NATIONAL PLANT GENOME INITIATIVE
FIVE-YEAR PLAN: 2014–2018

PRODUCT OF THE

National Science and Technology Council

MAY 2014

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

May 16, 2014

Dear Colleagues:

The enclosed report provides a five-year (2014-2018) plan for the National Plant Genome Initiative (NPGI). Implementation of this plan will build on significant advances made under the first three NPGI strategic plans. The NPGI will continue to advance the frontiers of plant science as wells as accelerate basic discovery and innovation related to economically important plants and processes that enable enhanced management of agriculture, natural resources, and the environment to meet the Nation's needs.

In developing this five-year plan, the Interagency Working Group on Plant Genomes (IWGPG) — a Working Group under the Life Sciences Subcommittee of the National Science and Technology Council's Committee on Science — received input from a diverse set of stakeholders and many sectors of the scientific community, including the National Academy of Sciences, as well as industry, professional societies and producer/grower organizations.

Great progress has been made in the past sixteen years in the domain of plant genomics, including with respect to increasing public access to relevant scientific information and data. Much more remains to be done, however, to capitalize on these previous advances and integrate them fully into the fabric of the national plant genomics infrastructure — including tools, people, and collaborative partnerships both domestically and internationally. In addition, it is critical to continue and accelerate research efforts in plant genomics in order to address the challenges brought about by the impacts of climate change and increasing demand for plant-based food, fuels, and materials.

The 2014-2018 NPGI Strategic Plan will help ensure that America remains a leader in harnessing the power of the plant genome for enhanced human nutrition, production of feedstocks for bioenergy, resilience to climate change, and enhanced global food security for a safer and more sustainable world. I thank the NPGI Working Group for all of its contributions and look forward to continuing progress in this important domain.

Sincerely,

John P. Holdren

John P. Holdren
Assistant to the President for Science and Technology
Director, Office of Science and Technology Policy

About the National Science and Technology Council

The National Science and Technology Council (NSTC) is the principal means by which the Executive Branch coordinates science and technology policy across the diverse entities that make up the Federal research and development (R&D) enterprise. One of the NSTC's primary objectives is establishing clear national goals for Federal science and technology investments. NSTC prepares R&D packages aimed at accomplishing multiple national goals. The NSTC's work is organized under five committees: Environment, Natural Resources, and Sustainability; Homeland and National Security; Science, Technology, Engineering, and Mathematics (STEM) Education; Science; and Technology. Each of these committees oversees subcommittees and working groups that are focused on different aspects of science and technology. More information is available at www.whitehouse.gov/ostp/nstc.

About the Office of Science and Technology Policy

The Office of Science and Technology Policy (OSTP) was established by the National Science and Technology Policy, Organization, and Priorities Act of 1976. OSTP's responsibilities include advising the President in policy formulation and budget development on questions in which science and technology are important elements; articulating the President's science and technology policy and programs; and fostering strong partnerships among Federal, state, and local governments, and the scientific communities in industry and academia. The Director of OSTP also serves as Assistant to the President for Science and Technology and manages the NSTC. More information is available at www.whitehouse.gov/ostp.

About the Interagency Working Group on Plant Genomics

The Interagency Working Group on Plant Genomics (IWGPG) serves as part of the internal deliberative process of the NSTC and provides overall guidance and direction for plant genomics research across the Federal government agencies. As laid out in the group's charter (Appendix A), the IWGPG's goal is to develop a five-year plan for the National Plant Genome Initiative (NPGI) for 2014–2018 through activities that (1) identify strategic research needs and resource gaps for the development of sustainable systems for food, bioenergy, and industrial feedstock production, considering the minimization of inputs such as water, energy, pesticides, and fertilizer, the effects of climate change, increased agricultural productivity, and minimization of environmental impacts through plant genomics and (2) identify opportunities for Federal agency coordination, cooperation, public/private partnerships, and associated opportunities to enhance training, education, and public outreach through the engagement of stakeholder communities.

About this Document

This plan was developed by the IWGPG, an interagency working group under the Life Sciences Subcommittee (LSSC) of the NSTC Committee on Science (CoS). This document is published by the OSTP.

Copyright Information

Printed in the United States of America, 2014.

National Science and Technology Council

Chair

John P. Holdren
Assistant to the President for Science and Technology
Director
Office of Science and Technology Policy

Committee on Science

Co-Chairs

Francis Collins
Director
National Institutes of Health

Cora Marrett
Acting Director
National Science Foundation

Philip Rubin
Principal Assistant Director for Science
Assistant Director for Social, Behavioral, and Economic Sciences
Office of Science and Technology Policy

Executive Secretary

Mayra N. Montrose
National Science Foundation

Interagency Working Group on Plant Genomics

Co-Chairs

Catherine Ronning
Biological and Environmental Research
U.S. Department of Energy

Jane Silverthorne
Directorate for Biological Sciences
National Science Foundation

Executive Secretary

Jack Okamuro
Agricultural Research Service
U.S. Department of Agriculture

Members

Noah Engelberg
Office of Management and Budget
Executive Office of the President

Adam Felsenfeld
National Human Genome Research Institute

Kenneth Haymes
Environmental Protection Agency

Randy Johnson
Forest Service
U.S. Department of Agriculture

Edward Kaleikau
National Institute of Food and Agriculture
U.S. Department of Agriculture

Warren Johnson
Smithsonian Conservation Biology Institute
Smithsonian Institution

Gail McLean
Basic Energy Sciences
U.S. Department of Energy

Kay Marano-Briggs
U.S. Geological Survey

Diane Jofuku Okamuro
Directorate for Biological Sciences
National Science Foundation

Pablo Rabinowicz
Biological and Environmental Research
U.S. Department of Energy

Jean Ristaino
Jefferson Science Fellow
U.S. Agency for International Development

Kay Simmons
Agricultural Research Service
U.S. Department of Agriculture

Carlos Peña
Office of Science and Technology Policy

Acknowledgements

The IWGPG acknowledges the following for their input and assistance: stakeholder groups who participated in the NPGI listening sessions; organizations that shared their plans and reports as input into the NPGI plan development process; agency staff for providing information and materials; Dr. Karen Cone for liaising with the LSSC; Dr. Carlos Peña for guidance and input during the process; and OSTP staff for logistical support.

Table of Contents

Executive Summary 9

Introduction 10

Major Accomplishments for 2009–2013 11

Plan for 2014–2018 15

Agency Contributions 32

Appendix A 37
IWGPG Charter

Appendix B 41
Accomplishments for 2009-2013

Appendix C 53
Stakeholder Input

Appendix D 54
Sequenced Plant Genomes

Appendix E 56
Databases and Tools Currently Supported Through the National Plant Genome Initiative (NPGI)

Appendix F 57
Developing Country Impacts of National Plant Genome Initiative (NPGI)-Supported Resources

Appendix G 60
Glossary

Appendix H 61
Acronyms

Executive Summary

The goal of the National Plant Genome Initiative's (NPGI) is to develop a basic knowledge of the structures and functions of plant genomes and to translate this knowledge into a comprehensive understanding of all aspects of economically important plants and plant processes of potential economic value. By bridging basic research and plant performance in the field, the NPGI seeks to accelerate basic discovery and innovation to improve the practice of agriculture, reduce the demands on environmental resources, and address challenges posed by global climate change.

Since 1998, the NPGI has been coordinated through successive five-year plans, each of which was developed by the Interagency Working Group on Plant Genomics (IWGPG) using inputs from diverse stakeholders. This document reports on the accomplishments of the NPGI for the past five years (2009–2013) and outlines a strategic plan for the NPGI for the next five years (2014–2018).

This plan was developed using inputs from the broader plant-research communities, which are listed in Appendix C. It prioritizes government activities and investments to ensure the most effective use of resources while advancing research and its downstream applications to meet societal needs through development of critical research tools and infrastructure, increased public access to data and publications emerging from NPGI-funded activities, and enhanced partnerships with the private and international sectors. The NPGI will continue to adhere to the guiding principles first articulated in 1998, including timely access to the outcomes of NPGI investments.

There are six major objectives for the next five years of the NPGI. Objectives 1 and 2 focus on open-access data and knowledge sharing and expanding the interoperability of tools and databases, while objectives 3–6 focus on enhancing the application of genomics for agriculture, acceleration of plant breeding, improvement of the practice of agriculture, reduction of the demands on environmental resources, and addressing challenges posed by global climate change.

Objective 1: Develop a new generation of databases and tools to enable every step in the innovation chain from basic discovery to translation into practical application and widespread use.

Objective 2: Create a network of plant germplasm resources to meet the Nation's research and breeding needs for the 21st century

Objective 3: Build tools to advance our understanding of plant biology and translate this knowledge into precision plant breeding for the development of sustainable systems for food, bioenergy, and industrial feedstock production

Objective 4: Empower the workforce to use a new generation of tools and resources

Objective 5: Build public-private partnerships to advance the translation of basic discoveries into practical application and support innovation in the practice of agriculture

Objective 6: Strengthen international partnerships to bring the benefits of new discoveries to all

I. Introduction

In 1998, the National Plant Genome Initiative (NPGI) was established as a coordinated national research program by the Interagency Working Group on Plant Genomics (IWGPG), with representatives from the U.S. Department of Agriculture (USDA), U.S. Department of Energy (DOE), National Institutes of Health (NIH), National Science Foundation (NSF), Office of Science and Technology Policy (OSTP), and Office of Management and Budget (OMB). In 2012, the IWGPG was reorganized under the Life Sciences Subcommittee (LSSC) of the Committee on Science (CoS), of the National Science and Technology Council to include broader representation of the Federal agencies that have activities relevant to NPGI's goals, and it was charged with developing a strategic plan for the next five years. This plan contains recommendations for Federal agencies regarding open access, data sharing, and interoperability of tools and databases to improve the practice of agriculture and reduce the demands on environmental resources while addressing challenges posed by global climate change.

The NPGI, from its inception, has operated under the principle that all Federally funded research resources—including data, software, germplasm, and other biological materials and research tools—should be openly accessible in a timely manner. Through this commitment to public access, the NPGI's accomplishments have led to a wealth of publicly available resources that have advanced research in the public and private sectors. The development of new technologies, some through NPGI investments, has greatly accelerated the pace of discovery and data generation. Also, as sequencing technologies become more advanced, genome-scale datasets can be gathered at high throughput for any crop. This wealth of data is revolutionizing the ways in which plant scientists perform research, but the absence of standards governing data generation and management for all common data types is hampering data-mining efforts. The standardization of data, including metadata such as experimental and environmental conditions, is critical for the correct interpretation of results, and, without coordination at the national and international levels, the full value of these resources will not be realized. In addition, the IWGPG is cognizant of the continuing cyberinfrastructure developments in biology and the continued need for interoperability. In this new five-year plan for the NPGI, the IWGPG reaffirms its commitment to the principle of open access and contributes to the Administration's efforts to increase public access to the outcomes of Federally funded scientific research.[1]

As the impacts of the genomics revolution continue to ripple across all aspects of plant research, the sheer volume of information available to researchers has become overwhelming. Having access to this information is only part of the challenge for the research community. A need for training also exists at all levels—from undergraduate students to senior-level researchers—in how to access, integrate, and archive genomic data and resources into learning and research. Through the activities outlined in this plan, the NPGI will continue to play a central role in coordinating Federal support of training and leveraging investments across intramural and extramurally funded activities.

---

[1] Memorandum on Science and Technology Priorities for the FY 2015 Budget from OMB and OSTP, www.whitehouse.gov/sites/default/files/microsites/ostp/fy_15_memo_m-13-16.pdf.

II. Major Accomplishments for 2009–2013

The NPGI, in its plan for 2009–2013, solidified its role as global leader and advocate for innovation in plant genomics. It continued to foster interagency coordination and cooperation that ultimately resulted in historic breakthroughs in plant biology through genomics as well as major advances in enhancing global food security, feedstocks for bioenergy, human nutrition, and resilience to climate change through applied genomic technologies for crop improvement.

NPGI investments in four strategic areas have transformed plant research and accelerated innovation in crop breeding:

- A dramatic expansion in genome sequencing of economically important plants and release of associated germplasm;
- Open access to the resulting plant genome sequence information, sequence data, and associated functional data;
- The development of high-performance cyberinfrastructure and software systems for sequence data analysis and trait analysis and utilization; and
- The development of tools and technologies to accelerate trait prediction and plant breeding at a genome-wide scale.

As a consequence of NPGI open data and open resource policies, the outcomes of its investments have been leveraged by university, government, non-governmental organizations (NGOs), and industry (Table 1). Major NPGI partners in generating data and providing open data access include the USDA National Plant Germplasm System (NPGS), Germplasm Resources Information Network (GRIN), and crop genome database system; the NSF iPlant Collaborative (iPlant) and Extreme Science and Engineering Discovery Environment (XSEDE); the DOE Systems Biology KnowledgeBase (KBase), Joint Genome Institute (JGI), and Bioenergy Research Centers (BRCs); the NIH National Center for Biotechnology Information (NCBI) and The Genome Institute at Washington University (TGI); Cold Spring Harbor Laboratories (CSHL), the National Center for Genome Resources (NCGR); Iowa State University; Washington State University; the European Union (EU) European Bioinformatics Institute (EBI); and the International Maize and Wheat Improvement Center (CIMMYT).

Available data on numbers of students and other trainees, scientific publications, cultivar and germplasm releases, software releases, and patent applications resulting from NPGI funding were collected from member agencies and compiled to provide an estimation of measurable outcomes (Table 2). Approximately 4,500 undergraduate and graduate students and postdoctoral researchers have received training in plant genomics, as have elementary/middle school teachers who will help cultivate the next generation of scientists. Nearly 6,500 peer-reviewed papers were published between 2009 and 2013, many in high-impact journals such as *Science*, *Cell*, *Genome Research*, *Journal of Cell Biology*, *Nature*, *Proceedings of the National Academy of Sciences*, and *PLoS Biology*, ensuring a broad dissemination of results across a diverse scientific community. NPGI investments have contributed significantly

Table 1. U.S. and International Initiatives Benefitting from NPGI Open Data and Open Genetic Resources

Initiative or Program	Participants
Feed The Future	U.S. Government Initiative[2]
Open Data for Agriculture	G8 countries[3]
Basic Research to Enable Agricultural Development (BREAD) Program	NSF, Bill & Melinda Gates Foundation[4]
Germplasm Enhancement of Maize (GEM)	USDA, seed industry partners[5]
Coordinated Agricultural Project (CAP) Program	USDA[6]
Consultative Group on International Agricultural Research (CGIAR) Cooperative Research Programs	USAID, Bill & Melinda Gates Foundation[7]
Plant Feedstock Genomics for Bioenergy	DOE and USDA[8]
Plants Engineered to Replace Oil (PETRO)	DOE Advanced Research Projects Agency – Energy (ARPA-E)[9]
Challenge Program HarvestPlus	CGIAR, Bill & Melinda Gates Foundation[10]
Durable Rust Resistance in Wheat (DRRW)	U.K. Department for International Development (DFID), Bill & Melinda Gates Foundation[11]
EU-US Biotechnology Task Force	NSF, USDA, NIH, DOE, Environmental Protection Agency (EPA), National Oceanic and Atmospheric Administration (NOAA), European Commission[12]
EU Trans-national Infrastructure for Plant Genomic Science (transPLANT)	EU Framework 7[13]

to the public germplasm collections, including improved bean, potato, tomato, wheat, and barley cultivars and hundreds of improved lines or populations. Approximately 1,000,000 plant accessions have been distributed. At least 19 scientific software packages have been released and are publicly available, and nearly 90 patent applications have been submitted, indicating the economic potential resulting from NPGI-supported research.

---

[2] www.feedthefuture.gov.

[3] https://sites.google.com/site/g8opendataconference/home.

[4] www.nsf.gov/pubs/2013/nsf13035/nsf13035.jsp.

[5] www.public.iastate.edu/~usda-gem.

[6] The CAP is a type of Research, Education, Extension, or Integrated Project supporting large-scale, multi-function team projects that conduct targeted research, education, and/or extension in response to emerging or priority area(s) of national need.

[7] www.cgiar.org/our-research/cgiar-research-programs.

[8] DOE and USDA Biomass Genomics Research http://genomicscience.energy.gov/research/DOEUSDA.

[9] http://arpa-e.energy.gov/?q=arpa-e-programs/petro.

[10] www.harvestplus.org.

[11] http://wheatrust.cornell.edu.

[12] http://ec.europa.eu/research/biotechnology/eu-us-task-force/index_en.cfm.

