus (QTL) in tomato is responsible for the difference between the small-fruited wild type species on the left and the commercial cultivar shown on the right. Reprinted with permission from Frary, A., Nesbitt, T.C., Frary, A., Grandilllo, S., van der Knaap, E., Cong, B., Liu, J., Meller, J., Elber, R., Alpert, K.B., and Tanksley, S.D.* Science ***289****: 85-88. Copyright © 2000 American Association for the Advancement of Science.*

***Coordinating plant genome informatics and related activities:*** As the results of NPGI supported projects begin to be placed in the public domain, it has become apparent that although individual research groups are developing various informatics tools to meet their own needs, these tools were not being widely shared primarily due to a lack of practical dissemination mechanisms. The NPGI awardees have now established a web-based resource for sharing information about research and related activities funded through the NPGI. This resource will serve several roles. First, it will provide an internal resource for rapid communication of ideas and issues among members of ongoing projects. For example, software developed by an individual project for its own use might have a broader application. Small improvements in an experimental protocol made by an individual project might be of interest to others. This website can be used to share experimental details among a broad group of plant genome researchers. Also, recruitment of students, training opportunities, and efficient management strategies might be shared on this site. Issues about the societal impact of the plant genome could be debated in an informal forum, as well. Second, it will serve as a resource to members of the community seeking information about NPGI projects or opportunities for participation. Third, it will serve as a site for informing the community about new initiatives and soliciting input on a variety of issues. In summary, this resource will facilitate the interaction between members of NPGI projects and broaden the participation of the community in sharing the resources generated by these projects.

## 6. Impact on plant science research and the community

The impact of NPGI funding to the plant science community is being felt beyond scientific and technical advances of the field. A large infusion of funding into a traditionally underfunded area such as plant sciences has made a major change in the way the overall research community views the field. The publicity surrounding the NPGI has highlighted the scientific as well as practical significance of studying plants. Plant scientists themselves have become reenergized about their research and that excitement is infectious. The maize research community has seen a resurgence in interest from early career scientists, including some of the best and the brightest students seeking to enter the field of plant genomics.

The NPGI has provided an opportunity to the plant science research community to conduct research in ways that were not possible or not thought of before. An emphasis on collaborative research as well as on a large-scale infrastructure building activities for the first two years of the NPGI has resulted in scientists collaborating across institutional, geographical and disciplinary boundaries. For example, the wheat genomics project involves principal investigators from 12 institutions in 9 states. The research directly supported with the NPGI funding is currently carried out by researchers at universities, colleges, private research institutions, and government laboratories across the country. Although much of this is accomplished through coalitions of large, multi-institutional consortia, the availability of Internet-accessible whole genome research resources combined with increasingly computer-literate student populations will increase participation of individual laboratories located at small institutions.

The NPGI has played a role in drawing scientists from other disciplines, including information science, computer science, mathematics, statistics, chemistry, engineering, and animal and microbial genomics. Within the plant science community, the NPGI has brought together scientists from all subdisciplines of biology from molecular/cell biology and genetics to biochemistry and physiology, to plant pathology and nematology, and to systematics, population biology and ecology.

Finally, the NPGI is helping to accelerate the plant genome revolution. Having

the information about the complete genome sequence of a reference plant has changed the way individual researchers conduct research. Individual investigators no longer have to devote a year to isolate, clone, and sequence a gene. They can simply use the sequence information in public databases to proceed directly to study biological questions. Also, biologists now have a choice of approaches. In the pre-genomics era, investigators studied one gene at a time. In the post-sequence era, it is possible to use a systems approach and to study a network of genes. Both approaches are complementary and important to advance the biological sciences.

## B. Impact Beyond Science

### 1. Human resource development and education

Numerous opportunities for productive exchange of scientific data and strategies have arisen through the NPGI. The NPGI-supported activities are now a core of many major conferences and workshops, both on specialized technical topics and also with broader scientific and policy agendas relating to plant genome research. The 8th Plant and Animal Genome Meeting held in January 2000 enabled scientists to participate in workshops ranging from specialized topics on horticultural plants, sugarcane, barley, poultry, and cattle, to general subjects on database construction, transcriptional profiling, and genomic computing techniques.

