



Designing an Agricultural Genome Program

DETAILS

50 pages | 6 x 9 | PAPERBACK
ISBN 978-0-309-06039-4 | DOI 10.17226/6081

AUTHORS

Board on Biology, National Research Council

BUY THIS BOOK

FIND RELATED TITLES

Visit the National Academies Press at NAP.edu and login or register to get:

- Access to free PDF downloads of thousands of scientific reports
- 10% off the price of print titles
- Email or social media notifications of new titles related to your interests
- Special offers and discounts



Distribution, posting, or copying of this PDF is strictly prohibited without written permission of the National Academies Press. (Request Permission) Unless otherwise indicated, all materials in this PDF are copyrighted by the National Academy of Sciences.

Designing an Agricultural Genome Program

Board on Biology
Board on Agriculture
National Research Council

NATIONAL ACADEMY PRESS
Washington, D.C. 1998

NATIONAL ACADEMY PRESS • 2101 Constitution Avenue • Washington, D.C. 20418

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competencies and with regard for appropriate balance.

This report has been reviewed by a group other than the authors according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

This report has been prepared with funds provided by the Office of Energy Research and the Office of Health and Environmental Research of the US Department of Energy under agreement number DE-FG02-94ER61939.

International Standard Book Number 0-309-06039-7

Additional copies of this report are available from:

National Academy Press
2101 Constitution Avenue, N.W.
Box 285
Washington, DC 20055
800-624-6242; 202-334-3313 in the Washington Metropolitan Area
<http://www.nap.edu>

Copyright 1998 by the National Academy of Sciences. All rights reserved.

Printed in the United States of America

STEERING COMMITTEE

DALE BAUMAN, Cornell University

MICHAEL CLEGG, Chair, University of California, Riverside

RONALD SEDEROFF, North Carolina State University

Science Writer

ROBERT POOL, Arlington, Virginia

Staff

PAUL GILMAN, Acting Director, Board on Biology

MICHAEL PHILLIPS, Director, Board on Agriculture

MARY JANE LETAW, Project Officer

JULIEMARIE GOUPIL, Project Assistant

KATHLEEN BEIL, Project Assistant

BOARD ON BIOLOGY

MICHAEL T. CLEGG, *Chair*, University of California, Riverside
JOHN C. AVISE, University of Georgia, Athens
DAVID EISENBERG, University of California, Los Angeles
GERALD D. FISCHBACH, Harvard Medical School
DAVID J. GALAS, Darwin Technologies, Seattle, Washington
DAVID V. GOEDDEL, Tularik, Inc., San Francisco, California
ARTURO GOMEZ-POMPA, University of California, Riverside
COREY S. GOODMAN, University of California, Berkeley
BRUCE R. LEVIN, Emory University, Atlanta, Georgia
OLGA F. LINARES, Smithsonian Tropical Research Institute, Balboa, Panama
ELLIOT M. MEYEROWITZ, California Institute of Technology, Pasadena
ROBERT T. PAINE, University of Washington, Seattle
RONALD R. SEDEROFF, North Carolina State University
DANIEL S. SIMBERLOFF, Florida State University
ROBERT R. SOKAL, State University of New York, Stony Brook
SHIRLEY M. TILGHMAN, Princeton University
RAYMOND L. WHITE, University of Utah, Salt Lake City

Staff

KATHLEEN BEIL, Project Assistant

COMMISSION ON LIFE SCIENCES

THOMAS D. POLLARD, *Chair*, The Salk Institute for Biological Studies, La Jolla, California
FREDERICK R. ANDERSON, Cadwalder, Wickersham and Taft, Washington, DC
JOHN C. BAILAR, III, University of Chicago
PAUL BERG, Stanford University
JOHN E. BURRIS, Marine Biological Laboratory, Woods Hole, Massachusetts
SHARON L. DUNWOODY, University of Wisconsin, Madison
URSULA W. GOODENOUGH, Washington University, St. Louis, Missouri
HENRY W. HEIKKINEN, University of Northern Colorado, Greeley, Colorado
HANS J. KENDE, Michigan State University, Lansing, Michigan
SUSAN E. LEEMAN, Boston University School of Medicine
THOMAS E. LOVEJOY, Smithsonian Institution, Washington, DC
DONALD R. MATTISON, University of Pittsburgh
JOSEPH E. MURRAY, Wellesley Hills, Massachusetts
EDWARD E. PENHOET, Chiron Corporation, Emeryville, California
EMIL A. PFITZER, Research Institute for Fragrance Materials, Inc., Hackensack, New Jersey
MALCOLM C. PIKE, Norris/University of Southern California Comprehensive Cancer Center, Los Angeles, California
HENRY C. PITOT, III, University of Wisconsin, Madison
JOHNATHAN M. SAMET, John Hopkins University, Baltimore, Maryland
CHARLES F. STEVENS, The Salk Institute for Biological Studies, La Jolla, California
JOHN L. VANDERBERG, Southwest Foundation for Biomedical Research, San Antonio, Texas

Staff

PAUL GILMAN, Executive Director

BOARD ON AGRICULTURE

DALE E. BAUMAN, *Chair*, Cornell University, Ithaca, New York
JOHN M. ANTLE, Montana State University, Bozeman
SANDRA S. BATIE, Michigan State University, East Lansing
MAY R. BERENBAUM, University of Illinois, Urbana
LEONARD S. BULL, North Carolina State University, Raleigh
WILLIAM B. DELAUDER, Delaware State University, Dover
ANTHONY S. EARL, Quarles & Brady Law Firm, Madison, Wisconsin
ESSEX E. FINNEY, Jr., US Department of Agriculture, Mitchelleville,
Maryland
CORNELIA FLORA, Iowa State University, Ames
GEORGE R. HALLBERG, University of Iowa, Iowa City
RICHARD R. HARWOOD, Michigan State University, East Lansing
T. KENT KIRK, University of Wisconsin, Madison
HARLEY W. MOON, Iowa State University, Ames
WILLIAM L. OGREN, University of Illinois, Urbana
GEORGE E. SEIDEL, Jr., Colorado State University, Fort Collins
JOHN W. SUTTIE, University of Wisconsin, Madison
JAMES J. ZUICHES, Washington State University, Pullman