[13] www.transplantdb.eu.

Table 2. Quantitative Analysis of the NPGI's Major Accomplishments 2009–2013

TRAINING[14]	
Postdoctoral researchers	1,069
Graduate students	1,129
Undergraduates	2,322
Elementary and middle school teachers	4
Total	4,524
PRODUCTS (OUTPUT)[15]	
Plant accessions distributed	1,000,000
New cultivar/germplasm/population released	485
Software releases	19
Patent applications	89
Publications	6,430
High impact[16]	242

New and improved cyberinfrastructure and computational systems supported by the NPGI have significantly facilitated the analysis and interpretation of "-omics" and associated metadata by providing the research community an integrative computing environment and greater access to data and analytical tools.

The DOE JGI addresses key challenges in plant genomics, providing the research community access to the latest generation of genome sequencing and analysis capabilities. With 1,200 visitors per day and 120,000 data downloads per year, the DOE-JGI's plant computational resources have had a huge impact in plant-genomics research. The DOE Systems Biology KBase enables users to integrate, analyze, and model data to design new experiments. Although KBase is still under development, it has already issued 8,000 access credentials, including 100 active users and 150 developers and code contributors.

NSF's iPlant is a community-driven cyberinfrastructure that provides a venue for researchers to access and query diverse datasets through user-friendly tools and resources. iPlant provides a broad range of web-based services, including Atmosphere cloud resources, the Discovery Environment, and the DNA Subway, and various training events. With total data storage capacity of 427 terabytes, iPlant executes approximately 18,300 jobs monthly for 360 to 610 active users and provides new registrations at the rate of 450 to 500 per month, although these numbers do not reflect the growing number of stand-alone projects such as the Comparative Genomics (CoGe) platform.

The USDA plant genome and genetic database system, including the Maize Genetics and Genomics Database (MaizeGDB), SoyBase, Gramene, Legume Information System (LIS), and GrainGenes, has 3,000 visitors daily to its sites, and the ARS/NSF Trait Analysis by Association,

---

[14] Compiled from USDA Agricultural and Food Research Initiative (AFRI), Agricultural Research Service ARS, DOE Office of Science (SC), and NSF data.

[15] Compiled from USDA AFRI, ARS, DOE SC, U.S. Forest Service (USFS) data.

[16] Impact factor ≥ 9.0.

Evolution, and Linkage (TASSEL) software platform counts 3,000 active users and 20,000 downloads annually.

The accomplishments made from 2009–2013 (summarized in Figure 1) illustrate how the NPGI has democratized plant research, enhanced the utilization of knowledge and resources generated through NPGI public and private partnerships, and dramatically accelerated domestic and global crop improvement. Detailed descriptions of the advances made under each objective can be found in [Appendix B](#).

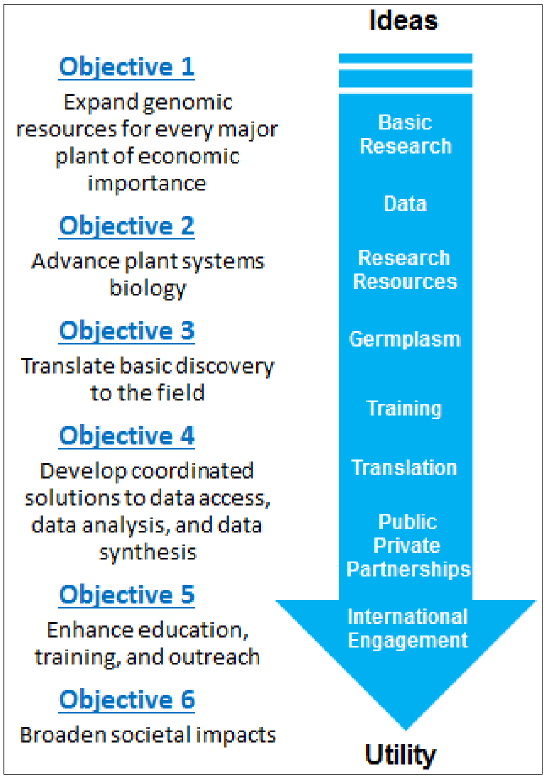

Figure 1. [Major Accomplishments for 2009–2013](#)

III. Plan for 2014–2018

The NPGI’s goal, per the IWGPG Charter (Appendix A) is to develop a basic knowledge of the structures and functions of plant genomes and to translate this knowledge into a comprehensive understanding of all aspects of economically important plants and plant processes of potential economic value. By bridging basic research and plant performance in the field, the NPGI seeks to accelerate basic discovery and innovation in economically important plants to improve the practice of agriculture, reduce the demands on environmental resources, and address challenges posed by global climate change.

Plant genomic research will play a critical role in addressing the needs of a growing world population and the environmental effects of climate change. In the United States, solutions to the challenges of reliably producing sufficient food and a sustainable supply of renewable energy on a shrinking arable land area will also rely heavily on the outcomes of basic research and their translation into the field. Since the publication of the National Plant Genome Initiative Five-Year Plan: 2009–2013, considerable progress has been made toward achieving the plan’s stated goal of expanding genomic resources for every economically important plant. As new DNA sequencing technologies and other new genomics technologies accelerate the production of large-scale genomic datasets, including transcriptomes, proteomes, and metabolomes, the challenges of managing these data are strategically important. While the wealth of data already available will revolutionize the way in which plant scientists perform research, standards for data generation and collection are incomplete, and this lack of standardization significantly affects efforts to mine available datasets to derive reliable and accurate information.

Agricultural productivity in the 21st century will face unprecedented challenges due to the changing global climate, including unstable patterns of precipitation and temperature, intensified biotic and abiotic stressors, and soil erosion.[17] Growing enough food, fuel, and fiber to accommodate a growing world population in a changing climate without further compromising the environment presents complications in the form of increased fertilizer use and subsequent runoff, increased pesticide/ herbicide use, and loss of biodiversity.[18] These issues are not independent of each other. Trends toward decreasing crop yields due to increased drought, heat, and pathogen stress may motivate the use of mitigating strategies that include higher inputs of increasingly scarce resources and pesticides. A more complete understanding of plant response to the surrounding environment, particularly the molecular and genetic mechanisms that sense and regulate this response, is critically needed to accelerate development of widely adaptable crops. With increased drought tolerance, nutrient-use efficiency, and resistance/tolerance to pathogens and insects, these improved varieties, ideally, will tolerate and even thrive under suboptimal conditions and require few inputs.

---

[17] C. L. Walthall, J. Hatfield, P. Backlund, L. Lengnick, E. Marshall, M. Walsh, S. Adkins et al., *Climate Change and Agriculture in the United States: Effects and Adaptation*, USDA Technical Bulletin 1935 (Washington, D.C.: United States Department of Agriculture, 2012), www.usda.gov/oce/climate_change/effects_2012/effects_agriculture.htm.

[18] Executive Office of the President, President’s Council of Advisors on Science and Technology, *Report to the President on Agricultural Preparedness and the Agriculture Research Enterprise* (Washington, D.C., December 2012), www.whitehouse.gov/sites/default/files/microsites/ostp/pcast_agriculture_20121207.pdf.

The NPGI, as prescribed in its Charter, promotes the use of existing and novel genomic technologies to address these issues of strengthening cropping efficiency and yields to promote food security while reducing environmental impacts. Such approaches are well suited to investigate the very complex relationship between plant genotype and the surrounding environment, including biotic (e.g., the numerous organisms that share the environment) and abiotic variables. A more complete understanding of these interactions will result from studies of whole ecosystems under multiple environments, enabled by a computational infrastructure and tools that allow integrative analysis of "-omics" and associated metadata and the development of models that will predict how various genotypes perform under specific environments. Such models will take into account the natural variation that allows plants to adapt to a changing world, including intercellular and intracellular communication, regulatory mechanisms, and metabolic and developmental pathways. This knowledge can then be manipulated to facilitate the development of more resilient crops, either through traditional breeding or targeted biological design.

In the coming years, standardization of data, including metadata such as experimental and environmental information associated with genomic datasets, will be essential if the full value of these data is to be realized. This problem spans all scales of the research endeavor, from local to global. Data generators and end users will need to be engaged in the development of the standards to which all should adhere, and coordination at national and international levels will be needed to ensure that all tools and resources are of maximum utility and known quality to avoid overlap and redundancy of effort. In the development of this new five-year plan, the IWGPG was tasked in the Charter to engage the plant science community in setting priorities for genomics tools and resources, defining new and sustainable strategies to meet community needs and priorities, advancing biological innovation and breakthrough discovery, and improving coordination among Federal agencies and international plant genomics partners.

The IWGPG, in accomplishing its goals, undertook an inventory of the current plant genomics tools and resources, including but not limited to analytical tools, genetic tools, sequences, and databases as prescribed in the Charter. Strategic research needs and resource gaps for the development of sustainable systems for food, bioenergy, and industrial feedstock production were identified by considering the

- Minimization of inputs such as water, energy, pesticides, and fertilizer;
- Effects of climate change and increased agricultural productivity;
- Minimization of environmental impacts using plant genomics;
- Opportunities for Federal agency coordination and cooperation and for public/private partnerships; and
- Associated opportunities to enhance training, education, and public outreach.

A wide range of stakeholder input was used in generating this plan, including stakeholder reports and white papers, stakeholder workshop discussions, an open workshop held at the Plant and

Animal Genome Conference XXI in San Diego, California, January 12, 2013, and comments received via a wiki site.[19]

The objectives of the NPGI Five Year Plan: 2014–2018 (diagrammed in Figure 2 and described in detail below) focus on open access, data sharing, and interoperability of tools and databases to improve the practice of agriculture, reduce the demands on environmental resources, and address challenges posed by global climate change.

Objective 1:	Develop a new generation of databases and tools to enable every step in the innovation chain from basic discovery to translation into practical application and widespread use.
Objective 2:	Create a network of plant germplasm resources to meet the Nation's research and breeding needs for the 21st century
Objective 3:	Build tools to advance our understanding of plant biology and translate this knowledge into precision plant breeding for the development of sustainable systems for food, bioenergy, and industrial feedstock production
Objective 4:	Empower the workforce to use a new generation of tools and resources
Objective 5:	Build public-private partnerships to advance the translation of basic discoveries into practical application and support innovation in the practice of agriculture
Objective 6:	Strengthen international partnerships to bring the benefits of new discoveries to all

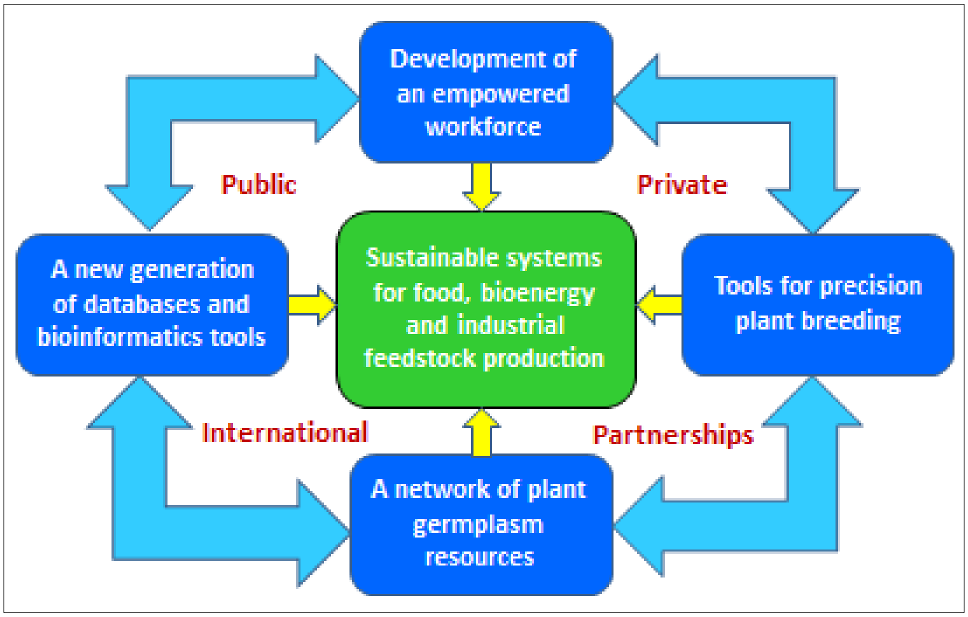

Figure 2. U.S. Federal Agencies, Public, and Private Partners Will Continue to Leverage Outcomes from NPGI Objectives

---

[19] https://extwiki.nsf.gov/x/DQAOAQ.

In the coming five years, the NPGI will incorporate new metrics for measuring the impacts of its investments on such areas as use of research resources and data, training outcomes, and downstream economic impacts. Collection of these datasets will be facilitated by the project reporting systems at research.gov but will require additional investments to collect and analyze the data. In addition to providing *quantitative measures* of the impacts of NPGI investments, this information will be valuable in determining priorities for future investments.

Objective 1: Develop a new generation of databases and tools to enable every step in the innovation chain from basic discovery to translation into practical application and widespread use.

Build the next-generation plant database system

Advances in genomic, genetic, and molecular technologies have revolutionized plant research and led to a rapid increase in the size and complexity of plant datasets. The management, analysis, and interpretation of data are now priority targets for innovative solutions in plant genomics research supported by the NPGI.

Diverse data repositories contain information important to understanding the genetic and molecular control of crop traits (e.g., weather, production management, and disease statistics). Researchers from multiple research locations and disciplines must be able to find, access, query, analyze, and effectively visualize these resources.

Researchers and breeders require a new generation of centralized access points to the genomic, genotypic, and phenotypic data stored at multiple databases. NPGI-supported crop genetic and genomic databases will also serve as “portals” that provide open access to these data and to the latest bioinformatics tools, workflows, and analytical methods available. These portals will also provide greater access to distributed high performance computing infrastructure supported by the NSF and the DOE (e.g., iPlant, Extreme Science and Engineering Discovery Environment (XSEDE), and KBase) necessary for analyzing large datasets.

Through funded projects, the NPGI will enable researchers and those responsible for data management to more effectively cooperate and coordinate in creating environments that better enable user interaction with diverse datasets through data discovery and exchange. New community standards, ontologies, nomenclatures, data representations, and protocols that allow systems to interoperate will enhance the user’s ability to compare and analyze data in new and meaningful ways.

With NPGI support, improved strategies and methods for analyzing large-scale genomic data, high-throughput phenotype or an observable characteristic or trait that describes an organism. data, and environmental and experimental data will be explored. Rigorously applied data standards will enable interdisciplinary collaborations that are needed to properly catalog and to integrate the array of data types. Integrative analyses will need to be combined with high-throughput data production.

Improvements in crop genome databases and analytical tools and the establishment of data standards and ontologies will enable researchers and breeders to access and use data to understand how genotype and environment determine phenotype—critical information needed to rapidly and efficiently design superior crops.

Computational modeling of plant metabolic, developmental, and genetic processes is gaining momentum. The NPGI will focus support on the development of technologies for modeling, analyzing, and visualizing metabolic and regulatory networks that will lead to a predictive understanding of the vast biological complexity of plant systems.

To improve the mechanistic understanding of how plant genomes function and to enhance the utility of crop sequence information for crop improvement, the NPGI will continue to support efforts to improve genome analysis and sequence annotation. For example, the Plant Encyclopedia of DNA Elements (pENCODE) is an important new step toward that goal. The primary goal of pENCODE will be to build a comprehensive parts list of functional elements for the genomes of major crops, including elements that act at the protein and RNA levels and regulatory elements that control cells and the circumstances in which a gene is active.

Enhance tools for data visualization and analysis

The NPGI will support the increase in interoperability between crop genome database systems, thereby enhancing comparative genomics across species and leveraging data and knowledge between related plant species and between distantly related species. In addition, the NPGI will support the incorporation of new tools for visualizing and interrogating increasingly diverse sets of genomic and experimental data.

Interoperability between crop and pathogen genome databases will also be needed as breeders search for effective and durable resistance genes. Rapid evolution of plant pathogens has made the search for durable resistance to plant disease a perennial challenge. The “red queen” analogy is often used to describe the endless search for disease resistance in crop plants. Comparative genomics across plant and pathogen boundaries are needed to ensure sustainable disease resistance in crop plants.

The NPGI will support outreach and training to expand access to resources that enable users to share and publish their data, develop tools and workflows for analyzing large complex datasets, and increase long-term sustainable, actively curated, interconnected resources that support plant genetics, genomics, and plant breeding.