As the number of projects supported by the NPGI has increased, a demand for researchers trained in genomics has increased steadily. The corresponding explosion in industrial investments in genomics and gene discovery activities has increased employment opportunities for skilled genomics researchers. Additionally, it is becoming evident that novel strategies for experimental validation of integrated gene function will only be successful with sufficient training in biochemistry and plant physiology. NPGI-supported projects can and do play a critical role in training people in genomics and associated areas such as bioinformatics, and in building a pipeline of skilled researchers.

### 2. International collaboration

The international rice genome research consortium is making steady progress, in which the US groups funded by the NPGI are taking an active role.

The NPGI also serves as a point of reference and contact for national and international collaborations in plant genomics research. An example is the NPGI project on *Medicago* genomics, a legume closely related to important food and forage crops. The project has built a productive international collaboration, coordinating its activities with similar projects in Europe. More recently, collaborations with scientists in Hungary and the Netherlands will extend comparative genomics from *Medicago* to alfalfa and pea. A wealth of tools is emerging, including libraries, new genetic markers, and ESTs, to speed future development of improved legume crops worldwide. Further information about the tools being generated by the *Medicago* project can be found at http://chrysie.tamu.edu/medicago/mtdb/.

The IWG was approached by the State of Sao Paulo Research Foundation for advice to develop a research collaboration between the NPGI and their genome program. As a result of this dialogue, Brazil has joined the international rice genome sequencing project and talks are underway to develop a collaboration in sugarcane genome research between Brazilian scientists and interested US groups.

### 3. Technology transfer and industrial partnerships

Plant genomic research has benefited from a dynamic interchange between university-based and private industry research. In addition to Monsanto's public release of the rice genome rough draft sequences mentioned elsewhere in this report, Cereon Genomics, LLC, recently took a groundbreaking step in making privately-generated data available to the public sector. Their collection includes single nucleotide polymorphisms (SNPs) and small insertions or deletions (INDELS) between closely related *Arabidopsis* ecotypes. These data, available to

researchers from academic and not-for-profit institutions through The *Arabidopsis* Information Resource (TAIR) website (at http://www.arabidopsis.org/cereon/), represent a tremendous scientific resource that will greatly facilitate the isolation of genes by map-based cloning techniques.

As more and more genes for economically important plant traits are identified with the use of the resources developed through the NPGI, it is expected that transfer of genomic information to technological applications will occur for improved agriculture, health, energy, or environmental remediation. Scientists and industry managers are seeking new ways to pair genome resources, such as sequences of ESTs or genomes, with high throughput screening methodology or novel transgenic applications.

The number of U.S. utility patents referencing *Arabidopsis* has sharply risen since the beginning of the NPGI (see Figure 6). The number is expected to dramatically escalate with the completion of the *Arabidopsis* genome sequence this year.

Some specific examples include:

(1) A patent on the discovery and use of genes that determine whether the growing shoot of a plant will develop into a flower, so-called floral meristem identity genes. One immediate application of this invention allowed for earlier flowering in trees, such that insertion of the gene into poplar shortened the flowering time from the usual 6 years to a mere 6 months. This promises to shorten the time for breeding new trees with improved wood quality and pest resistance.

(2) A patent on seed-specific promoters - the specific regions of genes that control when and where certain genes will be active. A plant transformed with a gene controlled by this type of promoter will express that gene only in the seed, thus allowing the accumulation of desired products in the seed but not any other part of the plant. This has potential for enhanced and efficient production of plant products in a form amenable to commercial processing.

(3) A patent on a gene that encodes for p-hydroxyphenyl pyruvic acid dioxygenase (pHPP) - pHPP is a key enzyme involved in the biosynthesis of a wide variety of plant products including vitamin E and carotenoids. This invention will allow for modification of these products in plants to suit specific nutritional, agricultural, or energy demands.

