

*This paper was presented at a colloquium entitled “Protecting Our Food Supply: The Value of Plant Genome Initiatives,” organized by Michael Freeling and Ronald L. Phillips, held June 2–5, 1997, sponsored by the National Academy of Sciences at the Arnold and Mabel Beckman Center in Irvine, CA.*

## Potentials of the National Corn Genome Initiative

(*Zea mays* L./genomics/expressed sequence tags/synteny/physical mapping)

E. H. COE\*

United States Department of Agriculture, Agricultural Research Service, Plant Genetics Research Unit, and Department of Agronomy, University of Missouri, Curtis Hall, Columbia, MO 65211-7020

**ABSTRACT** The present paper summarizes future needs in information and tools, technology, infrastructure, training, funding, and bioinformatics, to provide the genomic knowledge and tools for breeding and biotechnological goals in maize. The National Corn Genome Initiative (NCGI) has developed through actions taken by the National Corn Growers Association (NCGA) and participation in a planning process by institutions, companies, and organizations. At the web address for the NCGI, <http://www.inverizon.com/ncgi>, are detailed analyses of goals and costs, impact and value, and strategy and approaches. The NCGI has also produced an informative and perceptive video suitable for public groups or schools, about agricultural contributions to life and the place of maize in these contributions. High potential can be expected, from cross-application of knowledge obtained in maize and other cereals. Development of information and tools for all crops, whether monocots or dicots, will be gained through an initiative, and each crop will be positioned to advance with cost-effective parallels, especially for expressed sequences, markers, and physical mapping.

Cross-application of knowledge among crop species is the potential of this colloquium. In the beginnings of molecular mapping, reflecting this potential, markers were applied with much success across species. For example, maize probes were early used to begin mapping in the forage grass tall fescue (1), and they contributed to mapping in sorghum (2), wheat (3), and rice (4, 5), among others. Of course, before that there had been an oft-repeated theme of cross-species research revealing long-understood facts of common genetic mechanisms; of homology between morphological parts; of parallel themes in developmental processes; and of universality of biochemical processes, as much as in today's molecularly determined homology or orthology and its partner, synteny.

This paper is a summary of the derivation, perspectives, and projections of the National Corn Genome Initiative (NCGI), and some considerations of the potentials for all crops. Steady interest and encouragement from the National Corn Growers Association (NCGA), aided by James McLaren of the consultant firm Inverizon, Inc., has led to a fleshed-out plan for the NCGI, through participation of institutions, companies, and organizations in the planning process. At the web address for the NCGI, <http://www.inverizon.com/ncgi>, are detailed analyses of goals and costs, impact and value, strategy and approaches, and an appeal for advocacy. The NCGA has also produced an informative and perceptive video suitable for public groups or schools, about agricultural contributions to life and the place of maize in these contributions. This video may be obtained from NCGA, Attention: Amy Mercer, 1000

Executive Parkway, Suite 105, St. Louis, MO 63141. Capture of the public imagination is key to advancement of plant science, and advancement of plant science is key to U.S. agricultural leadership and its economic foundation. The NCGA, by spearheading this appeal, is performing a visionary service to science and to U.S. and world need, for which strong credit must be given.

The NCGA document opens with this quotable outlook:

The past and the future, have already been written—in the genetic code. If only we could afford to read the book.

In March 1997, based upon groundwork developed from three years of planning and design, the following Overview and Request for Action were adopted:

The National Corn Growers Association believes that the future of corn is written in corn's genetic code. The National Corn Genome Initiative, supported by NCGA, private sector agribusiness, and trade associations, calls for federal funding of \$143 million, over a five year period, for corn genome mapping. A corn genome map will give us the particular locations of genes that control important traits (such as stress tolerance, resistance to disease or pests, yield and nutritional value). These genes can then be incorporated into improved crop varieties to address a diverse range of producer and consumer problems and needs. To compete in the global corn market, U.S. agriculture must continually strive to efficiently and economically improve corn production capabilities—such as combating serious threats from disease, pests, and climate changes—without harming the environment. Modern biotechnology through plant genome mapping, sequencing and trait identification holds the key to achieving this goal. . .

Aside from protecting U.S. interests and access to important biotechnology and gene patents, the benefits of a corn genome map include:

- significant reductions in crop losses and reliance on pesticides
- improved nitrogen-use efficiency limiting potential for nitrates in the water supply
- improved animal nutrition leading to healthier meat and increased meat productivity
- significant reductions in environmental problems confronted by livestock producers such as modifying the digestibility of phosphorus in feed corn

Abbreviations: cDNA, DNA complementary to messenger RNA; EST, expressed sequence tag; NCGA, National Corn Growers Association; NCGI, National Corn Genome Initiative.