Increase access to high-throughput genotyping

Advances in DNA sequencing technologies have revolutionized the production of DNA sequence information, including plant genotyping. The NPGI will continue to support the development of new tools and resources for genotyping an expanding portfolio of crops for food, feed, fuel, fiber, and renewable energy production.

Until very recently, researchers working in many crops lacked access to sufficient sequencing capacity and bioinformatics support to generate enough genetic markers for mapping and analyzing priority traits. The NPGI’s support for genotyping technologies and bioinformatics will facilitate the production and characterization of valuable resources: DNA sequence-based markers, high-resolution genetic maps, and genetic stocks and exotic germplasm that will enhance gene discovery and cultivar improvement for major and minor crops.

With support from the NPGI, advances in high-throughput genotyping and the analysis of very large datasets will enable scientists to understand how alleles can combine synergistically and how best to select for these combinations, rather than selecting individual genes for their average effects.

Expand field-based phenotyping

In contrast to genotyping technologies, trait measurement is now often the critical bottleneck to genetic analysis and breeding progress. Thus, the NPGI will support the development of innovative high-throughput, semi-automated, and automated field-based phenotyping methods.

Improving the efficiency and accuracy of phenotypic evaluations will enable larger populations of plants to be assayed over more environments, thereby improving the power of gene function analysis. Phenotypic and genetic data from large population studies will be combined in a coordinated effort to develop improved models for predicting the performance of untested materials.

Phenotyping for plant diseases across environments will require sophisticated measures for disease detection and molecular analyses. Phenotypic evaluations of crop plants for disease will require coordinated efforts in pathogen monitoring and tracking so that field testing is done in appropriate regions.

Promote genome engineering toolbox development and utilization

The NPGI will continue to support the analysis of plant gene function by building on its toolbox of technologies for manipulating plant gene activity, including RNA interference, misexpression, and genome editing using artificial restriction endonucleases (e.g., transcription activator-like effector nucleases (TALENs)). Efficient pipelines for high-throughput crop transformation and the targeted manipulation of gene activity are needed to test novel hypotheses. New technologies for plant genome engineering will be optimized for the application to plants (e.g., Multiplex Automated Genome Engineering (MAGE)[20] and the clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9.[21]

Enhance tools for biochemical screening

More efficient and effective biochemical screens, along with enhanced genotyping and phenotyping methods, superior functional genetic analyses, and the capacity to extend findings

---

[20] H. Wang, H. Yang, C. S. Shivalila, M. M. Dawlaty, A. W. Cheng, F. Zhang, and R. Jaenisch, "One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering," *Cell* 153, no. 4 (2013): 910–918. www.sciencedirect.com/science/article/pii/S0092867413004674.

[21] K. Belhaj, A. Chaparro-Garcia, S. Kamoun, and V. Nekrasov, "Plant genome Editing Made Easy: Targeted Mutagenesis in Model and Crop Plants Using the CRISPR/Cas System," *Plant Methods* 9 (2013): 39–49. www.plantmethods.com/content/9/1/39.

from model plants to crops, will accelerate gene discovery and translate into cost-effective breeding tools.

Objective 2: Create a network of plant germplasm resources to meet the Nation's research and breeding needs for the 21st century

The ready availability of plant genome sequences, improved technologies for "resequencing" genomes and for identifying genetic markers, and novel statistical genetic approaches (e.g., genome-wide association analyses) converged during the last five years to enable plant researchers and breeders to achieve unparalleled advances in research and genetic improvement across a broad spectrum of major and specialty crops. This NPGI-supported research also expanded substantially the demand for genetic stocks and other genetic resources conserved in U.S. NPGS genebanks. Not only did the NPGS distribute record numbers of samples—more than one million—during the last five years, it also became home to valuable new genetic stocks and structured populations generated by NPGI-supported research. These important resources require long-term stewardship to safeguard their integrity for breeding and research. The NPGS genebanks not only conserve plant genetic resources, but also generate, preserve, and deliver associated descriptive genotypic and phenotypic data. The demand for those data has similarly expanded substantially during the last five years. The following proposed goals constitute a mixture of tried-and-true approaches and innovative steps for creating an enhanced network of plant genetic resources needed to underpin the entire NPGI enterprise.

Expand new populations for genomic research, trait analyses, and crop improvement

NPGI researchers will continue to generate and analyze new, specially-structured experimental populations (e.g., Nested Association Mapping (NAM)[22] and Multiparent Advanced Generation Intercross (MAGIC))[23] that reveal the genetic and molecular bases for key target traits and will continue to exploit the crop diversity in farmer varieties, wild crop relatives, older varieties, and elite breeding stocks in the NPGS genebank collections. Additional genotypic and phenotypic information for these populations will generate new testable hypotheses and insights for plant breeding methods that may accelerate rates of genetic gain and yield genetically enhanced populations improved for valuable agricultural traits, many of which are complex and quantitatively inherited. Accumulating genotypic and phenotypic data will reveal in the current NPGS genebank collections the genetic gaps that will be filled by additional field explorations (especially for wild crop relatives) and inter-institutional exchanges of genetic stocks and improved germplasm.

---

[22] M. McMullen, S. Kresovich, H. S. Villeda, P. Bradbury, H. Li, Q. Sun, S. Flint-Garcia, et al., "Genetic Properties of the Maize Nested Association Mapping Population," *Science* 325, no. 5941 (2009): 737–740., www.sciencemag.org/content/325/5941/737.full.

[23] P. X. Kover, W. Valdar, J. Trakalo, N. Scarelli, I. M. Ehrenreich, M. D. Purugganan, C. Durant, and R. Mott, "Multiparent Advanced Generation Inter-Cross to Fine-Map Quantitative Traits in *Arabidopsis thaliana*," *PLoS Genetics* 5, no. 7 (2009): e1000551, doi:10.1371/journal.pgen.1000551.

Support effective stewardship of genetic stocks and genetic diversity for research and breeding

For more than 60 years, NPGS genebanks have safeguarded, regenerated, and distributed—free of charge and restriction—genetic stocks for major crops and for the genetic diversity included in more than 14,700 species. NPGI support for open access to those diverse genetic resources and information has been critical to the progress in understanding how plant genomes function, in supporting plant breeding that enhances global food security and human nutrition, crop resilience and adaptation to global change, and in increasing production and conversion of biomass for biofuels and fiber. Without ready access to that germplasm, powerful genomic, phenomic, and bioinformatic tools cannot be applied effectively to attain the NPGI's goals.

NPGI-supported research has developed genetic stocks of mutant resources for the genetic analysis of crop and model plant systems. By combining classic mutagenesis strategies with contemporary rapid and inexpensive sequencing and genotyping approaches, geneticists can now "sequence-index" chemically induced and ionizing radiation-induced mutant stocks. The NPGS will require expanded curatorial capacity to maintain and distribute those mutant stocks and mutagenized populations (including insertion mutants and misexpression lines) so important for analyzing gene function through reverse genetics.

Maintaining and distributing large volumes of new genetic stocks and structured populations (e.g., recombinant inbred lines and mapping populations) have challenged the curatorial capacity of NPGS genebanks. During the next five years, the NPGI will need to identify new sources of support for those genebanks and explore new models and mechanisms for inter-institutional and inter-sector collaboration that enable genebanks to safeguard these valuable materials and continue to make them widely available.

Culture collections of plant-associated microbes represent an essential resource for NPGI researchers. Microbial collections play diverse and critical roles in understanding plant resistance to diseases. These publicly held resources are critical in developing strategies to control plant diseases that impact the vitality of U.S. agriculture, represent an important link between past and present disease epidemics, facilitate the identification of emerging diseases, and provide data that are essential for forensic investigations. The expanded curation of important plant microbes that can be useful for the NPGI is urgently needed.

Develop new phenotyped and sequence-indexed genetic resources

To exploit the genetic potential in genetic resources as revealed by DNA sequencing, more and higher quality phenotypic data must be generated and recorded for genebank accessions. NPGI researchers will generate phenotypic data via evaluations in multiple test environments according to standards that meet the needs of geneticists, breeders, and genebank managers. The NPGS's Crop Germplasm Committees, including breeders and other crop experts, can be enlisted to devise these standards for phenotyping. Researchers at genebanks and in crop genetics or breeding programs will require expanded capacity to conduct such evaluations.

High-quality evaluation data, when available, will enable NPGI researchers to combine that information with genotypic data to genetically dissect complex phenotypes (e.g., yield and adaptation) and discover and verify key allelic variants. Breeders can then recombine and select those variants and improve breeding populations. As research capacity and accumulated knowledge permit, genomic selection and related breeding by prediction methods will be extended and evaluated empirically for more complex genetic architectures (e.g., polyploids), subpopulation structures, and a range of specialty and other crops that until now have lacked these genetic tools.

Promote interlinked, web-based information resources to facilitate wider germplasm use

Effectively incorporating the genetic potential from non-elite sources such as landrace or domesticated varieties and wild crop relatives requires not only better information about their agronomic value and genetic structure, but also the means to retrieve, interrelate, visualize, and analyze diverse data (genotypic, phenotypic, taxonomic, and ecogeographical) from a common computational environment. Building from and enhancing existing information management systems such as GRIN-Global (for genebank accession data) and crop genome databases (Gramene, MaizeGDB, Soybase, LIS, and GrainGenes) during the next five years, the NPGI will develop an information management environment that will enable phenotypic, genotypic, genomic, and genetic resource data to be cross referenced. Ultimately, the integrated datasets will serve as predictive tools to enable plant breeders to more rapidly assess agronomic merit based on evaluation and genomic data for key traits and to apply genome-wide association analyses and genome-based selection.

Continued NPGI support for open access to diverse genetic resources and information is critical for sustaining progress in understanding how plant genomes function and for crop improvement in support of global food security, enhancing resilience and adaptation to global climate change, increasing the production and conversion of biomass for biofuels and fiber, and advancing human nutrition as outlined in the IWGPG Charter.

Objective 3: Build tools to advance our understanding of plant biology and translate this knowledge into precision plant breeding for the development of sustainable systems for food, bioenergy, and industrial feedstock production

An important goal of the NPGI-supported research is to generate new knowledge that enables breeders to develop improved varieties of crops that are able to adapt to a greater range of weather extremes and associated stresses. Understanding how genetic diversity can be employed through genomics to enable plants to respond and adapt to abiotic and biotic environmental challenges will facilitate tremendous applications for agriculture.

Develop a lifecycle systems view of plant growth and development from cells to plants

In the past five years, the NPGI has supported projects that developed a systems view of gene networks at the level of an organ or a seed but have yet to integrate those data at the whole-plant level. To fully determine the function of a gene, researchers need to understand its role in a broader, whole-plant context. Technological advances in proteomics and metabolomics will be critical toward achieving the goal of a systems-level understanding of plant processes that will enable plant modification for the development of improved crops with desired traits.

Plant productivity is governed by the capacity of leaves to assimilate carbon and nitrogen, the capacity of sink tissues to utilize photosynthate for growth and storage, and the ability to tolerate environmental stresses. Climate change—rising global temperatures, more frequent and extreme droughts, tropospheric ozone, and climate-associated change in disease and pests—threatens plant productivity. To accelerate breeding gains and enhance plant genetic adaptability, a systems-level understanding of plant growth and development at the genome level must be sought to enable breeders to avoid negative alleles and increase the frequency of positive alleles.

Develop next-generation epigenomics toward understanding the epigenomic code

Significant advances in our understanding of epigenetic mechanisms are expected in the next five years, which could lead to an elucidation of the epigenomic code–epigenetic modifications of a genome at a genome-wide scale. Genome-wide analysis of histone and DNA modifications at single-base resolution are currently available, and new and improved technologies are being developed that could greatly expand the types of information collected and its rate of collection. Single-molecule DNA sequencing technologies hold the promise of producing DNA modification information as the DNA sequence is being generated. The resources and knowledge obtained will allow researchers to achieve a comprehensive understanding of the regulatory networks that control the molecular mechanisms that result in chemical changes in DNA and chromatin proteins. As a consequence, it will be possible to manipulate the stability and effects of epigenetic phenomena on plant phenotype. The ability to control epigenetic inheritance will enable researchers to take advantage of epigenetic mutations to advance breeding programs for crop improvement.

Epigenetic modifications are also responsible, in part, for silencing the foreign DNA introduced into plant genomes through genetic transformation techniques. The ability to control the regulation of epigenetic mechanisms will facilitate plant genome engineering in the design of new and improved plant varieties.

Improve our understanding of plant mineral metabolism, including nitrogen- and phosphorus-use efficiency

Plants respond physiologically to nutrient availability in the soil, ranging from toxic levels of metals to insufficient nutrients and water. Nitrogen (N) and phosphorus (P), acquired through the roots, are critical for crop growth and are the two most costly nutrients in U.S. agriculture. Efforts on an international scale will be required to fully understand the points at which plant improvement can be used to address local constraints in increasing plant yield caused by deficiencies in these two key nutrients. New knowledge of the genes and mechanisms that enable crop plants to adapt to extreme nutritional challenges should lead to higher crop productivity in unfavorable growth environments and will improve the resiliency of agriculture to climate change while promoting sustainable intensification of production.

Develop a deeper understanding of metabolic partitioning

Many metabolites are subject to constant turnover. To fully understand function, we need to know not only the type and quantity of the metabolites present, but also how these metabolites are being used. Better tools and models need to be developed to determine and quantify metabolic fluxes (fluxomics), including the ability to measure absolute and relative abundances.

There is a need to develop more advanced techniques to assess signal molecule and carbon and nutrient transport processes that determine the allocation of signals and resources to different parts of the plant. Hundreds of genes that encode transporters of unknown substances appear to be involved in intra- or intercellular or vascular transport. Tools are needed to study transport and allocation processes, and computational methods are needed to integrate transport and metabolic models.

Taking advantage of powerful synchrotron light sources available at user facilities, new ultra-high-resolution technologies will be developed for real-time element and metabolite imaging in living cells and at the interface between plants and their associated microbial communities.

Expand our understanding of plant microbe interactions to better inform system-level approaches to plant breeding

Researchers are only now beginning to understand how microbes in the soil influence plant development, growth, and adaptation to a changing environment. Knowing how communities of beneficial microorganisms (bacterial and fungal) in the rhizosphere and endosphere interact with plant roots will be critically important to the future development of sustainable agricultural systems. These complex relationships involve various types of intercellular communications, and the extent of the response varies with the genotype of the plant and the microorganism. Advances in imaging and metagenomics will provide needed new technologies to improve our understanding of plant-microbe interactions at the molecular level. New genome-level research is needed to understand mechanistically how plant defenses adapt to continuously evolving pathogen virulence factors.

Develop new crop models to integrate variation, physiology, and environmental information toward predictive modeling

The development of statistical networks and models that incorporate phenotypic, "-omic," and associated metadata will enable a fuller understanding of plant processes, the influence of environmental interactions on these processes, and the ability to predict phenotype from genotype in the field and in multiple environments. As articulated in the American Society of Plant Biologists (ASPB) Plant Science Research Summit report,[24] a more complete understanding of gene regulatory processes will allow the prediction of plant adaptation under different environments. This capability will be enabled through enhanced data modeling and through analysis and interpretation and will require a stable data management infrastructure.

---

[24] American Society of Plant Biologists, *Unleashing a Decade of Innovation: A Vision for 2015–2025* (Rockville, MD, 2013), http://plantsummit.wordpress.com/summit-report.

Develop and deploy new tools and methods for high throughput phenotyping

A strong demand exists across the plant research and breeding community for new high-throughput phenotyping tools to enable large-scale field trials for agronomically meaningful traits. Biochemical and physiological assays typically require destructive sampling of plant tissues. New tools and methods—including sensing, imaging, and automation—for the non-destructive high-throughput measurement of complex physiological traits in the field will need to be developed. When these tools are combined with low-cost high-throughput genotyping, they will revolutionize the pace at which new discoveries in plant research can be deployed in fields across multiple environments and locations.

Analyze and evaluate genetic diversity across diverse environments

Genetic diversity supplies the raw material for plant breeders. NPGI objectives are designed to generate new knowledge for understanding (1) how genetic variation performs in specific environments, (2) how plant phenotype is determined by genetic and environmental variation, (3) how to develop and validate crop models that integrate genetic variation, physiology, and environments, and (4) how to develop genome-wide predictions that include biological and physiological mechanisms. This knowledge will allow breeders to utilize all available germplasm in developing crops that are widely adaptable in a broad range of environments and locations and to predict how specific combinations of genes and environments will perform in the field.