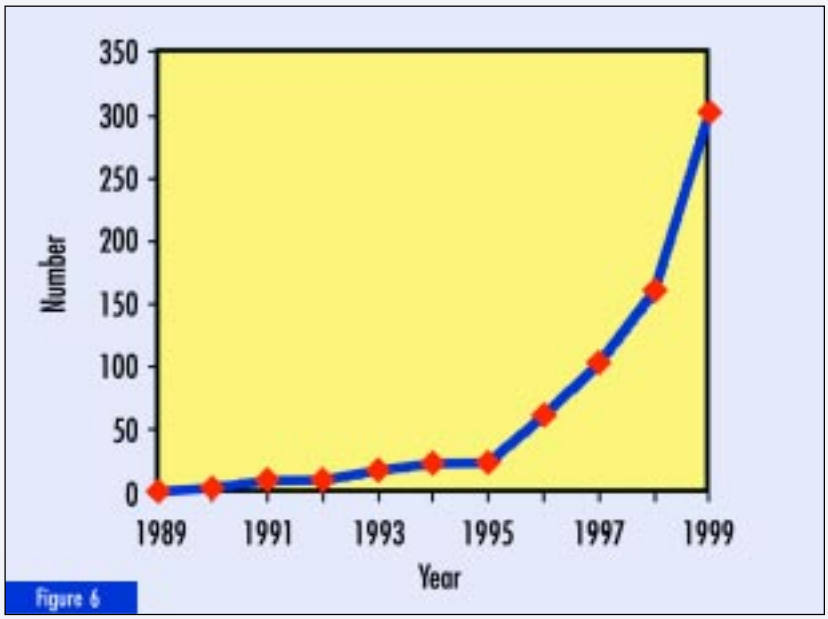



***Figure 6.*** *Number of US Utility Patents Referencing Arabidopsis. Data compiled from the Patent Full Text and Image Database, the US Patent and Trademark Office.*

### 4. Societal impact

The societal benefits of fundamental plant genome research are emerging in a number of exciting technological advancements announced this year. The new millennium ushered in a dramatic new application of agricultural biotechnology, that has the potential to provide affordable food and that will serve as better sources of essential nutrients to millions of people in developing countries. The genetic engineering of golden-colored rice grains to provide increased levels of vitamin A or sweet potatoes with resistance to virus attack represent just the beginning of the impact of plant biotechnology to the society. Recent public debates over genetically modified organisms and plant biotechnology have heightened awareness in the plant genome research community. The NPGI researchers are now aware that they must be involved in communicating to the public the importance of plant genome research and its societal impact.

# IV. New Opportunities and Challenges for the NPGI

In recent years, there has been a fundemental shift in the approach to analyzing large and complex genomes. The availability of high quality complete sequence information in key model plant species offers scientists a unique opportunity to exploit colinear gene relationships and compare the structural organization of selected genomic regions. The occurrence of unequal gene density may significantly affect future strategies for analysis of larger crop genomes, and may necessitate the development of novel strategies for limited, high-throughput sequencing of condensed, gene-rich genome segments. It is anticipated that the trend toward comparative genomics will continue, and a major challenge is to develop adequate research tools to provide for comparisons across species, e.g., for all grasses (including rice, wheat, barley, maize, and sorghum), and for genetic and physical maps as well as sequence data.

With the availability of increasing genome information, there is a need to integrate fundamental mechanisms of genome evolution and remodeling across a diverse spectrum of living things, from microbes to plants to animals. The biological function of ubiquitous genetic elements such as transposons, which may act on an evolutionary scale as agents of genome expansion and change, needs to be evaluated with respect to environmental pressures and organismal survival strategies. Recent evidence suggests that changes in the corn and barley genomes occurred within a relatively short evolutionary time frame by transposon-mediated genome expansion or deletion. An analogous phenomenon has been observed in *Drosophila* and humans. There is a fundamental need to understand the function of such universal elements of genomes that act in such different ways to confer an evolutionary advantage. Large-scale database management and algorithms for whole genome comparisons across different species will be increasingly necessary to address these types of questions.

Fostering of partnerships between the IWG agencies must continue, both to reduce redundancy and to capitalize on the special resources available to each agency. There exists a need to integrate the considerable wealth of germplasm and data being generated and administered by the various plant genome projects in a way that makes full use of the unique expertise and talents of each investigator. The genomics tools being generated will assist breeders as they face the agricultural challenges of the new century. Similarly, the genomics tools will help advance research to improve plants as energy feedstocks. Considerable investment into this area of research will provide an added synergy to the NPGI and strengthen the US position in plant genome research world-wide.

One of the most exciting outgrowths of the NPGI was the envisioning of an ambitious plan to identify the function of every gene and protein in *Arabidopsis thaliana* over the next ten years. This multinational effort, called the 2010 Project, seeks to use the information from the *Arabidopsis* genome sequencing project to analyze and dissect every component of a plant's metabolism, structure, and development. This information would be collected, collated and stored in a clickable database for investigators to examine fundamental integrated processes, essentially in the form of a complete "wiring diagram" of all the biological pathways in a virtual plant.