Staff

PAUL GILMAN, Executive Director
MICHAEL J. PHILLIPS, Director
SHIRLEY B. THATCHER, Senior Project Assistant

Preface

In 1993 the National Research Council's Board on Biology established a series of forums on biotechnology. The purpose of the discussions is to foster open communication among scientists, administrators, policy-makers, and others engaged in biotechnology research, development, and commercialization. The neutral setting offered by the National Research Council is intended to promote mutual understanding among government, industry, and academe and to help develop imaginative approaches to problem solving. The objective, however, is to illuminate issues, not to resolve them. Unlike study committees of the National Research Council, forums cannot provide advice or recommendations to any government agency or other organization. Similarly, summaries of forums are precluded from reaching conclusions or recommendations, but instead, are intended to reflect the variety of opinions expressed by the participants. The comments in this report reflect the views of the forum's participants as indicated in the text.

For the first forum, held on November 5, 1996, the Board on Biology collaborated with the Board on Agriculture to focus on intellectual property rights issues surrounding plant biotechnology. The second forum, held on April 26, 1997, was also conducted in collaboration with the Board on Agriculture. On April 3, 1997, Bruce Alberts, President of the National Academy of Sciences, met with U.S. Department of Agriculture (USDA) Acting Under Secretary for Research Cathy Woteki, USDA National Research Initiative Chief Scientist Ron Phillips, and National Science Foundation Assistant Director for Biological Sciences Mary Clutter to discuss a new and expanded plant and animal genome mapping and sequencing project. It was agreed that through the auspices of the

forum on biotechnology the National Research Council could identify issues and obstacles to a broad genome project with numerous plant and animal species as its subjects. An attempt would also be made to identify problem areas and breakthroughs in similar programs (e.g., the human genome project and the Japanese rice genome project).

Participation at the “Forum on Designing an Agricultural Genome Program” by representatives of the U.S. Department of Energy, U.S. Department of Agriculture, National Science Foundation, National Institutes of Health, and Congressional staff suggests that this project is important to many federal bodies. Scientists from industry, academe, and federal agencies shared their experiences in mapping and sequencing programs in diverse areas of genomics including human, mouse, *Arabidopsis*, livestock, and maize.

Examination of genomics has been a focus of other activities of the Board on Biology. In 1988 Bruce Alberts, chair of the Board’s Committee on Mapping and Sequencing the Human Genome at that time, released a report outlining how, when, and why we should map and sequence the DNA in the human genome, and strategies for implementation and management of the project. We hope the present forum proceedings will generate further interest in an agricultural genome project and other issues of biotechnology.

Dale E. Bauman, Chair
Board on Agriculture

Michael T. Clegg, Chair
Board on Biology

Contents

Introduction: Genes and Agriculture	1
Summary of Proceedings	4
Learning from Experience, 4	
Dealing with a Multitude of Genomes, 9	
Organizing the Project, 18	
Appendixes	
A. Program and Discussion Questions	23
B. Forum Agenda	25
C. Participant Biographies	27
Boxes	
1: Maps and Sequences, 6	
2: Considering the Social and Ethical Issues, 10	
3: Saved by the Weed, 12	
4: Gene Chips, 16	
5: Public and Private Genomes, 20	

The National Academy of Sciences is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The National Academy of Engineering was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. William A. Wulf is president of the National Academy of Engineering.

The Institute of Medicine was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Kenneth I. Shine is president of the Institute of Medicine.

The National Research Council was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. William A. Wulf are chairman and vice chairman, respectively, of the National Research Council.

Introduction: Genes and Agriculture

In February, a discovery at the University of Kiel in Germany could potentially affect farmers around the world. German scientists found, in the genome (DNA) of a wild beet, a gene that makes the plant resistant to nematodes, tiny worms that destroy the root systems of such crops as sugar beets, cauliflower, and oilseed rape. This taste of nematodes for plant roots is a particularly expensive one for farmers, costing them an estimated \$100 billion in worldwide crop losses each year, and there is little they can do to fight the worms except rotate their crops and spray them or the soil with pesticides. Now, however, the discovery of the nematode-resistance gene offers hope for an alternative. Classical breeding techniques can be used to transfer that gene into beets, but to get resistance to nematodes into unrelated crops such as corn or soybeans will require the isolation of the gene followed by genetic engineering of these other crops.

Those nematode-resistant crops will be just one member of a coming generation of “supercrops” and “superlivestock,” genetically engineered plants and animals that promise to transform agriculture as profoundly as anything since our distant ancestors learned to save seeds from one year to plant the next. With their toolkits of genetic engineering techniques, agricultural researchers can modify organisms to improve quality and convenience, lower the cost of production, and add a variety of useful traits. Some such products are already on the market. The FlavrSavr™ tomato, for instance, is genetically modified to ripen more slowly so that it can be picked closer to peak ripeness. And Roundup-ready™ soybeans have been altered to be immune to the herbicide Roundup™. Farmers who pay the extra \$5 a bag for seeds of this type can spray their entire field with Roundup and kill only the weeds.