Improve our understanding of gene and protein functions

Although the complete genomes of many important crop plants have been sequenced and annotated in the last five years (Appendix D), the determination of gene function in these genomes is far from complete. For example, currently 30%–34% of the *Arabidopsis* genome is annotated as "unknown" with respect to Gene Ontology (GO) function.[25] New approaches need to be developed to determine the function of such unknown and hypothetical genes and, importantly, to translate this information to other plant species that have incomplete or no genome sequence. Comparative genomic analyses can often be used to determine the function of conserved genes, but clearly defining taxon-specific pathways and processes will require a better understanding of the plant's physiology and biochemistry. Research at the protein and metabolite level will help clarify the role of currently unknown genes through direct determination of function. Coordinated studies of genes, proteins, and metabolites will be necessary to more completely understand whole-system functions and proteomic/metabolomic capabilities (e.g., high-throughput biochemical and/or transport assays) to determine the substrates of proteins that need to be developed and improved to keep pace with the advances in genomics. A toolkit for 21st century plant functional genomics will include an improved gene expression atlas, near-isogenic lines (NILs) for mapping and identifying complex trait genes, new mutation/modification methodologies to associate genes with phenotype, and a better understanding of regulatory networks through integrated transcriptomic, proteomic, and metabolomic datasets.[12]

---

[25] P. Lamesh, T. Z. Beradini, D. Li, D. Swarbeck, C. Wilks, R. Sasidharan, R. Muller, et al., "The Arabidopsis Information Resource (TAIR): Improved Gene Annotation and New Tools," *Nucleic Acids Research* 40 (2012): D1202–D1210. www.ncbi.nlm.nih.gov/pmc/articles/PMC3245047.

Objective 4: Empower the workforce to use a new generation of tools and resources

Advances in genomic technologies now allow plant genomes to be profiled at extremely high resolution and at an increasingly affordable cost. When these data are combined with advances in the genetic and molecular understanding of plant traits and how plants systems are constructed, the fusion of crop breeding with crop engineering will begin. A significant challenge for the next five years will be to train a new workforce and give this workforce the knowledge, analytical tools, and super computing resources necessary to analyze a new generation of experiments that contain trillions of data points.

Expand support for cross-disciplinary graduate training programs

The NPGI will continue to support the training of the next generation of plant scientists at all levels. Its support will provide for cross-disciplinary postdoctoral and graduate training programs with longer and larger stipends. In addition, the NPGI will continue to support programs like the NSF Mid-Career Investigator Awards in Plant Genome Research (MCA-PGR) to provide for the training and increased participation of established mid-career investigators primarily trained in fields other than plant genomics.

Train a new generation of plant scientists and breeders to design and invest in next-generation database systems

iPlant, KBase, and the national system of crop genome and genetic databases will empower the new generation of plant scientists and breeders.

iPlant. iPlant will enable a new generation of scientists to use high-performance cyberinfrastructure to solve computationally intensive genetic and breeding problems through education, outreach, and training (EOT). Examples include seminars and "road shows" that introduce iPlant tools and services to potential users at research and academic institutions around the country; Data-to-Publication (D2P) Workshops to which researchers, post-docs, and graduate students bring their own data and leave with analyses and insights suitable for publication; hack-a-thons at which super-users and science-savvy developers work in groups to expand iPlant tools and services; and mini-workshops at professional meetings that provide a hands-on introduction to iPlant tools and services. Information on future workshops is available at the iPlant website.[26]

KBase. To accelerate the sharing of genomic data, tools, and scientific information, KBase will provide training via several methods, both in-person and online.[27] This training will include two types of hands-on sessions: boot camps (focused on developing computational tools) and workshops (focused on tackling specific scientific problems). Web-based seminars (webinars) offer instruction on KBase tools and services and are held several times a month.

National crop genome and genetic databases. In partnership with scientists working in the field, public database curators will build on advanced programs for training researchers in the

---

[26] iPlant Events Calendar, www.iplantcollaborative.org/learn#upcoming.

[27] KBase Training Resources, http://kbase.us/about/training/.

latest technologies in plant genomics and breeding. Genotyping by sequencing (GBS), for example, is revolutionizing plant genetics and conventional plant breeding. However, expertise in GBS data analysis is currently limited and thus in high demand by geneticists and breeders in industry and the public sector. With support from the NSF and the USDA, Cornell University hosts a highly successful series of GBS workshops several times per year for the plant and animal research communities.[28,29]

Expand international traineeships for students on NPGI projects

Currently, many federally funded grants do not support assistantships for international graduate students. Many of the future food security issues will involve deploying crops to areas of the world that most likely will be significantly affected by climate change. We need to develop the human and institutional capacity in the developing world and expand partnerships with laboratories in the United States. The Borlaug Higher Education Agriculture Research Development (BHEARD) program funded by the U.S. Agency for International Development (USAID) will provide full graduate assistantships for international students to work on plant-related programs in U.S. laboratories. Expansion of programs such as BHEARD and the USAID-NSF co-funded Partnerships for Enhanced Engagement in Research (PEER) student program could greatly expand opportunities to globalize U.S. research and train international students.

Objective 5: Build public-private partnerships to advance the translation of basic discoveries into practical application and support innovation in the practice of agriculture

The President’s Council of Advisors on Science and Technology (PCAST) report entitled *Report to the President on Agricultural Preparedness and the United States Agricultural Research Enterprise* states that public investment in agricultural research should enhance the public good and target long-term challenges that “cannot be easily monetized in the short term.”[30] In support of expanding public-private relationships, the ASPB Plant Summit Report noted that strong public-private relationships will “increase possibilities for adding discovery value to privately created and held data” and “contribute to the visioning and implementation of new training paradigms” (i.e., training of a new generation of agricultural scientists).[31]

Enhance coordination of public support for basic research and hand-off for downstream translation by companies

The three projects highlighted below show how the investment in public-private partnerships will provide genomic information and genetic resources at the “pre-competitive” level for the benefit

---

[28] Education and Outreach, Institute for Genomic Diversity, Cornell University, www.igd.cornell.edu.

[29] Genotyping-By-Sequencing Workshops, Institute for Genomic Diversity, Cornell University, www.igd.cornell.edu/index.cfm/page/Education/workshops/GBSworkshop.htm.

[30] Executive Office of the President, President’s Council of Advisors on Science and Technology, *Report to the President on Agricultural Preparedness and the Agriculture Research Enterprise* (Washington, D.C., December 2012), 44, www.whitehouse.gov/sites/default/files/microsites/ostp/pcast_agriculture_20121207.pdf.

[31] American Society of Plant Biologists, *Unleashing a Decade of Innovation: A Vision for 2015–2025* (Rockville, MD, 2013), 27, http://plantsummit.wordpress.com/summit-report/.

of the entire research community and simultaneously support the President's Executive Order and Open Data Policy, which require data generated by the government to be made available to entrepreneurs, researchers, and others who can use these data to generate new products and services, build businesses, and create jobs.[32]

Germplasm Enhancement of Maize (GEM)[33] will continue to drive maize breeding forward by linking 35 companies and 20 public breeding programs. Maize breeders want a broad genetic base for their breeding programs, but adapting exotic strains to U.S. conditions is slow and costly. GEM meets the needs of U.S. maize producers by facilitating the integration of exotic material through cooperative public and private-sector research. Genomic data are being incorporated to characterize maize genetic diversity and elucidate the genetic bases of maize adaptation.

The North American Collaborative Oat Research Enterprise (CORE) project exemplifies a highly successful public-private partnership for species with a complex genome and limited resources.[34,35] CORE is a jointly funded project in oat genomics supported by the USDA (National Institute of Food and Agriculture (NIFA), Agricultural Research Service (ARS)), North American Millers' Association, General Mills, and the Prairie Oat Growers Association.

VitisGen[36] is a large, multi-disciplinary, collaborative project dedicated to using genomic technologies to decrease the time and cost involved in developing a new generation of grape varieties that are less vulnerable to weather extremes and disease.

To meet the demands of the growing global population, the demand for wheat in developing countries is projected to increase 60 percent by 2050. While total wheat acres are increasing, global wheat productivity is decreasing, mostly due to increasing average temperatures. Current climate change models suggest that wheat yields could decline by as much as 50 percent by 2050 in South Asia. In one example from many, Arcadia Biosciences has aligned its commercial wheat technology program with USAID and CIMMYT to develop heat-tolerant wheat varieties for South Asia.

New opportunities will be sought that enable public-private-sector collaboration in research involving non-proprietary genetic resources, varieties, and related genomic information. These in-kind resources from the private sector could be leveraged with NPGI-supported research for training and new research that are more relevant to commercial standards. Such partnerships could also strengthen the incorporation of more diverse genetic resources into commercial-quality germplasm.

Encouraging industry collaboration is one route from the public sector to deployment but may invite challenges in handling intellectual property and restrictions that could affect collaborating scientists. Policies that allow access to resources yet protect a company's profitability will need

---

[32] Executive Order on Open Data, www.whitehouse.gov/blog/2013/05/09/landmark-steps-liberate-open-data.

[33] www.public.iastate.edu/~usda-gem.

[34] CORE project, http://wheat.pw.usda.gov/CORE600/chado/analyses.

[35] R. E. Oliver, N. A. Tinker, G. R. Lazo, S. Chao, E. N. Jellen, M. L Carson, H. W. Rines, et al., "SNP Discovery and Chromosome Anchoring Provide the First Physically-Anchored Hexaploid Oat Map and Reveal Synteny with Model Species." *PLoS ONE* 8, no. 3 (2013): e58068, doi:10.1371/journal.pone.0058068.

[36] VitisGen Project, www.vitisgen.org.

to be devised to encourage public-private collaboration, thereby accelerating scientific discovery and commercialization.

Partnership between the public and private sectors will also benefit the next generation of plant scientists. Agricultural, food, and environmental research has a critical need for new, young researchers. To attract promising students to the field, it must be apparent that good jobs and careers are available in industry and academia. Future plant scientists also can be encouraged to consider entrepreneurial opportunities in starting new food ventures, protecting the environment, and pursuing media communications. Encouraging industry collaboration on plant genomics' grants will allow students to see first-hand the opportunities that training and degrees in plant-related fields can provide.

Objective 6: Strengthen international partnerships to bring the benefits of new discoveries to all

The advances made in the past five years through the NPGI have been greatly enhanced by international partnerships, especially with respect to agriculture in developing countries and increased access to germplasm and data. The challenges of managing big data extend far beyond national boundaries and provide an opportunity to leverage investments and expertise. However, the full potential of these efforts will be realized only if all of the data are freely accessible.

Promote open public genomics resources and analysis tools

The NPGI will continue to uphold its commitment to open access to public genomics resources and analysis tools. Through participation in the U.S. Action Plan for Open Data in Agriculture,[37] the NPGI agencies will ensure that the outcomes of funding will continue to be accessible to all. These efforts will be extended beyond the NPGI, to international partnerships wherever possible.

Coordinate informatics efforts on common datasets

Partnerships with international bioinformatics efforts in plant genomics (e.g., transPLANT) and the *Arabidopsis* Informatics Portal) offer an opportunity to coordinate investments in the development and enhancement of common datasets. Not only does coordination leverage investments, but it also promotes the development of common data standards and interoperability.

Promote core genome annotation efforts

While tremendous advances have been made in reducing the cost of sequencing and making sequencing technology widely available, efforts to annotate genomes have lagged behind, in part, because of cost but also because of a lack of available expertise and tools. The NPGI can promote annotation through the development of enhanced training opportunities and tool development and

---

[37] https://sites.google.com/site/g8opendataconference/home.

through existing cyberinfrastructure investments, international partnerships, and new programmatic activities.

Promote sharing of germplasm internationally

The NPGI will continue to support open access to genetic resources developed with NPGI funding and to international collaboration that safeguards diverse germplasm and makes experimental materials widely available.

IV. Agency Contributions 2014–2018

The NPGI Objectives for 2014–2018 are designed to meet strategic needs in genomics and genetics for understanding plant biology and to fill critical resource gaps for the development of high-yielding, climate-change-resilient, sustainable crop production systems for food, bioenergy, and industrial feedstock production. Member agencies of the IWGPG are committed to supporting the NPGI Objectives for 2014–2018, as summarized in Table 3.

Table 3. Summary of Agency Contributions to the NPGI: 2014–2018

	Objectives					
	1	**2**	**3**	**4**	**5**	**6**
Agency	Develop a new generation of database systems	Create a network of plant germplasm resources	Build tools to advance understanding of plant biology and translate knowledge for breeding	Empower the workforce to use new generation of tools and resources	Build public-private partnerships for translation to the field	Strengthen international partnerships
DOE	+		+	+	+	
NSF	+		+	+		+
USAID		+	+	+	+	+
ARS[38,39]	+	+	+	+	+	+
NIFA	+	+	+	+	+	+
USFS		+	+		+	
Smithsonian Institution (SI)	+	+	+	+	+	+

Note: The "+" signs indicate that the agency programs support the objective.

U.S. Department of Agriculture (USDA) Agricultural Research Service (ARS), U. S. Forest Service (USFS), and National Institute of Food and Agriculture (NIFA)

Open data ecosystem. The ARS will continue to serve as a global leader for open data, providing open access to this knowledge stream for a diverse global clientele through its public database system. The ARS will continue to provide strategically important genomic and genetic database support for the plant research and breeding community under this NPGI Five-Year Plan. The ARS will also continue to support the development of new tools to efficiently capture, manage, and utilize genomic, genetic, and phenotypic data via innovative data-mining strategies. Methods to interconnect diverse data systems will be deployed and will enable researchers to better associate specific genes with agriculturally important traits and build upon genetic advances in one crop to speed genetic gain in others. ARS, iPlant (NSF), KBase (DOE), and NCBI (NIH) scientists will

---

[38] United States Department of Agriculture, *Research, Education, and Economics Action Plan* (Washington, D.C.: Office of the Under Secretary for Research, Education, and Economics, 2012), www.usda.gov/documents/usda-ree-science-action-plan.pdf.

[39] United States Department of Agriculture (Agricultural Research Service), "National Program 301: Plant Genetic Resources, Genomics, and Genetic Improvement Action Plan 2013–2017," www.ars.usda.gov/research/programs/programs.htm?np_code=301&docid=22479.

continue their collaborations to enhance the ecosystem of open data and open access.

Germplasm. The ARS genebanks contain untapped sources of genetic variability necessary to improve crop quantity, quality, and production efficiency.[40] The ARS will continue to strategically acquire and conserve germplasm, develop new screening methods for identifying favorable traits, ensure that germplasm is distributed where and when it is needed, and safeguard these collections for future generations. Coordination for genetic resource development will be through the IWGPG agencies, and coordination for plant germplasm will be through the USDA NPGS in cooperation with the user community. The USFS also maintains genebanks of a limited number of conifer populations and other important at-risk forest species.

Crop genetic improvement. NIFA will continue to support research in basic plant biology to increase our understanding of important traits for crop improvement, such as drought tolerance, disease resistance, and nutrient utilization efficiency. NIFA will also support translational research for sustainable food production by providing opportunities for pre-breeding and germplasm enhancement, cultivar development, selection theory, applied quantitative genetics, and participatory breeding.

ARS, in close cooperation with public and private-sector collaborators, will improve and broaden the genetic base of U.S. crops to reduce genetic vulnerability. ARS scientists will devise and apply new technologies and innovative tools to develop superior new crop varieties and enhanced germplasm and will accelerate the deployment of high-value traits into breeding populations.

The USFS will work with cooperative forest tree breeding programs to understand the genes underlying important traits that impact forest health and productivity. These insights are used to develop forest trees that are more resilient to the stresses imposed by climate change, diseases, and insects.

Genomics. NIFA research, education, and extension opportunities in plant genomics, genetics, and breeding will address future challenges on food security, bioenergy, climate change, and sustainable intensification. By supporting investigations at the molecular, cellular, and whole-plant levels, NIFA will provide innovative and environmentally sound approaches to improve plants and protect them from biotic and abiotic stresses. NIFA will coordinate with Federal partners on areas especially relevant to plant genome structure and function to connect genotype to phenotype and to support new tools, resources, databases, training, and outreach to breeders, research scientists, students, farmers, and growers.

Training and education. ARS and its university cooperators will continue to mentor and train the next generation of crop breeders, geneticists, bioinformaticists, genome researchers, and genetic and genomic resource and information managers. Developing the next cohort of researchers is particularly important because of the projected strong demands for scientific personnel. NIFA will continue to support students and postdocs through grants to colleges and universities. In addition, NIFA's Fellowships Grant Program provides support to pre- and post-doctoral fellows in agricultural research, education, and extension.