In the first few years of the NPGI, there has been an emphasis on supporting research collaboration groups and developing research tools/resources for the entire community. Some activities require an economy of scale and will continue to be best performed by collaborating groups. These groups have produced research tools and information that are freely available on line, creating an unprecedented opportunity for individual investigators located at any kind of institution to use them and to become full participants in the plant genome revolution. Existing IWG agencies' research programs are beginning to see this level of expanded participation. Clearly, it is an opportune time for the NPGI to promote full participation of the plant biology community in plant genomics.

The increasing impact of intellectual property issues on the exchange of materials and data between public and private sector researchers, and on the educational environment in academic institutions, has been and still is a concern for the NPGI. Also, recent public debates over genetically modified organisms and plant biotechnology have highlighted the need for increased communication with the public about the significance and relevance of plant genome research to the society. It is time for the NPGI to address these issues.

## Recommended Investment for the Next 3 Years (FY2001-2003)

In the original long-range plan for the NPGI, the IWG emphasized the importance of adjusting the goals and budget predictions "in such a rapidly advancing field as the genomics, where new technologies and discoveries may dramatically affect the costs and the direction of the science." Based on the progress described in this report and scientific opportunities emerging as a result of advances in the field of genomics, the IWG recommends a total investment of $600M for the next three years.

- $50M for continued participation in sequencing the rice genome — Availability of Monsanto's rough draft rice genome sequence will likely help accelerate the rate of production sequencing of the rice genome by the international consortium. However, it will not help the gap-closing and finishing process, which is proving to be significantly more difficult in rice than in other model systems of smaller genome size. Additional funding will allow the US sequencing groups to take the lead in developing new methods to sequence difficult regions of the rice genome, contributing to the overall goal to generate a contiguous, completed sequence of the rice genome by the year 2004.

- $100M for high-throughput sequencing of gene-rich regions of selected genomes of economic importance — This will assist gene discovery in organisms of large genome size such as corn, wheat, and barley.

- $325M for functional genomics, including: $50M for studying, at the genome-wide level, the interactions between plant and plant-associated microbes (both beneficial and detrimental); $75M for functional genomics of the model plant, *Arabidopsis* (the 2010 Project); and, $200M for functional genomics of plants important to agriculture, energy, health and the environment.

- $100M for informatics and data management

- $25M for training and education of the next generation of genome researchers, and activities to address societal impacts of plant genome research.

# V. Appendix: Glossary

**Base Pair (bp):** Nucleotide bases (adenine, thymine, guanine and cytosine) are the building blocks of DNA. Two molecules of nucleotide bases are held together by weak bonds. Two strands of DNA are held together in the shape of a double helix by the bonds between base pairs. The number of base pairs is used to describe the size of a DNA molecule.

**Centromere:** The constricted region near the center of a chromosome. This is the region of the chromosome where the two sister chromatids are joined to one another.

**Chromatin:** A complex of protein and DNA that make up chromosomes.

**Chromosome:** The self-replicating genetic structure of cells containing the cellular DNA that bears in its nucleotide sequences the linear array of genes. In prokaryotes, chromosomal DNA is circular, and the entire genome is carried on one chromosome. Plant genomes consist of a number of chromosomes whose DNA is associated with different kinds of proteins.

**Clone:** An exact copy made of biological material such as a DNA segment (a gene or other region), a whole cell, or a complete organism.

**Cloning Vector:** A piece of DNA, such as a plasmid, into which a DNA segment can be inserted, transferred into an organism, and replicated or reproduced.

**Comparative Genomics:** The practice of comparing gene or protein sequences with each other across whole organism genomes, in the hope of elucidating functional and evolutionary significance

**DNA (deoxyribonucleic acid):** The molecule that encodes genetic information. DNA is a double stranded molecule held together by weak bonds between base pairs of nucleotides. The four nucleotides in DNA contain the bases: adenine (A), guanine (G), cytosine (C), and thymine (T). In nature, base pairs form only between A and T and between G and C; thus the base sequence of each single strand can be deduced from that of its partner.