How quickly this transformation of crop production and animal husbandry takes place will depend in large part on how much effort—and what type of effort—is put into basic research on the genetics of crops and livestock. Although plenty of research has already been done, among dozens of different species there has been no overall plan or coordination. Not surprisingly, then, progress toward genetically improved crops and livestock has been slower than many would hope. The gene for resistance to nematodes, for example, took eight years to discover. It was found only because of a lucky recombination of chromosomes when domestic sugar beets were cross-bred with wild beets that have a natural nematode resistance. Even now, although we know that wild beets have a nematode resistance gene we have no idea where it is located in the genome or how to isolate it.

Last fall, three government agencies took an important first step toward a more focused agricultural research effort. The National Science Foundation and the Departments of Agriculture and Energy established a program to track down the entire genetic code, including an estimated 20,000 genes, of *Arabidopsis thaliana*, a member of the mustard family. Although *Arabidopsis* itself has no commercial value—it is a weed—delineating its entire genome will give researchers working on other plants an important foundation for their own studies. It was chosen because working from what is known about the genes of *Arabidopsis*, they will be able to answer genetic questions about other species, such as the location of a particular gene, much more quickly.

Now members of the agricultural research community believe it is time to start a broader project: an agricultural genome program. Building from the base provided by the *Arabidopsis* effort and using expertise and technology developed in other ventures, such as the Human Genome Project, it should be possible to generate a huge—and hugely valuable—amount of information on the genetics of agricultural species in a relatively short time.

With that in mind, the National Research Council (NRC) held a workshop on April 26, “Designing an Agricultural Genome Program.” Catherine Woteki, Acting Under Secretary for Research, Education and Economics at the Department of Agriculture, summarized the purpose of the workshop this way: “We in the scientific community need to provide our best estimates of the scientific gain to be made from these investments [in agricultural genome studies]. Over the last decade we have been making a very major investment in genomic activities of importance to agriculture. We have taken the approach of not setting priorities among the major commodities of interest to agriculture. We have a decade’s worth of experience with that approach.”

But now, she noted, the Department of Agriculture has taken a different tack by joining with the National Science Foundation and the Department of Energy to fund the *Arabidopsis* program. “Having made that decision, let’s take the time to go back and examine what we have learned scientifically from the 40 or 50 different species that we have invested in over this past decade in genome-mapping related activities. Can that help us to define a more focused agenda for

the next five to ten years? Or should we continue with the approach that the USDA has taken, which is investing in the *Arabidopsis* mapping activity and then continuing the across-the-board approach in other plant species?"

Or, as Dale Bauman, chair of NRC's Board on Agriculture, put it: "What are the lessons we have learned to date, and how can we use them?"

In particular, the participants in the workshop were asked to assume that an agricultural genome project will be established and to discuss what features such a project should have in light of experience with the Human Genome Project and other genetic research over the past decade or so. How should resources be allocated among the dozens of commercially important species? Which scientific strategies will give the greatest return on investment? Would it be useful, for instance, to develop target organisms besides *Arabidopsis*, delineating their genomes completely? How should an agricultural genome project be organized and coordinated? And are there other issues, such as ethical or social concerns, that should be considered? By talking about such issues ahead of time, it should be possible to avoid some of the mistakes of the past and to make the most of research in the future.

By the end of the workshop, several concepts and lessons from the past had emerged that were significant. No one disagreed that now is a good time to launch a broad, coordinated agricultural genome project. A number of participants believed it should, for instance, be "multi-tiered," with the genomes of some organisms deciphered totally and others done to various degrees of completeness. Many participants also seemed to believe that the project should be planned in great detail before much money is spent on it, since timing and coordination among the different components will be a key factor in its effectiveness.

The following is a summary and synthesis of the discussions that took place during the workshop and of the recommendations offered by the various participants.

Summary of Proceedings

LEARNING FROM EXPERIENCE

A decade ago, a genome project was a new and untested idea. Indeed, in 1986 when the Department of Energy announced it would fund a Human Genome Initiative, the idea was met with a great deal of skepticism about whether such a project could be done at a reasonable cost and whether the project, even if it could be done economically, would be of great value to science. Over the past ten years, however, such reservations have vanished, as researchers working on many different species have learned more about what goes into a genome project and more about what comes out. Today, biologists know the complete genetic sequences for a number of microorganisms, including yeast, several bacteria, and scores of viruses. And there are several programs under way that seek to acquire the much larger—and more difficult to obtain—genomes of plants and animals. Targets for these projects include a weed, a worm, the fruit fly, the mouse, and, of course, humans. Much of what has been learned from these efforts can be applied directly to an agricultural genome project.

The most straightforward borrowing will be of tools and techniques. Mapping and sequencing a genome is the same task whether the genetic material comes from yeast, people, or rose bushes—the genes of all are composed of the same chemical building blocks, just put together in different ways. This means, said Daniel Drell, a biologist with the Department of Energy's Human Genome Program, that the advanced genomics technologies developed for other species can easily be put to work on an agricultural project. "There is no point in reinventing the wheel," he said. "The Department of Energy and the National Insti-

tutes of Health have achieved quite a bit in terms of technology development, informatics, resource development. We are delighted to make available all the lessons we have learned.”

How to apply these scientific tools—which information is most important to pursue, which techniques are best suited for certain types of studies, how research should be ordered and prioritized, and so on—is a more complicated issue, but it is an issue to which lessons from other genome work still apply. In this case, however, it does matter which genome efforts serve as models. The most appropriate lessons will come from genome work on multi-cellular organisms rather than bacteria and other single-celled creatures, because of the difference in the size and complexity of their genetic codes.