---

[40] S. McCouch, G. J. Baute, J. Bradeen, P. Bramel, P. K. Bretting, E. Buckler,J. M. Burke et al., "Agriculture: Feeding the Future," *Nature* 499, no. 7456 (2013): 23–24, doi:10.1038/499023a.

Department of Energy (DOE) Office of Biological and Environmental Research (BER) and Office of Basic Energy Sciences (BES)

BER supports genomics and systems biology research to advance understanding of the fundamental principles that guide plant biology, which enables a predictive understanding of plant systems relevant for sustainable bioenergy production. BES supports fundamental biological and biochemical research to understand energy capture, conversion, and storage in plants, which provides foundational knowledge related to DOE's mission to efficiently capture and utilize solar energy and to convert renewable resources into fuels and other energy-enriched products. BER and BES provide widely accessible support to single- and multiple-investigator projects at academic and private institutions and at the national laboratories. Furthermore, unique DOE user facilities for genomics, structural biology, high performance computing, and environmental studies allow investigators to access the most advanced tools of modern science.

Genomics and systems biology. BER's Genomic Science program will continue to support genomics and systems biology research and the development of novel, high-throughput genomics and imaging technologies, which promotes interdisciplinary approaches to understand plant systems from the molecular to the ecosystem level. The DOE JGI will continue to provide genome sequencing and analysis capabilities and, in partnership with the National Laboratories and academic institutions, will develop new capabilities such as large-scale DNA synthesis and next-generation single-cell genomics.

With the NIFA, BER will continue to support the DOE-USDA Plant Feedstock Genomics for Bioenergy program. This interagency program is committed to accelerating breeding and improving biomass feedstocks through fundamental research in biomass genomics, which will lay the scientific foundation for the use of lignocellulosic materials for bioenergy and biofuels.

Enhancing photosynthetic efficiency. BES's Photosynthetic Systems program will continue to uncover the fundamental science of the biological capture of sunlight and its conversion to and storage as chemical energy in plants. Understanding these processes can guide efforts to increase photosynthetic efficiency for the enhanced generation of biomass as a renewal energy source and to develop new systems for conversion of solar energy into electricity or fuels.

Biological engineering. BER will support research on plant genomic engineering technologies, not only to accelerate the development of plant varieties optimized for biofuel production, but also to deliver new tools and resources. The three DOE Bioenergy Research Centers[41] will be key sources of innovation in this field, as each continues to pursue the basic research underlying high-risk, high-return biological solutions for bioenergy applications.

Computational biology. DOE, by taking advantage of its supercomputers, high-speed connectivity, and distributed computing resources, will encourage the development of new computational biology and modeling tools. DOE's KBase will be instrumental in empowering discoveries and generating new hypotheses. KBase will maintain and nurture its collaboration with NSF's iPlant Collaborative and is pursuing such interactions with other publicly supported database and computational biology programs.

---

[41] BioEnergy Science Center (BESC) (Tennessee), Great Lakes Bioenergy Research Center (GLBRC) (Michigan), and Joint BioEnergy Institute (JBEI) (California). http://genomicscience.energy.gov/centers.

National Science Foundation (NSF)

The NSF supports basic research in all non-medical fields of fundamental science and engineering. NSF's goals—discovery, learning, research infrastructure, and stewardship—provide an integrated strategy to advance the frontiers of knowledge; cultivate a world-class, broadly inclusive science and engineering workforce and expand the scientific literacy of all citizens; build the nation's research capability through investments in advanced instrumentation and facilities; and support excellence in science and engineering research and education through a capable and responsive organization.

Genomics. The Plant Genome Research Program (PGRP) will continue to support genomics and systems biology research and the development of novel tools, technologies, and research resources to advance our understanding of the structure and function of economically important crop plants and plant processes of potential economic value. Support for plant research will continue to be available at all levels—from individual investigators through consortia of investigators working in virtual centers.

The PGRP will also continue to partner with the ARS wherever possible to ensure public access to the germplasm resources developed through NSF funding.

Databases and bioinformatics. The NSF will continue to promote the deposition of data emerging from PGRP-supported projects into long-term USDA-supported databases and will promote leveraging investments in analytical tools for data integration. NSF-funded cyberinfrastructure activities, such as iPlant, will continue to coordinate with U.S. counterparts including KBase and with international partners, wherever possible. In addition, activities supported through the NSF's Advanced Cyberinfrastructure program such as XSEDE (hwww.xsede.org) and FutureGrid (www.futuregrid.org) will continue to provide computing and advanced digital resources and services to researchers and educators in coordination with U.S. partners such as the Open Science Grid (www.opensciencegrid.org) and international partners such as the Partnership for Advanced Computing in Europe (PRACE) (www.prace-project.eu).

Training and education. The NSF will continue to promote the integration of training, broadening participation, and public outreach into its awards to develop a new generation of plant researchers empowered to use the resources and tools emerging from the NPGI. The NPGI Postdoctoral Fellowship and Mid-Career Investigator funding opportunities will be used to address critical workforce needs in areas such as plant breeding, bioinformatics, plant physiology, and biochemistry, in coordination with other NPGI-supported activities.

U.S. Agency for International Development (USAID)

International partnership. USAID will continue to support the President's initiatives to fight world hunger, including the Feed the Future and the New Alliance for Food Security and Nutrition programs. As part of these efforts, USAID is scaling up a comprehensive approach for fighting hunger and strengthening food security by (1) investing in cutting-edge scientific and technological research[42] to increase farmers' productivity and yields; (2) developing agricultural

---

[42] www.usaid.gov/what-we-do/agriculture-and-food-security/investing-agricultural-research-and-development.

markets, expanding trade, and using mobile phones to provide real-time prices to increase farmers' profitability; (3) helping farmers access capital so they can expand; (4) offering extension services that will educate farmers in the best agronomic techniques; (5) developing strategies so that countries can feed their populations without depleting natural resources; and (6) providing emergency food assistance so that vulnerable and malnourished populations can survive and quickly recover in times of crisis. The NPGI has helped foster many of USAID's efforts in development, including scaling up of climate and pathogen-resistant crop plants globally.

Smithsonian Institution (SI)

The SI supports basic research in plant genomics, taxonomy, biogeography, phylogenetics, and systems biology to advance our knowledge of functional and sustainable ecosystems and to document and preserve biodiversity.

International partnership. The SI will continue to expand partnerships and initiatives throughout the world and will increase global long-term, large-scale efforts to monitor plant biodiversity, including wild relatives of agricultural species, through efforts such as the Smithsonian Institution Global Earth Observatory (SIGEO) and the Center for Tropical Forest Science (CTFS), which are global networks of forest research plots committed to the study of tropical and temperate forest function and diversity. These initiatives monitor the growth and survival of about 4.5 million trees of approximately 8,500 species to increase scientific understanding of forest ecosystems, guide sustainable forest management and natural-resource policy, monitor the impacts of climate change, and build capacity in forest science.

Genomics. The SI will continue to increase the capacity for genomic research across all areas of plant biology, including projects in biodiversity, phylogenetics, population genetics, toxicology, environmental monitoring, complex plant animal interactions, and genomic analyses. These projects are linked with broad DNA barcoding initiatives to assist plant taxonomic research and rapid plant identification and complement the establishment of global networks to promote standardized collection of metadata, secure interoperability, and improved best practices.

Training and education. The SI will continue to provide professional training opportunities for teachers, students, and researchers interested in learning the latest techniques and networking with experts in the field. This training will include leveraging an active and networked community of educators and public engagement specialists that infuse public exhibits in museums with stories on genomic research, craft messages for specific public sectors, build public programs, produce media, and create collaborations with external organizations.

Appendix A: Charter of the Interagency Working Group on Plant Genomics (IWGPG)

CHARTER
of the
INTERAGENCY WORKING GROUP ON PLANT GENOMICS
SUBCOMMITTEE ON LIFE SCIENCES
COMMITTEE ON SCIENCE
NATIONAL SCIENCE AND TECHNOLOGY COUNCIL

Official Designation

The Interagency Working Group on Plant Genomics (IWGPG) is hereby reestablished by action of the National Science and Technology Council (NSTC), Committee on Science (CoS), Life Sciences Subcommittee (LSSC).

Purpose and Scope

Recognizing the critical role of plant genomics research in addressing the needs of a growing world population and an environment impacted by climate change, in 2009, the IWGPG published the five-year plan (2009–2013) for the National Plant Genome Initiative (NPGI).[43] Since then, good progress has been made toward achieving the Plan's stated goal of expanding genomic resources for every economically important plant,[44] and we expect that as sequencing costs continue to decline, genome sequence and related data will increase. While this wealth of data will revolutionize the way plant scientists perform research, there is currently a shortage of standards for data generation and collection—something that significantly hampers efforts to mine the numerous and inconsistent datasets for accurate information.

Standardization of data, including metadata such as experimental and environmental conditions, is critical for proper interpretation of results. This problem is compounded at a global scale where, without national and international coordination, the development of tools and resources is haphazard, inconsistent, and plagued by redundancies and gaps.

The purpose of the IWGPG is to engage the plant science community in prioritizing genomics tools and resources, defining new strategies that will meet community needs and priorities sustainably, advancing biological innovation and breakthrough discovery, and improving coordination among Federal agencies and international plant genomics partners.

---

[43] National Plant Genome Initiative: 2009–2013, National Science and Technology Council (Washington D.C., January 2009, www.whitehouse.gov/files/documents/ostp/NSTC%20Reports/NPGI%20Five-Year%20Plan%202009-2013.pdf.

[44] Crop Values 2011 Summary (February 2012) USDA, National Agricultural Statistics Service, www.nass.usda.gov).

Functions

The IWGPG will

1. Review NPGI accomplishments against the objectives defined in the NPGI 2009–2013 Strategic Plan and the development of coordinated solutions to data access, data analysis, and data synthesis as a means to foster biological innovations that support innovation in the practice of agriculture (January 2013).
2. Inventory current international plant genomics tools and resources, including analytical tools, genetic tools, sequencing needs, and databases (January 2013).
3. Identify strategic research needs and resource gaps for the development of sustainable systems for food, bioenergy, and industrial feedstock production. This inventory should consider the minimization of inputs such as water, energy, pesticides, and fertilizer; the effects of climate change and increased agricultural productivity; the minimization of environmental impact using plant genomics; and opportunities for Federal agency coordination, cooperation, public/private partnerships, and associated opportunities to enhance training, education, and public outreach through engagement of stakeholder communities as permitted by law (January 2013).
4. To the extent permitted by law, seek public input through a website (December 2012) and through a workshop organized at the Plant and Animal Genome Conference (January 2013).
5. Develop strategic objectives for the NPGI 2014–2018 Strategic Plan based on the prioritized research needs, defined resource gaps, and identified opportunities to guide the use of agency resources to high-priority activities (April 2013).
6. Produce the NPGI 2014–2018 Strategic Plan containing recommendations for Federal agencies regarding open access, data sharing, and interoperability of tools and databases to improve the practice of agriculture, reduce the demands on environmental resources, and address challenges posed by global climate change (July 2013).

Membership

The following NSTC departments and agencies are represented on the IWGPG:

Department of Agriculture

Department of Commerce[45]

Environmental Protection Agency

Department of Defense

Department of Energy (Co-chair)

Department of Health and Human Services

Department of the Interior

---

[45] Specifically, but not exclusively, to include the National Institute of Standards and Technology (NIST).

Department of State

Department of Veterans Affairs

National Aeronautics and Space Administration

National Science Foundation (Co-chair)

Smithsonian Institution[46]

U.S. Agency for International Development

The following organizations in the Executive Office of the President shall also be represented on the IWGPG:

Office of Management and Budget

Office of Science and Technology Policy

Cooperating departments and agencies shall include other such Executive organizations, departments, and agencies as the Co-chairs may, from time to time, designate.

Private-Sector Interface

The IWGPG may work with the President's Council of Advisors on Science and Technology to secure private-sector[47] advice and will recommend to the LSSC and/or the Director of the Office of Science and Technology Policy the nature of additional private-sector advice needed to accomplish its mission. The IWGPG may also interact with and receive ad hoc advice from various private-sector groups as consistent with the Federal Advisory Committee Act.

Termination Date

Unless renewed by the Co-chairs of the LSSC prior to its expiration, the IWGPG shall terminate no later than September 30, 2013.

Determination

I hereby determine that the establishment of the Interagency Working Group on Plant Genomics is in the public interest in connection with the performance of duties imposed on the Executive Branch by law, and that such duties can best be performed through the advice and counsel of such a group.

---

[46] Per the Memorandum of Understanding between the National Science and Technology Council (NSTC) and the Smithsonian Institution (SI), September 14, 2012.

[47] The Federal Advisory Committee Act, 5 U.S.C. App., as amended, does not explicitly define "private-sector," but the phrase is generally understood to include individuals or entities outside the Federal government such as, but not limited to, the following: non-Federal sources, academia, State, local or Tribal governments, individual citizens, the public, non-governmental organizations (NGOs), industry associations, and international bodies.

Approved:

Kathy Hudson
Co-chair of Subcommittee on Life Sciences, and
Deputy Director for Science, Outreach, and Policy
National Institutes of Health
Department of Health and Human Services

Date

Chavonda Jacobs-Young
Co-chair of Subcommittee on Life Sciences, and
Associate Administrator for
National Programs
Agricultural Research Service
Department of Agriculture

Date

Sharlene Weatherwax
Co-chair of Subcommittee on Life Sciences, and
Associate Director of Science for Biological and
Environmental Research
Office of Science
Department of Energy

Date

John C. Wingfield
Co-chair of Subcommittee on Life Sciences
Assistant Director for Biological Sciences
National Science Foundation

Date

Appendix B: Accomplishments for 2009–2013

Objective 1: Expand genomic resources for every major plant of economic importance

During the first ten years of the National Plant Genome Initiative (NPGI), the major genome sequencing goals focused on the generation of reference genomes for key model and crop plant genomes of increasing size and complexity and for their wild relatives. During the past five years, the dramatic decrease in sequencing costs that has resulted from the development of next-generation sequencing technologies, improved genome assembly, and innovative optical mapping strategies has made it possible to generate thousands of genome sequences derived from different plant accessions and individuals in populations and in more complex reference genomes from plants such as wheat, barley, switchgrass, and pine. New technological developments have also accelerated the understanding of the epigenome's role in phenotypic variation.

- The completion of genome sequences for diverse but strategically selected plant species over the past five years has accelerated basic and applied research efforts for crop improvement. The genome sequences of major crop plants (maize, soybean, tomato, potato, barley, grape, peach, strawberry, and cacao) and non-crop models (*Brachypodium* and *Setaria*) are now being used to understand the genetic basis of crop traits and accelerate rapid breeding for enhanced yield, drought tolerance, enhanced nutrient-use efficiency, and improved biofuel production (Appendix D).
- Newly available sequence resources range from gene-enriched sequences to high-quality whole-genome coverage. Genome sequencing of plant populations has revealed the genetic diversity and genetic opportunity inherent in model and crop plants that can now be used for crop improvement and has facilitated the mapping of new genes for the control of flowering time, nitrogen metabolism, enhanced nutritional value, conversion of biomass to biofuels, disease resistance, and growth in poor soils.
- The genome sequences of the wild relatives of economically important crop plants, including soybean, poplar, potato, and tomato, have provided powerful resources to help address grand challenge questions in plant biology, such as inbreeding depression, domestication, and the evolution of complex traits. These genome sequences have also been used to identify new genetic sources of agronomically important traits, such as disease resistance and enhanced human nutrition.
- Comparative database resources were expanded to provide access to sequence information and computational tools for strategically selected plants, which provided a means to perform comparative analyses at the genome, protein, metabolic pathway, genetic, and trait levels for basic and applied research (Appendix E).
- The 1001 Genomes Project,[48] which was initiated in 2008, has made tremendous progress toward cataloging genetic variation across whole genomes of *Arabidopsis thaliana* accessions to identify alleles underpinning phenotypic diversity across the

---

[48] www.1001genomes.org.

Northern hemisphere. These accessions are often locally adapted to the environments in which they grow, which allows insights into the genetic variation that leads to phenotypic variation.