**Ecotype:** A locally adapted population of a widespread species. Such populations show minor changes of morphology and other characteristics, which are genetically induced. They can still reproduce with other ecotypes of the same species.

**Epigenetic Phenomenon:** Heritable changes in gene expression that occur without a change in DNA sequence.

**EST (Expressed Sequence Tag):** A unique, short DNA sequence derived from a cDNA library. ESTs are useful for localizing and orienting the mapping and sequence data reported from many different laboratories and serve as identifying landmarks on the developing physical map of a plant genome.

**Expression Profile:** A snapshot of the genes expressed in specific tissues of specific organisms at specific points in time. This snapshot tells an investigator where, when, and to what extent a particular gene is expressed in a cell, tissue, or organism and what physiological pathways are active in the cell.

**Functional Genomics:** Studies of the relationship between the structure and organization of the genome and the function of the genome as it directs growth, development, physiological activities, and other life processes of the organism.

**GenBank:** A public database where DNA sequences are deposited and made public. It is operated and supported by the National Library of Medicine, part of the National Institutes of Health, and is part of an international consortium of a gene sequence database.

**Gene:** The fundamental physical and functional unit of heredity. A gene is an ordered sequence of nucleotides located in a particular position on a particular chromosome that

encodes a specific functional product (i.e., a protein or RNA molecule).

**Gene Density:** The number of genes in a given length of DNA (e.g., 1 gene per 5 kb of DNA).

**Gene Recombination:** A process of chromosome exchange during cell division, resulting in new combinations of genes at a specific location on chromosomes.

**Gene Silencing:** The process of inactivating a gene.

**Gene Transposition:** Transfer of a gene to a new position on the same or another chromosome or plasmid.

**Genetic Map:** A map of the relative positions of genetic loci on a chromosome, determined on the basis of how often the loci are inherited together.

**Genetics:** The study of the patterns of inheritance of specific traits.

**Genome:** All the genetic material in the chromosomes of a particular organism; its size is generally given as its total number of base pairs.

**Genome Project:** Research and technology development effort aimed at mapping and sequencing some or all of the genome of human beings and other organisms.

**High Throughput Biology:** An experimental approach that generates massive amounts of raw data at the production scale using highly automated technologies such as genome sequencing technology or microarray technology, and processes the data by a batch method using computational and other information management tools.

**Human Genome Project:** The national effort, led by DOE and NIH, was started in the late 1980s. It includes several projects to (1) determine the sequence of human DNA, (2) develop new computational methods for analyzing genetic map and DNA sequence data, and (3) develop new techniques and instruments for detecting and analyzing DNA. While the ultimate objective is to understand the structure, organization and function of the human genome, the Human Genome Project supports studies on several model microbial and animal genomes. No plant genomes are targeted by the Human Genome Project.

**Informatics:** The study of the application of computer and statistical techniques to the management of information. In genome projects, informatics includes the development of methods to search databases quickly, to analyze DNA sequence information, and to predict protein sequence and structure from DNA sequence data.

**Library:** An unordered collection of clones (i.e., cloned DNA from a particular organism), whose relationship to each other can be established by physical mapping.

**Linkage Map:** A map of relative positions of genes on a chromosome. Genes inherited together are close to each other on the chromosome, and said to be linked.

**Microarray Technology:** New approach to the study of how large numbers of genes interact with each other. This technology provides a quantitative assessment of how a cell's regulatory networks control extensive gene sets simultaneously. The method uses a robot to precisely apply tiny droplets containing functional DNA to glass slides. Researchers then attach fluorescent labels to DNA from the cell they are studying. The slides are put into a scanning microscope that can measure the brightness of each fluorescent dot; brightness reveals how much of a specific DNA fragment is present, an indicator of how active it is.

**Nuclear Genome:** The part of the genome of an organism that is found within the nucleus of the cell. It encodes most of the structural and regulatory sequences necessary for cellular function.

**Organelle Genome:** The part of the genome of an organism that is found within an organelle, such as a mitochondrion or a chloroplast.

**Physical Map:** A map of the physical locations of identifiable landmarks on DNA (e.g., restriction enzyme cutting sites, genes); distance is measured in base pairs. The highest resolution map would be the complete nucleotide sequence of the chromosomes.

**Promoter:** The part of a gene that contains the information to turn the gene on or off. The process of transcription (reading of DNA sequences to produce gene products) is initiated at the promoter.