\*To whom reprint requests should be addressed at: USDA/ARS/PGRU, 210 Curtis Hall, University of Missouri, Columbia, MO 65211-7020. e-mail: [ed@teosinte.agron.missouri.edu](mailto:ed@teosinte.agron.missouri.edu).

improved yields by 3–4% and resistance to diverse environmental conditions, such as drought

a 20% increase in production efficiency within 10 years adding at least \$4 billion in added farm value per year

laying the groundwork for similar advances for other cereal crops.

#### NCGA REQUEST FOR ACTION (March 1997)

Congress should set aside \$10 million in Agricultural Research Service Funds for the National Corn Genome Initiative. (FY98 Ag Appropriations)

Congress should establish a new “genome mapping” category within the competitively awarded National Research Initiative and appropriate not less than \$10 million. (FY98 Ag Appropriations)

Congress should ensure that not less than one-half of the research funds in the Fund for Rural America be set-aside for basic research projects and that priority be given to plant and animal genome mapping projects (Farm Bill Research Title Reauthorization)

Current events and activities of the NCGA, including information on responses in Congress and the Administration, are described in the Web document.

The route to this fleshed-out plan has involved three years of communication, discussion, shared thinking, and compromise among scientists and managers in the public sector and in biotechnology and seed trade industries, feed scientists and feed industry, grain producers, and grain marketers and processors. While the basic science defines what can be done, the fact that maize is a commodity of such economic importance defines the breadth of the stakeholders for the NCGI.

At a small group meeting in 1994, Bob Mustell of the NCGA stated, “The most important thing we could do is map the corn genome.” A small group of scientists met soon after, to consider scientific priorities and to plan how to proceed. The perspectives developed by that group of scientists have changed only to the extent that technologies have opened alternative or more efficient choices, but the objectives and the framework of highest-priority goals have not undergone significant modification. An action and planning group met a number of times subsequently and through 1995, developing two workshops that were held in 1996, involving scientists in genomics, breeding, informatics, biochemistry, physiology, food processing, and nutrition; and involving interests in production, marketing, and processing. The workshop outcomes are presented in the Web, and the 1997 document is constructed from those inputs. The Executive Summary of the document defines the mission, objective, goal, management, and opportunity:

The mission is to provide a national resource which will support and stimulate sustainable economic advances in crop production, with the major focus being on corn as a key driver of the U.S. agricultural sector.

The objective is to generate a corn genome map which can be utilized by the corn and other crop industries in the U.S. to enhance the amount and the quality of produce per acre.

The goal is to implement a directed and coordinated program to clone, sequence and map the approximately 50,000 genes which control growth, development, yield and quality in corn. There would also be associated work related to gene expression information, maintaining physical stocks, and the development of a computer-based informatics system to store, utilize and retrieve usable data.

A management group should be established with responsibility to:

develop direction on the most cost-effective approaches to be taken

request and review competitive grant applications

coordinate the competitive funding to ensure appropriate coverage without duplication

manage information inputs from participating laboratories

obtain intellectual property protection on behalf of the Government

manage access to the results in accordance with the agreed procedures

provide advice to other funding bodies on the key leverage points for application

serve as a source of reference to leverage the results into other major crops

The opportunity is that the genes of corn can be more precisely located than is possible today, that the function of many genes can be determined and tracked to a location, and that the use of this information will transform corn breeding into a pipeline which can deliver the products needed for the global customers of the 21st century. In other words, the vision is that with such a corn genome map, the U.S. can maintain a leadership position in world agricultural production and in environmental stewardship.

A few questions deserve consideration, relevant to the prospects for meeting the needs of multiple crops through a maize initiative:

(i) For the NCGI, \$143 million is requested by the NCGA. Is this sufficient for achievement of the goals, specifically for maize? Is it sufficient for extensions to other species? Are multiple species a necessity to an initiative grounded in maize, an extension or a luxury?

(ii) Is the available population of potential scientist participants sufficient for an initiative on this scale? Some of the most consuming tasks will likely be done in shops that are designed and geared up for comparable work in humans or in other species and will not require crop-specific participation; others are very biology dependent and will require specialists, the number of whom is limiting.

(iii) What are the opportunities for research to be undertaken in various crops consonant with a maize initiative? Because each species has uniquely accessible components of the puzzle (for example, rapidly developing physical maps in certain species and well developed trait analyses in others), multiple roles and mutual benefit will contribute most effectively.