Microorganisms have relatively small genomes. The yeast *Saccharomyces cerevisiae*, for example, has 16 chromosomes with a total length of 12 million base-pairs—a base-pair being an individual letter in the genetic instructions that are encoded in an organism’s DNA. Most of the bacteria whose genomes have been decoded have fewer than two million base-pairs. The genomes of plants and animals are, by comparison, huge. The common weed *Arabidopsis thaliana*, which is believed to have the smallest genome of any flowering plant, still boasts 120 million base-pairs. Corn’s genome has 2.3 billion base-pairs, and wheat’s is an oversized 16 billion. More than 100 times as large as *Arabidopsis*. Compared with wheat, the human genome is modest—only 3 billion base-pairs—but it is still a thousand times larger than that of the typical single-celled organism. The bigger a genome is, the more difficult it is to sequence, and not just because of the greater number of base-pairs. As the total amount of genetic material increases, the technical difficulties of handling and keeping track of all the DNA grow as well. For that reason, researchers looking to sequence the genomes of rice or corn or cows or pigs can learn most by gathering advice from their colleagues working on *Arabidopsis* and mice and humans.

That advice, as offered at the workshop, ranges from very broad and general to quite narrow and specific. On the broad end, there is general agreement about the right way to approach mapping and sequencing of multi-cellular organisms. The recommendations of Christopher Somerville, director of the Department of Plant Biology of the Carnegie Institution of Washington in Stanford, California, were typical: “When we began developing the *Arabidopsis* program, we realized the sequencing had to be preceded by other forms of information, particularly a very good map, because sequenced information is most useful when it is used in conjunction with mapping information on individual genes and mutants.” In mapping a genome, researchers piece together a large-scale picture of where genes and larger chunks of the chromosomes are. A sequence, on the other hand, is a letter-by-letter compendium of the genetic code, a listing of the base-pairs in the order in which they appear on a chromosome. (See Box 1.)

BOX 1 MAPS AND SEQUENCES

The genetic code of every living creature—plant, animal, or microorganism—is carried by long, twisting molecules of DNA. A single strand of DNA consists of a long sequence of small molecules called bases, and it is the order of these bases that encodes all genetic information. It is customary to represent the four bases that go into DNA as letters in a sort of chemical alphabet: A for adenine, C for cytosine, G for guanine, and T for thymine. In this way, a strand of DNA is written out as one long “word”: AATAGCTCC..., and so on for a hundred million or more letters in a single chromosome. In cells, DNA usually exists as a linked pair of complementary strands, with an A in one strand always paired side by side with a T in the other strand, and similarly for C and G. Thus the strand AATAGCTCC would normally be paired with its complement, TTATCGAGG.

Sequencing a stretch of DNA means listing the chemical bases in the exact order that they occur. Researchers usually speak of base-pairs instead of bases, but in practice they are equivalent since knowing one member of a base-pair automatically identifies the other. Once researchers have cataloged every base-pair in an organism’s DNA, they know that creature’s entire genetic code. From there they can track down the particular stretches of DNA devoted to genes, identify the sections involved in turning genes on and off, and in general decode the instructions for assembling and operating that particular creature.

The major obstacle to sequencing an organism’s genome—its total complement of DNA—is the sheer amount of DNA involved. Even bacteria tend to have genomes that are one or two million base-pairs long, and the genomes of plants and animals are generally hundreds of millions to billions of base-pairs long. Modern sequencing techniques can handle pieces of DNA no more than a thousand base-pairs in length, so researchers must chop up the long strands of DNA into small fragments before sequencing them, and this leaves them with the problem of how to piece together these fragmentary sequences. With various tricks it is relatively straightforward to assemble sequences that are thousands or tens of thousands of base-pairs long, but this is just a drop in the bucket on a chromosome that may be hundreds of millions of base-pairs long. Furthermore, it’s not easy to tell where on that chromosome a particular sequence fits.

This is where mapping comes in. As the name implies, a map gives researchers a way of finding locations along a chromosome. There are two basic sorts of maps, genetic and physical.

A genetic map approximates where on a chromosome particular genes or markers lie. Markers are short, easily identifiable stretches of DNA that vary from person to person, so that with simple tests researchers can usually tell which parent an individual inherited a particular marker from. When a mother’s and father’s chromosomes break up and later recombine to form the DNA for an offspring, genes and markers that are close neighbors on a chromosome tend to be inherited together. If Gene X is only 100,000 base-pairs away from Gene Y, and Bobby inherits Gene X from his mom, chances are very good that he’ll get Gene Y from her as well. By keeping careful track of how often traits and markers are inherited together, researchers can get a good idea of their relative locations on a chromosome. A genetic map of a chromosome may contain hundreds of genes and markers listed in the order they appear on the chromosome. The “genetic distance”

BOX 1 Continued

between any two of them—how likely it is that both will be inherited together—provides an approximation to the physical distance between them in terms of numbers of base-pairs.

A physical map, on the other hand, consists of a collection of overlapping stretches of the DNA itself. To produce a physical map, researchers start with many copies of a chromosome and chemically snip them into small pieces, each, say, just a few hundred thousand base-pairs long. Each piece is separated from the others and cloned, creating a set of many identical copies of this one bit of DNA, which can then be analyzed in various ways. Any given cloned stretch will likely share an overlap region with one or more other cloned stretches, and by identifying these overlaps, scientists can figure out that Clone X and Clone Y fit together with a certain common region. A complete physical map of a chromosome is a set of these clones that, together, cover the whole chromosome, along with a description of where on the chromosome the clones lie. Thus a physical map is more than a road map of the chromosome—it also includes physical bits of the chromosome that can themselves be manipulated and analyzed, and the libraries of clones that are part of a physical map are themselves very valuable to researchers.

Ultimately, one can completely sequence a chromosome by getting physical maps that have finer and finer resolution, until the individual clones are small enough to be sequenced. Then it is simply a matter of sequencing these millions of small bits of DNA and assembling them into a complete genome.