**QTL (Quantitative Trait Locus):** The location of a gene that affects a trait that is measured on a quantitative (linear) scale. These traits are typically affected by more than one gene, and also by the environment. Examples of quantitative traits are plant height (measured on a ruler) and body weight (measured on a balance).

**Reverse Genetics:** An experimental approach that begins with information about the primary DNA or protein sequence and uses this knowledge to generate targeted mutations (heritable changes) or altered expression levels. Observation of the resulting effects on the organism (a physical attribute such as the color of the flower or the shape of the leaf) yields information about the physiological function of the gene or protein. This is the reverse strategy to a classical genetics approach that proceeds from observing a defined genetic trait towards obtaining sequence information for a specific gene.

**Rough Draft Sequence:** A scaffold of known DNA sequence across an estimated 90% of the genome. This "working draft" sequence is of lower accuracy than the finished sequence and is not continuous across the genome. The data, however, is invaluable to many scientists as they strive to understand genes important to their own research. Generating the draft required researchers to sequence each piece of DNA only about 5 times, instead of the usual 10 times used for obtaining the highest quality sequence that is 99.99% complete and allows for only a single error in 10,000 bases.

**Sequencing:** Determination of the order of nucleotides (base sequences) in a DNA or RNA molecule or the order of amino acids in a protein.

**Signal Transduction:** The process by which an external signal is transmitted into and within a cell to elicit a response.

**SNP (Single Nucleotide Polymorphism):** A common, but minute, variation that occurs in DNA sequences of a genome. These variations can be used to track inheritance in families or species. SNP is pronounced "snip".

**Structural Genomics:** Studies of the structure and organization of the genome including DNA sequencing and physical and genetic mapping.

**Synteny:** The colinear relationships between the genomes of different organisms. Often chromosomal regions from related organisms, such as various grass species, contain corresponding genetic information and similar gene order.

**T-DNA:** The portion of the Ti plasmid of the bacterium, *Agrobacterium* that is transferred into the plant and integrated into the plant genome during infection by the bacterium. The wild-type T-DNA carries genes that cause the plant cells to grow as a tumor. T-DNA can be disarmed by removing these genes. The disarmed T-DNA can be used as a vector for genetic engineering in plants.

**Technology Transfer:** The process of converting scientific findings from research laboratories into useful products by the commercial sector.

**Transcription:** The synthesis of an RNA copy from a sequence of DNA (a gene); the first step in gene expression.

**Transgenic Organism:** An experimentally produced organism in which DNA has been artificially introduced and incorporated into the organism's germ line.

**Transposable Element or Transposon:** A heterogeneous class of genetic elements that can insert at new locations on chromosomes. They vary in structure, mechanism of transposition, the fate of the donor element, and their choice of target sites.

**Unigene Set:** A non-redundant set of EST clones.

## Abstract

The Interagency Working Group (IWG) for Plant Genomes was appointed in May, 1997, by the Office of Science and Technology Policy (OSTP), in response to a request from the Senate VA, HUD and Independent Agencies Appropriations Subcommittee. The charge was to identify science-based priorities for a national plant genome initiative and to plan for a collaborative interagency approach to address these priorities. The IWG recommended establishment of the National Plant Genome Initiative (NPGI) with the long-term objective to understand the structure and function of genes in plants important to agriculture, environmental management, energy, and health. The IWG issued the first year's progress report in October 1999, describing the status of the NPGI activities since the NPGI's inception (http://www.whitehouse.gov/WH/EOP/OSTP/html/genome/index.html).

The current report documents progress in the second year of the initiative, emphasizing selected project highlights and new scientific opportunities. Progress has been made in all areas including new scientific discoveries, development of research tools and resources that allow the entire scientific community to participate in the NPGI activities, and increased training opportunities for students. These advances have opened up new opportunities and presented new challenges. In order to take full advantage of new opportunities and to meet the challenges, it is recommended that additional investments be made across all participating agencies at the level, for the next three-year period, of a total of $600 million.


## For further information, contact:


**National Science and Technology Council Executive Secretariat at (202)456-6120 (voice) or (202)456-6026 (fax).**



*Also available on the NSTC Home Page via link from the OSTP Home Page at:*
http://www.whitehouse.gov/WH/EOP/OSTP/html/OSTP_Home.html


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