(iv) What will be done with the knowledge and tools? Of high potential is the prospect for increasing our ability to utilize existing, unexploited germ plasm to contribute to productivity. Preparation for that, however, requires trait evaluations of diverse materials and characterizations of their genomic variability. Unfortunately, technology for measurement and identification of desired traits is costly and time limiting at the precision necessary for genetic analysis—will parallel development of efficient plant, field, and product measurement techniques be needed?

(v) Will sufficient knowledge develop in parallel, for metabolic pathways, biochemical constituents, and regulation of genes, to allow the most advantage to be taken of the knowledge resource? Pathways and constituents that are in common among species, a long-recognized and widely exercised potency, are a salient paradigm for gene and genomic functions in common, and regulation is now beginning to be recognized to be a part of the orthology of systems among species. Expansion of such cross-species information is integral, and vital, to an initiative.

(vi) Will concord among researchers in different crop commodities be sufficient to draw national support for a mutually planned effort? Conversely, will the plan affect prospects for concord?

While not responding entirely to the above questions, considerations from the NCGI planning process, and some drawn

from discussions in this colloquium, can be distilled to the following research needs for the future, to make rapid progress in a comprehensive manner. Because the success of a genomics effort depends more on flexible guidance by a steering group than on absolutes, these briefs represent needs that appear today to be the highest priorities, and may well be reprioritized as changes are seen to be needed.

### Information and Tools Needed

One goal of a plant genome program may ultimately be a complete sequence of a cereal genome and of other selected crop genomes, achieved through international cooperation. The first-priority needs, however, are to derive planning and prioritization procedures, and to initiate those essential efforts that not only will enable serious applications of genomic information and tools but also will set the stage for sequencing and its definitive impact on crop genomics.

Seven first-priority essentials for an effective plant genome program are as follows:

(i) A large number of cDNA segments (e.g., 50,000–250,000) representing expressed genes should be isolated, in part from maize and in part from other cereals; sequenced and assembled (i.e., distinguished as unique vs. overlapping); located to position on genetic or physical maps; and defined to expression according to tissues, cell types, developmental stages, and targeted response conditions.

(ii) Efficient, high-density, robust markers for trait mapping and tracking, placed both genetically and physically, must be developed.

(iii) A physical map of the maize genome, constructed through large-fragment overlaps and/or cytogenetic methods, anchored with sequenced clones and matched with physical maps of one or more smaller-genome nodal species (e.g., rice or sorghum), is essential.

(iv) The power and utility of cross-species mapping (through synteny) and for predicting gene sequence and function in one species from another (synteny and orthology) (5–7) must be tested firmly. Expressed sequence tag (EST) clones, from cDNAs, in particular can be expected to contribute most effectively to this priority (4).

(v) A comprehensive resource is needed, of transposable-element-derived gene modifications from which targeted genes can be selected (8, 9).

(vi) Stocks and materials maintenance, organization, and distribution; databases and informatics; education and outreach are integral.

(vii) Need should be anticipated by scaling up of research in the areas of metabolic pathways, steps, dynamics, and flux, for substantial advancements in knowledge stimulated by the plant genome program.

### Technology Needed

(i) To selectively sequence genes of interest from any species without having to sequence large amounts of noncoding regions found in large genomes, new means are needed. This technology would facilitate the sequencing of orthologous gene family members from different species, and from gene variants found within germ plasm collections.

(ii) Efficient methods are needed for study of differences as small as single-base pairs, to be applied in more-direct approaches to genes that affect measurable traits (quantitative trait loci, QTLs).

(iii) Highly effective approaches are needed by which gene expression can be categorized and defined explicitly. These include comprehensive methods (e.g., arrays); methods that display function in place in tissues; methods for deriving clones for targeted functions or targeted traits; and tests of efficacy

and reliability of applications of transformation into bacteria, yeast, and other species.

(iv) More-precise techniques and tools, nondestructive whenever possible, are much needed for measurement of traits relevant to productivity.

(v) Programming and consensus are urgently needed on map representation, interspecies comparisons, and metabolic representation. These are, for the most part, informatics dependent and should integrate closely with the biological advancements.

### Infrastructure Needed

(i) A steering group should be established, to guide a multiagency (competitive grants, peer reviewed) program, initially to develop procedures by which planning and prioritization are done, to initiate the program, and to set continuity and representation procedures in motion. Membership should have scientists with strong experience in genomics research and should include private-sector and organizational representation.

(ii) Working groups can be developed by the steering group, to aid in planning and to provide inputs, and to be involved in communication within and outside the program; membership should be fluid, cross-commodity, and cross-discipline.