In theory, it might seem that the sequence is all one needs since all the information the genome holds is found in the order of the base pairs. But the practice is somewhat different. Maps are useful in a variety of ways. Before the entire sequence of a genome has been written out, they help researchers discover the genes responsible for particular traits or diseases. As the sequencing is going on, maps provide the physical material for the sequencing effort and allow researchers to piece together the sequences derived from small bits of DNA. Even after the sequence is complete, good maps are an indispensable guide to what is found where. One would no more throw away the maps and rely on the sequence than one would navigate a trip from Florida to California with neighborhood plots that showed every building, fence post, and fire hydrant along the way. “So,” Somerville said, “much of the support during the first few years [of the *Arabidopsis* project] was actually directed toward mapping. I think that is a good model to follow for the other genomes.”

For the *Arabidopsis* program, the map of choice was a “physical map,” a set of overlapping stretches of chromosome that covered the entire genome. In particular, the *Arabidopsis* researchers used yeast artificial chromosomes, or YACs—

chunks of *Arabidopsis* chromosomes modified so that they were accepted by yeast organisms as their own. When the yeast multiplied, they created many copies of each stretch of *Arabidopsis* DNA, establishing a permanent collection for study by as many scientists as were interested.

Future genome projects may want to use a slightly different type of physical map, suggested Neal Copeland, a mouse researcher at the National Cancer Institute. “As people [working on the mouse genome] have started using these YAC maps, it is clear that they are not the ultimate maps.” In particular, he said, because YACs are frequently unstable or chimeric, the resulting YACs “are no good for sequencing.” Instead of YACs, he recommended making maps with BACs—bacteria artificial chromosomes—which can be used for sequencing.

By borrowing technology and techniques from other genome work, researchers in an agricultural project should have everything necessary to map and sequence the genomes they’re interested in. “We do not have any technical problems that I am aware of,” Somerville said of the *Arabidopsis* effort. “One of our largest problems is just acquiring the financial resources to do it.” The only problem—if “problem” is the right word for it—is that the technology is improving so rapidly that most technical advice will be outdated in a few years. “In going back and looking at the projections for what would happen in the mouse and the human genome projects,” Copeland said, “what we have learned is that all these projections have been way underestimated. We have gone much beyond where anybody could have even dreamed that we would be right now.”

That technical proficiency is the root of the one major weakness that a number of workshop participants identified in current genome projects. As researchers in various laboratories collect more and more genetic data, some way must be found to transform this mass of often disconnected information into a unified whole—to collect it in one or a few places, organize it, catalog it, annotate it, and make it available to whoever wishes to use it. But too often this “bioinformatics” side of a genome project has failed to keep up with the generation of the data.

“Scientists who do genome research often neglect informatics,” said DOE’s Drell. “They just figure someone else is going to take care of it. As a consequence, they neither ask for nor get the kind of financial support that is necessary to curate data for deposition into a database. This is a major problem because what you get is a large collection of individual Web sites, and it is extremely difficult to find the data. The data officially are public. They are available if you know where to look for them, but they are very hard to get.”

“There is a whole variety of informatics issues,” Drell continued. “They have been a running headache for the genome projects, and they are too important to neglect. You should think about how to address them from the start. The most difficult and expensive solution of all is to not work them out first and then have to play catch-up afterwards.”

Many of the scientists at the workshop agreed with Drell that any agricul-

tural genome project should have plans in place from the beginning to take care of the data. The project should include large, central databases that are compatible with each other and easily accessible to researchers. Because the data will be valuable for years, if not decades, to come, it will be vital to plan for—and pay for—long-term curating of those databases. Furthermore, as David Galas, president of the Darwin Molecular Corporation and a member of the NRC's Board on Biology noted, the databases need to be more than just a collection of maps and sequences. "Sequence is fine, but what is really interesting is the biology that is in it." When biologists discover information about a particular gene in a database—what it does, how it is regulated, the structure of the protein that it encodes, and so on—they should be able to add that information to the database. "That is really a major challenge," he said, "and it is not cheap. There are a lot of different ways of going at it, but our ability to annotate, while getting better, is still a long way from being what we would like it to be."

Researchers must keep in mind that they will be collecting data not just on one species but on dozens. "We've got real problems that go well beyond what we have dealt with in the Human Genome Project, and that is in comparative genetics—genomics among a very large number of species," said David Galas. If an agricultural genome project is to be successful, the databases for the various species should be interconnected and compatible with one another, so that a researcher studying, say, tomatoes can easily find relevant information from *Arabidopsis* and other species. "That is a whole other level of difficulty," Galas said, as researchers on mouse and human genomes have discovered. Integrating the databases from those two species would be valuable for the same reasons that integration among agricultural genomes would be valuable, but so far such coordination is more hope than reality. "Quite frankly," he said, "we're still catching up."

Finally, the early consideration of social and ethical issues for an agricultural genome project is a step that other genome projects have proven to be valuable (see Box 2: Considering Social and Ethical Issues).

DEALING WITH A MULTITUDE OF GENOMES

An agricultural genome project will have much in common with the Human Genome Project. Much of the technology will be the same, the agricultural effort will demand the same massive databases and careful coordination between laboratories that the human program does, and both projects have the promise of potentially revolutionary payoffs. But, as Daniel Drell of the Department of Energy said, there will be a major difference, one that makes the agricultural project much more challenging. "There is only one species of human, so the human genome project is easier to define, and it only has to be done once," he said. Agriculture, by contrast, involves hundreds of species, dozens of them

BOX 2 CONSIDERING SOCIAL AND ETHICAL ISSUES

In the early 1980s, four drug companies announced that they would soon begin selling farmers a product to increase milk production in cows: recombinant bovine growth hormone, or rBGH. Except for being produced in bacteria, rBGH was the same molecule that cows themselves produced to trigger milk production, so the companies anticipated no problems with acceptance, and indeed, in 1984, the Food and Drug Administration pronounced milk from rBGH-treated cows safe for human consumption. Soon after, however, the artificial hormone became mired in controversy. Proponents of family farms claimed rBGH would shrink the nation's dairy herds and spell the end of small producers. Environmentalists and some scientists began raising safety issues they claimed the FDA had glossed over, including charges that drinking rBGH milk might increase the risk of breast cancer in women and that the increased use of antibiotics on rBGH-treated cows might lead to the development of antibiotic-resistant bacteria. In response, the FDA asked the National Institutes of Health to study the issue, and in 1990 an NIH panel gave rBGH a clean bill of health.