- New systems-wide technologies have allowed for the characterization and integration of plant proteomes, metabolomes, interactomes, and other "-omics" profiles. With these data, researchers can better elucidate genetic, regulatory, and metabolic networks that are critical to advancing our understanding of fundamental plant processes for crop improvement.
- Epigenetic modifications of chromosome structure and gene expression were identified on a genome-wide scale in *Arabidopsis*, maize, and rice, which provided a first glimpse of these numerous epigenetic "marks" and revealed that transposon-mediated gene silencing is widespread, varies in natural populations and in cycles within populations, is reprogrammed in pollen and in seeds, and, as predicted by Barbara McClintock, may contribute to phenotypic variation in plants.[49,50]
- Complete parental linkage maps of two tetraploid switchgrass genotypes have been constructed, which will enable the development of marker-assisted selection (MAS) strategies to improve switchgrass and other potential bioenergy grass species.
- A Smithsonian biorepository[51] that has the capacity to preserve more than four million samples with vouchers was created for genomic studies and linked to global long-term, large-scale efforts through the Global Genome Initiative to monitor plant diversity, including wild relatives of agricultural species.

Objective 2: Advance plant systems biology

Recent advances in proteomics and metabolomics have generated large datasets for model plants and economically important crop plants. NPGI-funded projects are now able to integrate and analyze these data with other "-omics" datasets to provide a systems-level understanding of plant gene regulatory pathways underlying complex traits such as mineral element accumulation, the biochemical pathways of specialized metabolites potentially important for engineering strategies for biotic stress tolerance, and seasonal growth and phenology traits.

- Advances in next-generation sequencing technologies, computational tools, and genetic resources have democratized the use of genomics to enable gene discovery and the identification of genes and gene networks associated with any trait of interest or economic value. Identification of a multi-gene network of nitrogen response genes may provide for new strategies to increase nitrogen use efficiency (NUE) in crop plants. The genome-scale functional gene network, RiceNet,[52] generated using heterogeneous genomics and proteomic datasets, allows the prediction of the function of genes that are necessary for plant defense against bacterial pathogens. Quantitative genetics and genomics, together

---

[49] www.plant-epigenome.org.

[50] http://genomevolution.org/wiki/index.php/EPIC-CoGe.

[51] www.mnh.si.edu/ggi.

[52] www.functionalnet.org/ricenet.

with high-throughput, cutting edge two-dimensional (2D) and three-dimensional (3D) image capture and analysis tools, were used to identify genes that regulate the growth, development, and function of root systems in response to different environmental and soil conditions.

- Research supported through the NPGI is providing a deeper understanding of biosynthetic pathways to allow the development of novel strategies to enhance plant feedstocks for biofuel. Switchgrass engineered to contain the maize juvenility gene, Corngrass1, has significantly higher glucose release without expensive pretreatment and an innate mechanism that prohibits transgene escape. Newly identified gene networks and proteins necessary for the synthesis and accumulation of seed oils hold promise for the production of biodiesel. These findings promise new approaches for the improvement of dedicated bioenergy crops that will enable the production of renewable biofuels.
- Basic research funded through the NPGI has developed genome-scale gene expression datasets for economically important plants that have accelerated the analysis of gene networks for economically important complex traits. Three heterosis-linked genes identified in tomato are helping to elucidate this phenomenon that has been exploited by commercial agricultural entities for decades. Long-term funding for basic research in *Arabidopsis* and tomato has revealed a link between freezing and drought that can now be exploited for engineering improved stress tolerance in agriculturally important crop plants.
- Predictive models that can greatly facilitate crop improvement are being constructed. Computational methods that predict phenotype from genotype in maize with 90 percent accuracy were developed and are being used to accelerate breeding. Similarly, the network of genes in *Populus trichocarpa* that regulate lignin biosynthesis during wood formation will enable the development of new strategies for improving wood production and processing.[53]
- Significant advances have been made in the development of new computational tools for data integration and modeling, which furthers our understanding of the genes underlying complex traits. A consortium of databases and biological research communities has been established to develop and provide access to standard gene ontologies for use by the broader functional genomics community.[54] The Trait Analysis by Association, Evolution, and Linkage (TASSEL) software platform, together with the genotyping by sequencing (GBS) analysis pipeline, facilitates association mapping of complex traits in plants.[55] The Pathway Tools software platform allows the creation, curation, visualization, and analysis of metabolic pathways using diverse datasets. Improving Plant Genome Annotation (IPGA) supports the development of annotation tools to facilitate interpretation of the structure, function, and evolution of genomes of economically important plants, and the genome annotation software platform MAKER has been used to annotate plant and animal genome sequences.

---

[53] www.nsf.gov/awardsearch/showAward?AWD_ID=0922391.

[54] www.plantontology.org.

[55] www.maizegenetics.net/index.php?option=com_content&task=view&id=89&Itemid=119.

- In addition to Challenge Grant funding opportunities focused on important issues for crop improvement, such as understanding the mechanism of heterosis and improving genome annotation, novel review mechanisms have been used to promote the development of innovative approaches for addressing grand challenge questions. An “Ideas Lab” brings together scientists from diverse backgrounds to stimulate new ideas and strategies to solve long-standing problems in economically important crop plants. Through two Ideas Labs jointly funded by the National Science Foundation (NSF) and the Biotechnology and Biological Sciences Research Council (BBSRC) in the United Kingdom, U.S. and U.K. researchers are working together to test potentially transformative strategies for enhancing the efficiency and sustainability of photosynthesis and for increasing nitrogen availability to crop plants while maintaining or increasing yield, thus decreasing the need for applying fertilizer.

Objective 3: Translate basic discovery to the field

Translation of the basic research outcomes generated through NPGI funding has been promoted and accelerated through the development of germplasm resources for breeding.

- New genomics tools and approaches are being implemented in crop breeding programs, which reduces the time needed from years to weeks for researchers to make associations between genetic markers and traits. Such advances in marker discovery and platforms, built upon genotyped genomes that include maize, soybean, grape, rice, cotton, tomato, potato, fava bean, onion, blueberry, cranberry, and oat, are now being used by breeders to rapidly develop improved cultivars.
- Genomic selection projects funded through the NPGI have led to a substantial reduction in cost per sample for GBS and potential acceleration of breeding for economically important crops.
- Advances in genomics have facilitated the identification of genes that regulate nutritionally important traits. The discovery of two genes in corn that increase pro-vitamin A levels in grain 16-fold enabled breeders to develop and release three hybrid corn varieties of higher nutritional value in Africa in 2012, with more planned for 2013.[56]
- This acceleration is having a profound impact on breeding new varieties, with improved agronomic traits such as higher yield, drought resistance, and nutrient-use efficiency. The identification and cloning of the major flowering-time gene in sorghum are enabling the conversion of tropical sorghums to short stature, creating early flowering forms, and making a diverse range of new germplasm available to sorghum breeders.
- NPGI-supported research is providing a deeper understanding of biosynthetic pathways to allow the development of novel strategies to improve food oils. Such studies have led to high-oleic soybean marketed under the brand names Plenish^{TM} (DuPont-Pioneer) and VistiveGold^{TM} (Monsanto).

---

[56] www.harvestplus.org/sites/default/files/HarvestPlus%20Annual%20Report%202012%20Web.pdf.

- Genomics' approaches have led to a better understanding of the diversity of effectors and virulence factors found in plant pathogens such as the Irish potato famine pathogen *Phytophthora infestans* and the identification of a gene that confers resistance to soybean cyst nematode and, for wheat, root knot nematode resistance.[57] These approaches have also elucidated the nature of fungal pathogen resistance in Chinese chestnut—work that will lead to better pest resistance and improved crop yield.

- Molecular tools and genetic resources have been deployed to combat a new and deadly form of wheat stem rust, a fungal disease caused by Ug99, a pathogen that has the potential to spread rapidly and is threatening much of the world's wheat supply. Resistance genes effective against Ug99 were discovered in wild relatives contained in the U.S. Department of Agriculture (USDA) germplasm collections.[58], [59] Advanced molecular methods have been used to deploy these genes into new wheat varieties for the United States and the world, which provides protection from Ug99.

- NPGI-supported resources have also affected the breeding of slow cycling crops such as tree fruits, where breeding is an expensive, long-term process. The "FasTrack" system utilizes an early flowering and fruiting gene to accelerate the breeding cycle. This approach allows for the rapid incorporation of important traits into plums and other long-generation-cycle crops and provides tree fruit and other long-cycle crop breeders the ability to respond to new market demands, climate changes, diseases, and pests in a way never before possible.[60]

- The impacts of the NPGI "open data" and "open access" policies are global. The U.S. Agency for International Development (USAID), USDA, NSF and the U. S. Department of Energy (DOE) have collaborated to provide global access to genome sequence information, genetic data, and germplasm resources for crop improvement in countries with developing agricultural economies to address major challenges facing agriculture today. Appendix F includes examples of partnerships between NPGI agencies, the Consultative Group for International Agricultural Research (CGIAR), and the private sector to tackle important challenges such as stress tolerance, disease resistance, and the impacts of soil quality on the yield of important field crops, including wheat, corn, sorghum, cassava, chickpea, and pigeon pea.

---

[57] V. M. Williamson, V. Thomas, H. Ferris, and J. Dubcovsky, "An Translocation Confers Resistance Against Root-knot Nematodes to Common Wheat," *Crop Science* 53, no. 4 (2013): 1412–1418, doi:10.2135/cropsci2012.12.0681.

[58] S. Periyannan, J. Moore, M. Ayliffe, U. Bansal, A. Wang, L. Huang, K. Deal, et al., "The Gene *Sr33,* an Ortholog of Barley *Mla* Genes, Encodes Resistance to Wheat Stem Rust Race Ug99," *Science* 341, no. 6147 (2013): 786–788, doi:10.1126/science.1239028.

[59] C. Saintenac, W. Zhang, A. Salcedo, M. N. Rouse, H. N. Trick, E. Akhunov, and J. Dubcovsky, "Identification of Wheat Gene *Sr35* That Confers Resistance to Ug99 Stem Rust Race Group," *Science* 341, no. 6147 (2013): 783-786, doi:10.1126/science.1239022.

[60] http://ucanr.edu/sites/fastrack.

Objective 4: Develop coordinated solutions to data access, data analysis, and data synthesis

The wealth of resources generated through NPGI funding since 1998 and NGPI's continued commitment to open access to the data and resources emerging from its projects have benefits and challenges. While addressing grand challenge questions in plant science on a genome-wide scale is becoming possible, often the barrier is accessing and analyzing large datasets in new ways. Coordinated cyberinfrastructure investments made by the NPGI agencies over the past five years have led to the development of new approaches that are changing the way that plant research is conducted.

The iPlant Collaborative (iPlant)

The iPlant Collaborative[61] was supported by NSF in 2008 as a cyberinfrastructure collaborative for the plant sciences to enable new conceptual advances through integrative, computational thinking. iPlant's activities are community driven and involve plant biologists, computer and information scientists, and engineers who work together in integrated teams to identify grand challenge questions in the plant sciences that can be addressed through computational approaches. The iPlant Collaborative is led by the University of Arizona in partnership with the University of Texas at Austin, Cold Spring Harbor Laboratory (CSHL), and the University of North Carolina Wilmington. By design, its activities are developed and coordinated through a broad network of collaborators.

The iPlant Collaborative's activities were initially developed by the identification of grand challenge questions in partnership with the research community. Self-organizing Grand Challenge teams for each question work with the iPlant team to develop Discovery Environments, a cyberinfrastructure for open research and education. iPlant's activities are currently organized around two Grand Challenges—the iPlant Tree of Life (iPToL) and iPlant Genotype to Phenotype (iPG2P)— and additional seed projects. iPToL has organized outcomes of investments in phylogenetics and computer sciences to construct a cyberinfrastructure capable of scaling up phylogenetic tree inference methods by more than 100-fold to disseminate the data associated with trees of 500,000 plants. iPG2P focuses on elucidating the relationship between the information in the genome of a plant and its interplay with environmental inputs to produce the range of possible plant phenotypes. These efforts will be critical in predicting and understanding the impacts of a changing environment on crop performance in the field.

The iPlant tools and resources include access to cloud computing through Atmosphere and training tools such as DNA Subway, which makes high-level genome analysis broadly available to students and educators through a user-friendly interface. The Powered by iPlant program allows existing projects to use the iPlant cyberinfrastructure to provide services to their users by integrating with iPlant's authentication system, data store, job execution system, and semantic web services and by using its servers to host their resources. Current program members include BioExtract server, Cyberinfrastructure for Phylogenetic Research (CIPRES), Comparative

---

[61] www.iplantcollaborative.org.

Genomics (CoGe) platform, the Center for Genome Research and Biocomputing, PHYLOfantastic, TARGeT, the Integrated Breeding Portal, and BrachyBio Data Central.

DOE Systems Biology Knowledgebase (KBase)

In 2011, DOE funded KBase,[62] led by the Lawrence Berkeley National Laboratory. The KBase collaboration includes Argonne, Brookhaven, and Oak Ridge National Laboratories and CHSL, the University of California at Davis, Hope College in Michigan, the University of Illinois at Urbana-Champaign, and Yale University. The DOE Joint Genome Institute (JGI) and the DOE Bioenergy Research Centers (BRCs) also contribute to this effort. KBase also coordinates its activities with the iPlant Collaborative.

KBase is an open source and open architecture computational environment for integrating large, diverse datasets generated by the scientific community and using this information to advance predictive understanding, manipulation, and design of biological processes in an environmental context. It is meant to be a community resource that enables users to integrate a wide spectrum of genomics and systems biology data, models and information for microbes, microbial communities, and plants. Powerful tools within KBase can analyze and simulate data to predict biological behavior, generate and test hypotheses, design new biological functions, and propose new experiments.

KBase, in its first public release in February 2013, exceeded expectations, showing a hybrid high performance computing and cloud infrastructure built for 100 GBit/second data transfer over the Energy Sciences Network (ESnet). Currently, KBase has over 3 petabytes of storage and 12,000 cores for data processing within the Magellan cloud computing resource. KBase holds over 175 eukaryotic genomes, including many variants of poplar, *Arabidopsis*, sorghum, *Chlamydomonas*, *Brachypodium*, *Miscanthus*, and switchgrass, and many other algae and fungi. Furthermore, it features phenotypes for plant genome variants and 731 gene expression experiments in *Arabidopsis* and poplar. Also available at KBase are services for determining genetic variation, genotype-to-phenotype mapping, candidate gene filtering, trait modeling, pathway enrichment, and co-expression network analysis for gene function prediction, and it will soon provide services for plant metabolic modeling. KBase's infrastructure allows users to share data, analysis tools, workflows, and scientific conclusions with the research community.

Both iPlant and KBase coordinate with international plant genomics databases, including the Trans-national Infrastructure for Plant Genomic Science (transPLANT), [63] which is a European Union (EU) Seventh Framework Program that brings together 11 institutions involved in plant data integration, management, and analysis.

---

[62] http://kbase.us.

[63] http://transplantdb.eu.

Building a network of databases to facilitate data integration and linkage with germplasm resources

With iPlant and KBase serving as hubs for many of the NPGI-supported efforts to analyze and integrate large-scale data, NPGI-supported databases focus on resources for specific crops or on efforts that impact many crop genomes, such as annotation.

The goal of the Comprehensive Phytopathogen Genomics Resource[64] (CPGR) is to improve the value of plant pathogen genome sequences through uniform, updated, and value-added annotation. The data generated from the project are useful to breeders who are interested in developing pathogen-resistant crop plants.

- NPGI strategic investments in open data and germplasm enable public and private researchers to drive increases in crop productivity and resilience to disease and pests through enhanced access to knowledge, data, and germplasm. Each year, more than 500,000 users—domestic and international—mine the USDA crop genetic and genomic database system (Maize Genetics and Genomics Database (MaizeGDB), SoyBase, Gramene, Legume Information System (LIS), and GrainGenes) crop genome sequence information, mapping data, and analysis and visualization software to access genetic diversity that represents over 100 years of germplasm collection, improvement, and preservation.
- The Soybean KnowledgeBase[65] (SoyKB) is a web resource for soybean translational genomics that allows users to take full advantage of the *Glycine max* genome sequence and the wealth of genomics data being generated for soybean.
- The Biofuel Feedstock Genomics Resource (BFGR)[66] and genome browser is a web-based portal and database that provides tools for functional genomics and comparative analysis of plant biofuel feedstock species, many of which lack a genome sequence, to facilitate improvement of lignocellulosic biofuel feedstocks.

Objective 5: Enhance education, training, and outreach

A major goal of the NPGI, from its inception, has been to attract and train a new generation of plant breeders while supporting the continued development of plant breeding professionals. In the past five years, a Plant Breeding Training Network (PBTN)[67] has been established through which researchers can collaborate and learn via courses, seminars, and discussions. Specific activities to recruit new students into plant breeding have included collaboration with Minority Serving Institutions, mentoring undergraduates, creating educational materials with plant breeding examples, and producing educational materials and recruitment films. Specific activities to train graduate students included immersing students in plant breeding programs, creating an

---

[64] http://cpgr.plantbiology.msu.edu.