(iii) Banking and distribution of materials: Ready access should be given to what is available, with prompt response. Precision and accuracy would be expected to meet high standards for the research and utilization community.

(iv) Database functions and continuity (see below).

(v) Development of professional capacity is needed, possibly by coalitions with industry and grantees, for efficient accomplishment of goals.

### Training and Education Efforts Needed

(i) Postdoctoral training must be enhanced in biology of crop species; in genomics; and in informatics, together designed to interlace training in each area.

(ii) Public and stakeholders require information and must be enlisted in the areas of concerns, needs, and value.

(iii) Existing programs (National Science Foundation; Education) should be employed, through intercommunication.

### Funding Needed

Existing resources committed to plant science research in the public sector are modest, considering the value of crops as commodities, yet any projected funding raises questions about what will be their source. Predicated on that, levels of funding appropriate to make rapid progress in a comprehensive manner should be developed without assumptions as to their source. Redistribution of existing plant research funds is neither desirable nor suggested. The level of funding proposed for the NCGI is modest, yet is realistic, and unquestionably will contribute to advances in other cereals. Targeted work on the other cereals, or on dicot crops, will be made less demanding of resources by virtue of the knowledge and tools of maize. In some cases, funding from commodity-specific interests may become a significant source of targeted funds for other crops, stimulated by the results from the NCGI.

### Bioinformatics Needed

(i) For analysis and representation of EST or genomic sequence data, excellent software exists in shops that are conducting these types of effort already, and the data can be accessed, linked, and rearranged as needed for specific crop use. Depending on support to existing commodity databases and on developing capabilities under this program, smooth and trans-

parent connectivity may be expected to the new data and new packages, ensuring immediate currentness of data.

(ii) Databases should be coupled as closely as possible to research, data collection, and resource distribution, guided by biologists, and organized around species or closely related groups of species. Funding needs to be at approximately 15–20% of the program, at a level sufficient to support quality curation, service, outreach, documentation, systems, support, and specialized software where necessary. The program must encourage interaction and mutual development of the databases by the species curators, especially toward harmonizing representation of biological information and the exploitation of synteny and other powerful relationships.

(iii) Database coverage of “orphaned” crops (i.e., ones with no systematization of information currently) should be seriously considered, through interaction and mutual development efforts with specialists in these crops. Crops specific to the developing world should be included.

(iv) Database programs should be reviewed and evaluated regularly, with continued funding dependent upon performance.

(v) Interactions and linkages should be promoted, with databases for nonplant species (human, fruit fly, worms, yeast, bacteria, domestic animals) and with databases that are not specifically genetic (metabolism and biochemistry; commodity groups; germ plasm; botany; etc.).

Paced by the above plan, encompassing the essentials, technologies, infrastructure, and informatics, other crops,

whether monocots or dicots, will be positioned to advance with cost-effective parallels, especially for ESTs, markers, and physical mapping.

Ongoing stimulation and encouragement by Bob Mustell and the NCGA is very much appreciated by myself personally, as it is by the maize research and biotechnology community. Jim McLaren’s efforts in developing the plan and documentation, and in leading the communication and exchanges in workshops, have been central. I am grateful for helpful advice and insights toward the present paper, from Jeff Bennetzen, Sam Cartinhour, Rob Martienssen, Susan McCouch, Jim McLaren, Mike McMullen, Jo Messing, John Mullet, and Mary Polacco.

1. Xu, W. W., Sleper, D. A. & Krause, G. F. (1994) *Crop Sci.* **34**, 246–252.
2. Hulbert, S. H., Richter, T. E., Axtell, J. D. & Bennetzen, J. L. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 4251–4255.
3. Devos, K. M., Moore, G. & Gale, M. (1995) *Euphytica* **85**, 367–372.
4. Moore, G., Foote, T., Helentjaris, T., Devos, K., Kurata, N. & Gale, M. (1995) *Trends Genet.* **11**, 81–82.
5. Van Deynze, A., Nelson, J., Yglesias, E., Harrington, S., Braga, D., McCouch, S. & Sorrells, M. (1995) *Mol. Gen. Genet.* **248**, 744–754.
6. Bennetzen, J. L. & Freeling, M. (1993) *Trends Genet.* **9**, 259–261.
7. Bennetzen, J. L. & Freeling, M. (1997) *Genome Res.* **7**, 301–306.
8. Bensen, R. J., Johal, G. S., Crane, V. C., Tossberg, J. T., Schnable, P. S., Meeley, R. B. & Briggs, S. P. (1995) *Plant Cell* **7**, 75–84.
9. Das, L. & Martienssen, R. (1995) *Plant Cell* **7**, 287–294.