Still, the controversy continued. Ben & Jerry's ice cream stores announced they would not buy from any dairy using the still-experimental treatment. Several grocery store chains followed suit. Maine and Vermont passed labeling laws intended to identify milk products from cows given rBGH. It was not until 1994 that the FDA gave the final go-ahead for commercial sales of the recombinant hormone. Since then, the concerns have subsided, and rBGH is now used regularly in many herds.

The rBGH controversy offers a cautionary tale about the pitfalls of genetic engineering in agriculture. Although neither the cows nor the milk were altered, and although the only thing "unnatural" about the rBGH was that it had been produced in bacteria rather than cows, a significant part of the public was nevertheless uneasy about the use of the hormone.

In light of this history and the nervous reception given to the few genetically engineered food products to reach the market, such as the slow-ripening Flavr-Savr tomatoes, biological researchers need to think about the social and ethical implications of an agricultural genome project, said Daniel Drell, a biologist with the Department of Energy's Human Genome Program. "I think it would be naive for the agricultural community to imagine you are not going to get challenged." This is particularly true because the whole point of an agricultural genome project is to make possible wholesale manipulation of the genes of crops and livestock. Yet, as Nina Fedoroff of Pennsylvania State University pointed out, "People continue to have problems with the concept of designer foods. People continue to have problems with genetically engineered animals and plants."

Instead of assuming, as the makers of rBGH did, that the public will automatically welcome their efforts to improve agriculture, workshop participants urged those planning an agricultural genome project to consider the social and ethical implications of the work ahead of time and try to anticipate objections and difficulties. At a minimum, a task force on social and ethical issues should be an official part of the program. Scientists need to acknowledge that some of the concerns of the public are legitimate and that more technical information is not necessarily the answer. "The ultimate regulation," Drell said, "is going to be whether John and Jane Q. Citizen buy your product. If they are made nervous by what they imagine you have done to it, you will find out about it in the worst way."

economically important. The key issue for any agricultural genome program will be how to deal with this multitude of genomes.

In theory, of course, it would be nice to have the complete genome of every animal and every plant that is important to the food and fiber industries: cows, pigs, sheep, chickens, turkeys, corn, wheat, rice, soybeans, potatoes, tomatoes, strawberries, cotton, apple trees, orange trees, pine trees, oak trees, and many more. In practice, that isn't going to happen, at least not any time soon. Genome projects are expensive. The Human Genome Project is expected to cost \$3 billion by the time it is finished, and even the very short genome of *Arabidopsis* will take as much as \$75 million to decipher. As a practical matter, it will be possible to go after only a small fraction of all the agricultural genome information one would ideally like to have.

Fortunately, it's possible to make a small fraction go a long way, as long as the fraction is chosen correctly. Most of the agriculturally important crops are genetically related—distant cousins perhaps, but still cousins—and so information obtained about one can often be put to use with another. The same is true for farm animals. So the major question facing an agricultural genome project becomes: Which fraction of all the genetic information available should be tackled first to get the most bang for the buck?

A number of the participants in the workshop agreed that there is no better way to get bang for the buck than to sequence the entire genome of a representative organism, as is now being done for humans and for the weed *Arabidopsis thaliana* (see Box 3: Saved by the Weed). As David Galas, president of the Darwin Molecular Corporation, said, "The actual value of having a single species worked out in some detail is absolutely enormous and will be catalytic, independent of what is done beyond that." Having an entire genome to study—not just the sections coding for the genes, but the regions that turn the genes on and off and also the vast stretches of DNA that seem to serve no function—will allow researchers to learn details about the design and functioning of living organisms that can now only be guessed at.

But perhaps the greatest value of a complete genome is that it offers a catalog of all the genes at work in an organism. Humans have, for instance, an estimated 50,000 to 100,000 genes, but only about 10,000 of them have been identified and sequenced. *Arabidopsis* has approximately 20,000 genes, of which only about 2,000 have been fully sequenced. When their respective genome projects are complete, researchers will have a list of every gene as well as its sequence and its exact position on a chromosome. Furthermore, these catalogs will provide researchers with complete or nearly complete lists of genes for other animals and plants, even those that are only distantly related to humans or *Arabidopsis*.

"Our impression is that basically all plant genes are represented in *Arabidopsis*," said Chris Somerville of the Carnegie Institute of Washington. The exact structure of the genes varies from plant to plant—which accounts for the differences between species—but all plants seem to make do with essentially the

BOX 3 SAVED BY THE WEED

Ironically enough, the most important plant in agricultural research over the next decade may be a common weed. *Arabidopsis thaliana* is a small, nondescript member of the mustard family. Even when its white flowers are in bloom, it is not a plant that attracts much attention—except from biologists. Over the past fifteen years, *Arabidopsis* has become the plant world's equivalent to the laboratory mouse, and now it is poised to become the first plant to have its entire genome sequenced and made available for study.