[65] www.soykb.org.

[66] http://bfgr.plantbiology.msu.edu.

[67] http://passel.unl.edu/communities/pbtn.

online collaborative space, providing online courses and workshops, and supporting scientific meeting attendance.

- A website, housed at eXtension.org,[68] has been created to help train the next generation of breeders in marker-assisted breeding approaches and strategies. Accomplishments include the establishment and certification of the Plant Breeding and Genomics Community of Practice (PBGCoP),[69] the establishment of a shared Drupal workspace, the development of workshops that address the application of genomic resources to plant breeding, and the development of a plant breeding and genomics (PBG) social network (primarily through membership to the PBG newsletter). The PBGCoP has published educational materials on a variety of important topics in the PBG community, including experimental design, statistical inference, molecular biology, selection, genetic mapping, and quantitative trait loci and population development. In addition to traditional static tutorials, PBGCoP has worked to develop a self-paced, inquiry-based curriculum to connect breeders and students and to open access to computational tools and genomic data through hands-on learning.
- The NPGI has initiated the training of the next generation of researchers from the developed and developing countries in sequencing, annotation, and discovery of genes linked to functionally important traits in plants and pathogens that are useful for plant breeding. USAID has also recently funded the Borlaug Higher Education Agricultural Research and Development (BHEARD) program, which supports international graduate student training in “Feed the Future” countries in Africa (Uganda, Malawi, Ghana, Mali) and South East Asia (Bangladesh).
- The NPGI provides support for postdoctoral associates through several programs. For example, National Plant Genome Initiative (NPGI) Postdoctoral Research Fellowships are jointly sponsored by NSF, DOE, and ARS to provide recipients plant genomics training that emphasizes quantitative genetics, modern breeding approaches, and bioinformatics.
- The DOE Office of Science Early Career Research Program supports the development of individual research programs of outstanding scientists early in their careers (within 10 years of having received a Ph.D.), focusing on the disciplines supported by the DOE Office of Science, including plant systems.
- The recognition that new approaches and technologies are needed to advance the field of plant genomics led to the offering of the NSF Mid-Career Investigator Awards in Plant Genome Research (MCA-PGR). The MCA-PGR seeks to increase participation of mid-career investigators primarily trained in fields other than plant genomics, such as, but not limited to, plant physiology or plant biochemistry.
- As part of the NPGI, the USDA Agricultural Food and Research Initiative (AFRI) places a high priority on science that can create an impact through plant genomics, genetics, and breeding. The NIFA encourages AFRI-funded scientists to undertake outreach activities to accelerate the adoption of improved plant materials. As a result, scientists working on

---

[68] www.extension.org.

[69] www.extension.org/pages/32362/about-the-plant-breeding-and-genomics-community-of-practice.

these projects have reached over 3,400 people at in-person workshops and have developed webinars, training modules, videos, and other outreach materials that have been accessed over 370,000 times as of spring 2013.

- Diverse training and outreach activities continue to broaden the impacts of the research supported by the NPGI. These activities include workshops to train the community in the use of tools and resources at NPGI-supported databases, such as Gramene, Plant Ontology, MaizeGDB, and SoyBase; research training programs for K–12 teachers, such as the Iowa State University "Plant Genomics Education Outreach Program"; and educational videos such as "Tomato–Decoded" produced in collaboration with NBC Learn. In addition, museum exhibits, such as the "Maize: Mysteries of an Ancient Grain",[70] with its related online teacher materials, has been viewed by over 300,000 people at five sites in the past three years.

Objective 6: Broaden societal impacts

The NPGI has served as a focal point for the coordination of Federally supported plant genomics research, with some activities supported through non-governmental foundations and investments leveraged through international partners. These collaborations build on the extensive investments that have already been made in genome resources and data for crops of economic importance in the United States and have utility for additional crops and agricultural challenges that may arise in the future.

- In 2009, the NSF established Basic Research to Enable Agricultural Development (BREAD)[71] to support early-concept research focused on challenges facing smallholder farmers in developing countries. This five-year NSF program, supported by $24 million from the NSF and $24 million from the Bill & Melinda Gates Foundation, has provided support for tools that will accelerate plant breeding through genomic selection and the use of doubled haploids and has funded projects that are focused on disease resistance, including uncovering cryptic sources of wheat resistance to wheat stem rust and understanding the role of plant satellite DNA in viral diseases of cassava. BREAD integrates the NSF's peer review of bottom-up ideas and the Gates Foundation's extensive expertise in addressing developing world issues and provides funding for the United States and international partners. The complementary Partnerships for Enhanced Engagement in Research (PEER),[72] a USAID-funded competitive grants program managed in partnership with the NSF, provides an opportunity to support scientists from developing countries who are working with NSF-funded scientists at U.S. institutions.
- Federally supported plant genomics research linked with crop breeding is being leveraged by non-governmental organizations (NGOs), foundations, and international partners to protect the food supply. Plant genomics information and new computational approaches are enabling corn, rice, sorghum, legume, and wheat researchers to use next-generation

---

[70] www.museumoftheearth.org/maizeexhibit/Maize%20Info%203.9.12.pdf.

[71] www.nsf.gov/funding/pgm_summ.jsp?pims_id=503285&org=BIO.

[72] http://sites.nationalacademies.org/pga/dsc/peer/index.htm.

sequencing advancements for rapid whole-genome characterization of breeding lines and diverse genetic resources. The innovative plant genetics research supported by the NPGI agencies is changing the International Maize and Wheat Improvement Center (CIMMYT) Global Wheat Program[73] and many other international breeding efforts from a conventional breeding approach toward the more advanced genomic selection approaches that will accelerate genetic gains, especially for complex traits such as yield and heat tolerance.

- NPGI-supported researchers in the United States have leveraged resources and worked with an international scientific consortium to protect the global grain supply from the deadly Ug99 wheat stem rust strain. Ug99 protection has been markedly advanced by discoveries of new rust-resistant genes, molecular tools to accelerate crop breeding, and a new understanding of the fungal genome supported by the NPGI. Since 2005, ARS scientists have conducted Ug99 resistance screening in Eastern Africa in partnership with Kenyan and Ethiopian national research programs, CIMMYT, the International Center for Agricultural Research in the Dry Areas (ICARDA), and the Borlaug Global Rust Initiative.[74] Support from the Durable Rust Resistance in Wheat Project[75] (funded by the Bill & Melinda Gates Foundation and U.K. Department for International Development (DFID) and led by Cornell University) has helped develop the first Ug99-resistant wheat varieties in Kenya and Ethiopia. The USDA-USAID Wheat Productivity Enhancement Project in Pakistan has resulted in the first release of Ug99-resistant varieties for Pakistan and the resumption of cereal disease pathogen monitoring. USAID-supported projects continue to enable the rapid build-up of Ug99-resistant seeds for farmers and advance research that helps growers better manage cereal rust pathogens.

An opportunity for future collaboration with many of these partnerships involves the data they produce. Since its inception, the NPGI has operated under the guiding principles that all research resources, including data, software, germplasm, and other biological materials and research tools should be openly accessible in a timely manner. Investments in the nation's cyberinfrastructure made through iPlant, KBase and the ARS databases have had an effect well beyond the United States because this capital provides access to, and facilitates the use of, the NPGI data. The USDA has taken a leadership role in promoting open data for agriculture through its organization of the G-8 International Conference on Open Data for Agriculture, held April 29–30, 2013, at the International Finance Corporation in Washington, D.C.[76] Agencies participating in the NPGI contributed to the U.S. Action Plan for Open Data in Agriculture by including policies and implementation plans that serve as the prototype for action plans from all other G-8 countries. The NPGI agencies are listed as being responsible for promoting policies, investing in projects, and making all data, including crop genome databases and genetic resources, available. These

---

[73] http://intranet.cimmyt.org/en/programs-and-units/global-wheat-program.

[74] www.globalrust.org/traction/permalink/about2.

[75] www.wheatrust.cornell.edu/index.html.

[76] https://sites.google.com/site/g8opendataconference/home.

agencies are also involved in efforts to increase access to the results of Federally funded scientific research,[77,78] which builds on the NPGI principles.

---

[77] www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.

[78] www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government-.

Appendix C: Stakeholder Input

- Climate Change and Agriculture in the United States: Effects and Adaptation, U.S. Department of Agriculture (USDA) Agricultural Research Service, Climate Change Program Office (February 2013).
- Corn Breeding Research (CBR – NCCC167) Research Priorities (March 2013).
- Food, Fuel, and Plant Nutrient Use in the Future, Council for Agricultural Science and Technology (2013).
- Maize Genetics Executive Committee (March 2013).
- McCouch, S., G. J. Baute, J. Bradeen, P. Bramel, P. K. Bretting, E. Buckler,J. M. Burke et al. "Agriculture: Feeding the Future." *Nature* 499, no. 7456 (2013): 23–24.
- National Plant Genome Initiative Workshop, Plant and Animal Genome Conference (January 2013).
- Plant Science Research Summit, American Society of Plant Biologist (March 2013).
- Rosaceae Community Plant Genome Research Priorities (May 2013).
- Agriculture and Food Research Initiative (AFRI), National Institute of Food and Agriculture (NIFA) Stakeholder Listening Session (February 2012).
- National Research Council. *Research Universities and the Future of America: Ten Breakthrough Actions Vital to Our Nation's Prosperity and Security*. Washington, D.C.: The National Academies Press, 2012.
- Proposed U.S. Wheat Genomics Research Priorities, National Wheat Improvement Committee, Wheat Genomics Subcommittee (November 2012).
- Report to the President on Agricultural Preparedness and the Agriculture Research Enterprise, The President's Council of Advisors on Science and Technology (PCAST) (December 2012).
- Grand Challenges, Crop Science Society of America (2010).
- National Research Council. *Toward Sustainable Agricultural Systems in the 21st Century*. Washington, D.C.: The National Academies Press, 2010.
- Plant and Pest Systems, NIFA Stakeholder Workshop (April 2010).
- Soybean Genomics Research Program Strategic Plan: Implementing Research to Meet 2012–2016 Strategic Milestones (October 2010).

Appendix D: Sequenced Plant Genomes

	Plant	Species Name (genome size)	Family	Major Finding(s)	Publication
1	**Apple**	*Malus x domestica* Borkh cultivar “Golden Delicious” (742 Mb)	Roseaceae (roses)	Discovered gene families that control fruit development. 57,000 genes identified.	Nature Genetics 42, 833–839 (2010)
2	**Barley**	*Hordeum vulgare* L. (5.1 Gb)	Poaceae (grasses)	Provided an important new resource for genetic diversity for genome-enabled crop improvement. Estimated 30,000 genes total.	Nature 491, 711–716 (2012)
3	**Brachypodium**	Brachypodium distachyon (270 Mb)	Poaceae (grasses)	Expanded knowledge on grass evolution; emerging grass model system. 25,000 protein coding genes identified.	Nature 463, 763–768 (2010)
4	**Bread wheat**	*Triticum aestivum* (17 Gb)	Poaceae (grasses)	More than 90,000 genes were identified, significant loss of gene family members on domestication, expansion of growth and metabolism genes was noted.	Nature 491, 705–710 (2012)
5	**Cacao**	*Theobroma cacao* (430 Mb)	Malvaceae (mallows)	Enabled genome-assisted breeding, discovered expansion of gene that impact cocoa flavor (e.g., flavonoid and fatty acid synthesis genes).	Nature Genetics 43, 101-108 (2011)
6	**Castor bean**	Ricinus communis (350 Mb)	Euphorbiaceae (spurge)	Provided new information on the ricin toxin and lipid biosynthetic gene families. 31,000 genes predicted.	Nature Biotechnology 28, 951–956 (2010)
7	**Chickpea**	*Cicer arietinum* L. cultivar CDC Frontier (740 Mb)	Fabaceae (legumes)	Sequence comparison with 90 germplasm accessions shed light on breeding-associated genetic sweeps and selection (i.e., domestication). 28,000 genes identified.	Nature Biotechnology 31, 240–246 (2013)
8	**Chinese cabbage**	*Brassica rapa* ssp. *pekinensis* line Chiifu (284 Mb)	Brassicaceae (mustard)	Variable gene loss among triplicated genome segments may explain morphological plasticity in the genus. 41,000 genes identified.	Nature Genetics 43, 1035–1039
9	**Corn**	*Zea mays* ssp. *mays* L B73 (2.3 Gb)	Poaceae (grasses)	First detailed view of the large and repetitive genome of maize; advanced maize genome-assisted breeding and basic research. Estimated 32,000 genes identified.	Science 326, 1112–1115 (2009)
10	**Cotton**	*Gossypium raimondii* (760 Mb)	Malvaceae (mallows)	Important diploid reference genome for allopolyploid cotton. Estimated 37,000 genes identified.	Nature 492, 423–427 (2012)
11	**Cucumber**	*Cucumis sativus* var. *sativus* L.(370 Mb)	Cucurbitaceae (cucurbits)	Elucidated the evolution of genome structure in the Cucurbitaceae and provided insight into agronomically important traits.	Nature Genetics 41, 1275–1281 (2009)
12	**Medicago**	*Medicago truncatula* (375 Mb)	Fabaceae (legumes)	Whole genome duplication allowed the emergence of critical components for rhizobial nodulation. 62,000 predicted genes.	Nature 480, 520–524 (2011)
13	**Melon**	*Cucumis melo* L. DHL92 (450 Mb)	Cucurbitaceae (cucurbits)	Low number of known disease resistance genes. Candidate genes for fruit quality identified. 27,000 protein coding genes predicted.	PNAS 109, 11872–11877 (2012)

14	**Papaya**	*Carica papaya* cultivar SunUp (372 Mb)	Caricaceae	Genome sequence of a genetically engineered commercial crop. 24,000 protein coding genes predicted.	Nature 452, 991–996 (2008)
15	**Peach**	*Prunus persica* L. Batsch cultivar Lovell (270 Mb)	Roseaceae (roses)	Shed light on domestication by comparisons with 14 resequenced peach accessions. 27,000 protein coding genes predicted.	Nature Genetics 45, 487–494 (2013)
16	**Pigeonpea**	*Cajanus cajan* L. Asha (830 Mb)	Fabaceae (legumes)	Shed light on the role that different gene families, such as drought-tolerance genes, played during domestication in India. 48,000 protein coding genes predicted.	Nature Biotechnology 30, 83–89 (2012)
17	**Potato**	*Solanum tuberosum* L. DM (840 Mb)	Solanaceae (nightshades)	Sequence variations are frequent and likely cause inbreeding depression. 39,000 protein coding genes predicted.	Nature 475, 189–195 (2011)
18	**Rock cress**	*Arabidopsis lyrata* MN47 (210 Mb)	Brassicaceae (mustard)	68% larger in genome size than *A. thaliana* due mostly to small deletions in the smaller *A. thaliana* genome. 32,000 predicted genes.	Nature Genetics 43, 476–481 (2011)
19	**Sorghum**	*Sorghum bicolor* L. Moench BTx623 (730 Mb)	Poaceae (grasses)	24% of genes are grass-specific and 7% are sorghum-specific. 27,000 protein coding genes predicted. Advanced genome-assisted breeding and research for this important food, feed, and energy crop.	Nature 457, 551–556 (2009)
20	**Soybean**	*Glycine max* var. Williams 82 (1.1 Gb)	Fabaceae (legumes)	Ancient polyploid genome. 46,000 protein coding genes predicted. 78% of predicted genes located in chromosome ends. Advanced soybean genome-assisted breeding and basic research.	Nature 463, 178–183 (2010)
21	**Soybean (wild relative)**	*Glycine soja* var. IT182932 (900 Mb)	Fabaceae (legumes)	0.31% DNA sequence difference between wild and cultivated soybeans. *Glycine soja* and *Glycine max* divergence predates domestication.	PNAS 107, 22032–22037 (2010)
22	**Strawberry (woodland species)**	*Fragaria vesca* ssp. *vesca* Hawaii 4 (240 Mb)	Roseaceae (roses)	Allowed the identification of horticultural traits such as flavor, nutritional value, and flowering time. 33,000 protein coding genes identified.	Nature Genetics 43, 109–116 (2011)
23	**Tomato**	*Solanum lycopersicum* cultivar 'Heinz 1706' (900 Mb)	Solanaceae (nightshades)	Genome triplication and allelic variation may explain the generation of variation in agronomic traits in Solanaceae. 31,000 protein coding genes estimated for tomato.	Nature 485, 635–641 (2012)
24	**Tomato (wild relative)**	*Solanum pimpinellifolium*	Solanaceae (nightshades)	Closest wild relative to tomato; 0.6% nucleotide divergence but 8% divergent from potato.	Nature 485, 635–641 (2012)