In the early 1980s, a group of plant biologists decided to choose one plant that could, like the mouse, serve as a model for exploring the workings of many species. They settled on *Arabidopsis* for many of the same reasons that the mouse became a model animal: it is small and easy to maintain, it reproduces rapidly (its growing cycle is six weeks), and it has a variety of mutants. Furthermore, *Arabidopsis* has an exceptionally small genome—only 120 million base-pairs, believed to be the smallest of any flowering plant. This made it easier for researchers to track genetic mutations to a particular spot on one of the plant's five chromosomes, and it also made *Arabidopsis* a natural target for a genome project.

In the fall of 1996, three U.S. government agencies announced funding for an *Arabidopsis* sequencing effort, the first—and so far only—plant genome project in the United States. The National Science Foundation and the Departments of Energy and Agriculture awarded \$12.7 million over three years to three sequencing groups headquartered at Cold Spring Harbor, Stanford University, and The Institute for Genome Research. At about the same time, those three U.S. teams announced that they would cooperate with two other *Arabidopsis* sequencing groups, one at a laboratory in Japan and the other a consortium of 17 European labs. By agreeing to divide up the sequencing work and share the results, the five groups expect to delineate the entire *Arabidopsis* genome by 2004.

Like work on the mouse, the ultimate value of research on *Arabidopsis* lies in what it teaches scientists about other, more important species. All flowering plants are very similar genetically, so insights garnered from *Arabidopsis* can be applied to everything from roses and sunflowers to corn and tomatoes. Indeed, Christopher Somerville, an *Arabidopsis* researcher at the Carnegie Institution of Washington, said that essentially every gene found in any flowering plant has a counterpart in *Arabidopsis*. Thus if a corn researcher wished to study, say, the uptake of phosphorus from the soil by the roots of a corn plant, he would not have to search blindly through the corn genome to find the relevant genes. He could instead identify those genes first in *Arabidopsis*—a much easier process, given the size of its genome and its eventual complete sequencing—and from there track them to corn.

Although there is not agreement on such possibilities, the *Arabidopsis* Genome Initiative promises to be important for corn and many other crops, said Tony Cavalieri of Pioneer Hi-Bred International, the major supplier of corn seed in the United States. "You might say that the most important commitment [by the government] to corn genetics to date has been not the genomics work in corn that has been publicly funded, but the commitment to the *Arabidopsis* sequencing program."

same set of genes. “For example,” Somerville said, “we find the same genes controlling floral morphology [the shape and structure of flowers] in *Arabidopsis* as in snapdragon, which has a very different flower, and in maize, which has an even more distant, different-looking flower.” The story is the same for other genes.

“I think that *Arabidopsis* research is going to culminate in understanding what every plant gene does,” Somerville said. “The significance of that is going to be an ability to take that knowledge and map it on to all of the plants that we care about.” In other words, to study a particular trait in any plant, it should be enough to figure out which gene controls that trait in *Arabidopsis* and then look for the corresponding gene in the plant of interest. By pinpointing the genes that make *Arabidopsis* resistant to various diseases, for instance, researchers could track down those genes in other plants, especially strains with a strong resistance to particular diseases. Those genes could then be inserted into the DNA of commercial crops to give them the same resistance.

Leveraging the *Arabidopsis* genome in this way depends on an exciting recent discovery that during the evolution of the angiosperms, plants whose seeds are within a fruit, the order of genes on chromosomes appears to change slowly. Thus, closely related species such as green pepper and tomato have large regions of chromosomes with the same gene order. Similarly, rice, maize, wheat, barley, millet and sorghum have been found to have very similar gene order. This means that if the genome of one of the angiosperms were completely characterized it would be possible to extrapolate much of that information to all of the other plants in that group with a high degree of accuracy. Thus, for instance, if a gene for a trait such as drought tolerance was mapped in species-A, it would be possible to identify the DNA sequence of the gene in the fully characterized species (species B) based on the conserved gene order. This, in turn, would permit the isolation of the useful variant of the gene from species-A by using the DNA from species-B as a hybridization probe. Unfortunately, although this tactic will work for plants closely related to *Arabidopsis*, such as oilseed rape or brussels sprouts, it gets less useful for those not so closely related—wheat, say, or pine trees. As Jeff Bennetzen of Purdue University said, “I would agree with Chris [Somerville] that *Arabidopsis* will give us virtually all plant genes. The difficulty is the ties we will be able to make across these organisms.” Although there is some colinearity between even distantly related species, there is not nearly enough. A gene might be very close to a particular marker in *Arabidopsis* but nowhere near the corresponding marker in, say, soybeans. As a result, Bennetzen said, “You are not going to be able to transfer information across species quite that simply based on map position.”

For that reason, many workshop participants agreed that one of the goals of any agricultural genome program should be to develop target species other than *Arabidopsis*. The genomes of these organisms should be fully sequenced so that they too can serve as models for other, related species.

This raised the question, “Which species should be targeted?” Several criteria for choosing were suggested. First, since it would do little good to sequence another plant closely related to *Arabidopsis*, the target species should be selected from other parts of the plant world. All crop plants—indeed, all flowering plants—belong to one of two broad groups, monocots and dicots, which are distinguished by the number of leaves that first sprout from the seed—one in the case of monocots, two for dicots. *Arabidopsis* is a dicot, as are peas and beans, potatoes, tomatoes, carrots, strawberries, apples, artichokes, and beets. Monocots include wheat, rice, corn, bananas, pineapples, onions, and asparagus. “*Arabidopsis* is a really great model for all of the dicots,” Somerville said, “but we really should have extensive information on one monocot.” Other workshop participants agreed.