Appendix E: Databases and Tools Currently Supported Through the National Plant Genome Initiative (NPGI)

Database	Link
Arabidopsis Interactome Network Map	http://interactome.dfci.harvard.edu/A_thaliana/index.php?page=home
Biofuel Feedstock Genomics Resource	http://bfgr.plantbiology.msu.edu
Cacao Genome Database	www.cacaogenomedb.org
CottonGen	www.cottongen.org
Genome Database for Rosaceae (GDR)	www.rosaceae.org
Germplasm Resources Information Network (GRIN)	www.ars-grin.gov/npgs/searchgrin.html
Dendrome	http://dendrome.ucdavis.edu
Epigenetic natural variation in *Arabidopsis thaliana*	http://chromatin.cshl.edu/cgi-bin/gbrowse/epivariation
Gene Expression Omnibus (GEO)	www.ncbi.nlm.nih.gov/geo
GrainGenes	http://wheat.pw.usda.gov/GG2/index.shtml
Gramene	www.gramene.org
ionomicHUB (iHUB)	www.ionomicshub.org/home/PiiMS
iPlant Collaborative (iPlant)	www.iplantcollaborative.org
KnowledgeBase (KBase)	http://kbase.us/
Legume Information System (LIS)	www.comparative-legumes.org
Maize Genetics and Genomics Database (MaizeGDB)	www.maizegdb.org
Maizesequence.org	www.maizesequence.org/index.html
Papayabase	www.papayabase.org/
PhaseolusGenes	http://phaseolusgenes.bioinformatics.ucdavis.edu
Phytozome	www.phytozome.net
PlantGDB	www.plantgdb.org
Plant Genomes Central	www.ncbi.nlm.nih.gov/genomes/PLANTS/PlantList.html
Plant Metabolic Network (PMN)	http://plantcyc.org
Plant Ontology Consortium	http://plantontology.org
PlexDB	www.plexdb.org
Short Read Archive	www.ncbi.nlm.nih.gov/sra
Salk Institute Genomic Analysis Laboratory (SIGnAL)	http://signal.salk.edu
Solanaceae Genomics Resources	http://solanaceae.plantbiology.msu.edu
Solanaceae Coordinated Agricultural Project (SolCAP)	http://solcap.msu.edu
Sol Genomics Network (SGN)	http://solgenomics.net
SoyBase	http://soybase.org
The Arabidopsis Information Resource (TAIR)[79]	http://arabidopsis.org
The Triticeae Toolbox (T3)	http://triticeaetoolbox.org

---

[79] TAIR is scheduled to be replaced by the Arabidopsis Informatic Portal (AIP) in 2015.

Appendix F: Developing Country Impacts of National Plant Genome Initiative (NPGI)-Supported Resources

U.S. Agency for International Development (USAID) Consultative Group on International Agricultural Research (CGIAR) and NPGI Deploy Genomics for Crop Improvement in Developing Countries

In 2008, the CGIAR system began establishing multi-center CGIAR Research Programs (CRPs) that focus on five-year, results-oriented plans to solve major issues in agricultural development for poor farmers. The new CRPs are leveraging 13 years of investment in crop genomics by the NPGI to accelerate the development of higher yielding and more disease- and climate-resilient crop varieties. Several examples of highly successful NPGI CGIAR partnerships are described below. A full list of the CRPs with their project documents can be found at the CGIAR Fund website (www.cgiarfund.org/research_portfolio). The U.S. Government's global hunger and food security initiative "Feed the Future" supports nine out of 15 CRPs.

Food Security Innovation Center Program Area	CRP	Objectives	Lead Center
Climate-Resilient Cereals	3.1 Wheat	International Maize and Wheat Improvement Center (CIMMYT) program goals include boosting farm-level wheat productivity and fortifying crop resistance to globally important diseases and pests, enhancing adaptation to warmer climates, and reducing water, fertilizer, labor, and fuel requirements. The CIMMYT marker assisted wheat breeding program is enhanced by NPGI investments and collaborations in wheat genotyping and wheat genome sequencing. CIMMYT collaborates with NPGI-funded U.S. Department of Agriculture (USDA) and university scientists and the NPGI-funded databases GrainGenes and Gramene in design and planning, software development, and data analysis.[80, 81]	CIMMYT (International Maize and Wheat Improvement Center)
Climate-Resilient Cereals	3.2 Maize	A major focus at CIMMYT is to double the productivity of corn, make it more resilient to climate change, and reduce fertilizer, water, and labor costs. The CIMMYT marker-assisted corn breeding program leverages NPGI investments in the corn genome sequence and genotyping of corn breeding germplasm to accelerate and increase the efficiency of its breeders. CIMMYT collaborates with NPGI-funded USDA and university scientists for corn genotyping and phenotyping data production and analysis (15K/40K landraces and breeding lines genotyped to date). Genotyping results are hosted by the NPGI-funded databases Maize Genetics and Genomics Database (MaizeGDB)and Gramene.[82, 83]	CIMMYT

---

[80] www.gramene.org/collaborators.

[81] www.maizegdb.org/cooperators.php

[82] Gramene collaboration with CIMMYT, IRRI and IRIS, www.gramene.org/collaborators.

[83] MaizeGDB collaboration with CIMMYT, www.maizegdb.org/cooperators.php.

Climate-Resilient Cereals	3.3 Global Rice Science Partnership (GRiSP)	A primary objective at the International Rice Research Institute (IRRI) is to increase rice productivity and value for the poor through the accelerated development of improved varieties that are better adapted to climate change and are environmentally resilient. IRRI, the USDA Dale Bumpers National Rice Research Center, and university scientists share germplasm and associated data and knowledge of that germplasm. IRRI collaborates with the NPGI-funded Gramene database in designing and planning the database and creating software.[84]	IRRI
Advanced Approaches for Pests and Diseases	3.4 Roots, Tubers, and Bananas	An important goal of the breeding program at the International Institute of Tropical Agriculture (IITA)is conserving and accessing genetic resources and accelerating the development and selection of varieties with higher, more stable yield and added value. IITA, USDA, and university scientists are working together to deploy genomic technologies for cassava breeding. Molecular markers are developed and used for breeding by genomic selection.[85]	IITA
Productive Legumes	3.5 Grain Legumes	The International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) is applying molecular-marker-based strategies for crop improvement to maximize the benefits that grain legumes offer to smallholder farmers, especially women, by increasing income, securing food supplies, improving nutrition, and sustainably intensifying farming systems. ICRISAT collaborates with the NPGI-funded Legume Information System (LIS) database on chickpea and pigeon pea genome sequence data production, data analysis, the development of markers for breeding, and hosting data for the broadest community access.[86]	ICRISAT
Climate-Resilient Cereals	3.6 Dryland Cereals	This program is focused on improving the productivity of four major cereal crops: barley, finger millet, pearl millet and sorghum. It leverages NPGI investment in the sorghum genome sequence to produce well-adapted, higher yielding varieties with better disease resistance and tolerance to abiotic stresses for the drylands of Asia and Africa.	ICRISAT

CGIAR-university linkage program. CGIAR fosters partnerships between U.S. universities and CGIAR centers in agricultural genetic and genomic research.

Improved Maize for African Soils (IMAS). After drought, the major challenge for African maize farmers is low-fertility soil with poor nitrogen-supplying capacity. The goal of this project is to use the most advanced genetic technologies, transgenics, and marker-assisted breeding methods to address nitrogen deficiency in African maize production. Researchers at Pioneer Hi-Bred,

---

[84] www.cgiar.org/cgiar-consortium/research-centers/international-rice-research-institute-irri.

[85] ARS collaboration with IITA for cassava breeding by genomic selection, www.nextgencassava.org/genomic_select.html.

[86] LIS collaboration with ICRISAT, www.comparative-legumes.org/pages/collaborators.

CIMMYT, and partner institutions in Kenya and South Africa are currently identifying DNA markers that correlate with nitrogen use efficiency (NUE) in maize.

Cassava brown streak virus resistance. The IITA is working with National Agricultural Research Systems (NARS) in Africa and the Donald Danforth Plant Science Center to utilize the most advanced genetic technologies—transgenic and marker-assisted breeding methods—to develop cassava resistance to the cassava brown streak virus. Resistance genes have been identified and are currently being tested in confined crop field trials.

Potato late blight resistance and sweet potato weevil resistance. Through genomics-assisted conventional breeding and biotechnology, CGIAR supports the International Potato Center (CIP) to develop varieties of potatoes in Peru and Africa that are resistant to the fungal disease that causes late blight. CIP is working with NARS in Africa to develop sweet potato resistance to weevils through biotechnology approaches. Significant advancement has been made to expedite the discovery of new molecular variants for resistance.

Banana bacterial wilt resistance. Little natural resistance to Banana Xanthomonas Wilt (BXW) disease exists in the narrow germplasm base for the East Africa highland banana. To tackle this problem, IITA is working with NARS in East Africa and the Agricultural Biotechnology Support Program (ABSP) and has identified genes in peppers that confer resistance to BXW.

Appendix G: Glossary

Epigenome is the record of the chemical modifications to the DNA and histone proteins of an organism's genome. Epigenetic modifications are reversible modifications on a cell's DNA or histones that affect gene expression without altering the DNA sequence.

GBit/sec is the term used for gigabit per second or 10^9 bits per sec.

GBS or genotyping by sequencing is an important, cost-effective tool for plant genome sequence analysis that uses modern "next-generation" DNA sequencing technologies for breeding, population studies, germplasm characterization, and trait mapping in diverse organisms, including those that lack a reference genome sequence.

Genebank is a biorepository for the conservation and distribution of a wide diversity of genetic resources important for agriculture and industry. Genebank curators and researchers increase the value and utility of conserved materials and associated information by characterizing their genetic contents and evaluating their merits.

Genome is an organism's complete set of hereditary information or DNA sequence.

Genome editing is a form of genetic engineering in which DNA is modified in situ—inserted, replaced, or removed from the genome—using tailored nucleases or "molecular scissors."

Genomics is a discipline in genetics that applies DNA sequencing methods and bioinformatics to sequence, assemble, and analyze the function and structure of the complete genome of an organism.

Genotype is the genetic makeup of a cell, tissue, organ, or whole organism.

Metabolome is a comprehensive profile of the low molecular weight compounds of a plant cell, tissue, or whole organism.

NPGI is the National Plant Genome Initiative, a coordinated national research program created by the Interagency Working Group on Plant Genomes (IWGPG) – U.S. Department of Agriculture (USDA), National Science Foundation (NSF), U.S. Department of Energy (DOE), Office of Science and Technology Policy (OSTP), and Office of Management and Budget (OMB) in 1998.

Phenotype is a composite of a plant's morphological, physiological, biochemical, and agronomic traits that is determined by its genome sequence and its interaction with the environment.

Photosynthetic efficiency is the fraction of light energy converted into chemical energy during photosynthesis in a plant.

Proteome is the complete set of proteins expressed in a cell, tissue, organ, or whole organism.

RNA interference is the RNA-mediated silencing of gene expression in a cell, tissue, organ, or whole organism. RNA interference is also referred to as post-transcriptional gene silencing (PTGS).

Transcriptome is the set of all RNA molecules, including mRNA, rRNA, tRNA, and other non-coding RNAs produced in a cell, tissue, organ, or organism.

Appendix H: Acronyms

2D	two-dimensional
3D	three-dimensional
ASPB	American Society of Plant Biologists
AFRI	Agricultural Food and Research Initiative
AIP	Arabidopsis Informatic Portal
ARPA-E	Advanced Research Projects Agency – Energy
ARS	Agricultural Research Service
ASPB	American Society of Plant Biologists
BBSRC	Biotechnology and Biological Sciences Research Council
BER	DOE Office of Biological and Environmental Research
BES	DOE Office of Basic Energy Sciences
BESC	BioEnergy Science Center
BFGR	Biofuel Feedstock Genomics Resource
BHEARD	Borlaug Higher Education Agriculture Research Development Program
BRC	Bioenergy Research Center
BREAD	Basic Research to Enable Agricultural Development
BXW	Banana Xanthomonas Wilt
CAP	Coordinated Agricultural Project
CGIAR	Consultative Group on International Agricultural Research
CIMMYT	International Maize and Wheat Improvement Center
CIP	International Potato Center
CIPRES	Cyberinfrastructure for Phylogenetic Research
CoGe	Comparative Genomics (software platform)
CORE	Collaborative Oat Research Enterprise
CoS	Committee on Science
CPGR	Comprehensive Phytopathogen Genomics Resource
CRISPR	clustered regularly interspaced short palindromic repeats
CRP	CGIAR Research Program
CSHL	Cold Spring Harbor Laboratories
CTFS	Center for Tropical Forest Science
D2P	Data-to-Publication
DFID	Department for International Development
DOE	U.S. Department of Energy
DRRW	Durable Rust Resistance in Wheat
EBI	European Bioinformatics Institute
EOT	education, outreach and training
EPA	Environmental Protection Agency
ESnet	Energy Sciences Network
EU	European Union
GBS	genotyping-by-sequencing
GDR	Genome Database for Rosaceae
GEM	Germplasm Enhancement of Maize
GEO	Gene Expression Omnibus
GLBRC	Great Lakes Bioenergy Research Center
GO	Gene Ontology
GRIN	Germplasm Resources Information Network
GRiSP	Global Rice Science Partnership
ICARDA	International Center for Agricultural Research in the Dry Areas
ICRISAT	International Crops Research Institute for the Semi-Arid Tropics

iHub	ionomicHUB
IITA	International Institute of Tropical Agriculture
IMAS	Improved Maize for African Soils
iPG2P	iPlant Genotype to Phenotype
IPGA	Improving Plant Genome Annotation
iPlant	iPlant Collaborative
iPToL	iPlant Tree of Life
IRRI	International Rice Research Institute
IWGPG	Interagency Working Group on Plant Genomics
JEBI	Joint BioEnergy Institute
JGI	Joint Genome Institute
KBase	KnowledgeBase
LIS	Legume Information System (ARS, NCGR)
LSSC	Life Sciences Subcommittee
MAGE	Multiplex Automated Genome Engineering
MAGIC	Multiparent Advanced Generation Intercross
MaizeGDB	Maize Genetics and Genomics Database
MAS	marker-assisted selection
MCA-PGR	Mid-Career Investigator Awards in Plant Genome Research
N	Nitrogen
NAM	Nested Association Mapping
NARS	National Agricultural Research Systems
NCBI	National Center for Biotechnology Information
NCGR	National Center for Genome Resources
NGO	non-governmental organization
NIFA	National Institute of Food and Agriculture
NIH	National Institutes of Health
NIL	near isogenic line
NOAA	National Oceanic and Atmospheric Administration
NPGI	National Plant Genome Initiative
NPGS	National Plant Germplasm System
NSF	National Science Foundation
NSTC	National Science and Technology Council
NUE	nitrogen use efficiency
OMB	Office of Management and Budget
OSTP	Office of Science and Technology Policy
P	Phosphorus
PBG	plant breeding and genomics
PBGCoP	Plant Breeding and Genomics Community of Practice
PBTN	Plant Breeding Training Network
PCAST	President's Council of Advisors on Science and Technology
PEER	Partnerships for Enhanced Engagement in Research
pENCODE	Plant Encyclopedia of DNA Elements
PETRO	Plants Engineered to Replace Oil
PGRP	Plant Genome Research Program
PMN	Plant Metabolic Network
PRACE	Partnership for Advanced Computing in Europe
PTGS	post-transcriptional gene silencing
SC	DOE Office of Science
SGN	Sol Genomics Network
SI	Smithsonian Institution

SIGnAL	Salk Institute Genomic Analysis Laboratory
SIGEO	Smithsonian Institution Global Earth Observatory
SoyKB	Soybean KnowledgeBase
STEM	Science, Technology, Engineering, and Mathematics
T3	The Triticeae Toolbox
TAIR	The Arabidopsis Information Resource
TALEN	transcription activator-like effector nuclease
TASSEL	Trait Analysis by Association, Evolution, and Linkage
TGI	The Genome Institute at Washington University
transPLANT	Trans-national Infrastructure for Plant Genomic Science
U.S.C.	United States Code
USAID	U.S. Agency for International Development
USDA	U.S. Department of Agriculture
USFS	U.S. Department of Agriculture, Forest Service
XSEDE	Extreme Science and Engineering Discovery Environment