There was no consensus, however, on which monocot should be given priority. If the shortness of the genome is a consideration—as it was in the choice of *Arabidopsis*—then rice is the obvious candidate. Its genome is only 420 million base-pairs, several times larger than that of *Arabidopsis*, yet smaller than any other commercially important monocot. Several other countries, including Japan and Korea, have already begun sequencing the rice genome, and a U.S. effort could lead to the sort of international genome effort already at work on *Arabidopsis*. On the other hand, a number of workshop participants thought a corn genome project would make more sense. Although its genome is more than five times that of rice—2.3 billion base-pairs—corn is a much more valuable crop in the United States, thanks mainly to its use in animal feed. (Wheat is also a valuable crop, but the size of its genome—nearly seven times that of corn—rules it out as a target species. That oversized genome is due partly to wheat’s evolution as a merger of three separate species and partly to the genome having a great deal of repetitive DNA. Rice, by contrast, has a very efficient genome, with relatively little space on its chromosomes devoted to DNA that serves no function.)

Nor was there any consensus on how many target organisms should be chosen. Some thought one monocot—probably corn or rice—would do. Brian Larkins of the University of Arizona argued for both. “There are some good technical reasons why maize should be done and not simply assume that it is going to be easy to extrapolate the data from rice.” Comparisons of the DNA of rice and corn show, he said, that there is much less similarity in the structure of their genomes than once thought. And Bennetzen commented it may be necessary to go even further. “What a lot of us in the field think now is that we are going to need a number of nodal organisms, organisms that will allow you to study a whole series of species that are closely related.” Agricultural genome research would be apportioned into groups of related species, each with its own nodal organism, he said.

Because completely sequencing a genome is expensive, there is a limit to how many target species or nodal organisms can be delineated down to the last

base-pair. Just sequencing corn completely, for example, would run into the hundreds of millions of dollars. Entirely sequencing multiple target species would be prohibitively expensive.

Fortunately, most of the information provided by a complete sequence of a genome can be gained by creating detailed maps—both physical and genetic—of the genome, identifying and sequencing most of the genes (or, more accurately, cDNA clones for the genes), and then positioning the genes on the maps. This avoids sequencing the entire genome, which includes many sections that don't contain genes. "You can get virtually all of the information you need without sequencing through all the repetitive sequences and clusters of retrotransposons," said Nina Fedoroff of Pennsylvania State University. "The bottom line is that there are large chunks of clustered sequence that are probably not going to be useful." Since sequencing is a major part of the cost of a genome project—it's now about 30 cents to \$1 a base-pair, although the price is dropping rapidly—bypassing this low-value DNA cuts the cost of analyzing a genome sharply.

Furthermore, it's not necessary to completely sequence the genes themselves, either. It's possible, Somerville pointed out, to get a lot of the information about a gene by using expressed sequence tags, or ESTs, a version of the gene for which one sequences only a few hundred of the thousands of base-pairs that make up the full gene. (See Box 4: Gene Chips.) "Based on similar studies of human genes you only get five percent more information about the probable function of a gene, based on sequence analysis by fully sequencing the genes," Somerville said. That is, if a researcher has isolated a gene from, say, potatoes and wishes to discover its function by comparing it with a collection of known genes from *Arabidopsis* or another target species, the chances of success are only about five percent greater if the researcher works with a completely sequenced gene than with ESTs.

The use of ESTs makes it feasible to amass a surprisingly large amount of information about even those plants and animals that are not chosen as target species. "In almost every area of genomics, there is always this battle over what species are going to be done," said Craig Venter of The Institute for Genomic Research. "We find this quite a lot in the microbial world because everybody has their own pet species. . . . But instead of people fighting over whether it is going to be wheat or corn or cows, fundamental data can be put out in a very short period of time that will rapidly advance all of these areas simultaneously."

Specifically, Venter recommended, "you could generate for \$20 million spread over three to five years, 100,000 ESTs from each of forty different species. So there would be over four million ESTs from forty species that could be rapidly put into the public domain." Although there is invariably both redundancy and omission in a set of ESTs—some genes are represented by several or many tags, while other genes are missed altogether—a collection of 100,000 ESTs would pick out most of the important genes in a species. These ESTs could then, Venter said, be placed on physical maps of each of the forty genomes by

BOX 4 GENE CHIPS

An agricultural genome project would advance agricultural research in many ways. One of the most intriguing is that it would make possible the use of the so-called “gene chips” for dozens of different agricultural species, opening up an entirely new way of analyzing the inner workings of these plants and animals.

A gene chip offers a way of testing which of an organism’s genes are active. Of the tens of thousands of genes that make up a plant’s or animal’s genome, only a portion are turned on—that is, working to produce their particular proteins—at any given time. Perhaps a few dozen genes are very highly active, others are less busy, still others are barely working, and a majority are idle. This pattern of gene activity varies according to cell type—it will be quite different in a root cell than in a leaf cell, for instance—and it changes according to circumstances. An infection will alter the pattern of gene activity, for example, as will such environmental stresses as a lack of water or the presence of pollutants. By analyzing the patterns and how they change, researchers can get clues into how an organism functions—for example, finding genes for disease resistance by seeing which genes become active in the presence of an infection.

“We already have gene chips with thousands of genes,” said Christopher Somerville of the Carnegie Institution of Washington. “But we envision, in the very near future, small chips in which all of the genes will be placed on these chips and these can be used to measure the expression of all of the genes in the plant in a single experiment. These kinds of experiments are going to qualitatively change how we go about doing plant biology.”

To date, however, the gene chips can be made for only a small number of species—those plants and animals for which gene libraries already exist. When a gene is turned on in a cell, the cellular machinery uses the long strand of DNA that makes up the gene as a template, creating a complementary strand of messenger RNA, or mRNA. This mRNA in turn is used to direct the assembly of a protein.

using radiation hybrids—a technique in which chromosomes are broken apart by radiation and the fragments inserted into the cells of other organisms for handling. Furthermore, by using “gene chips,” microscopic arrays that contain probes for thousands of different genes, it will be possible to match up the ESTs from the various species against each other and also relate them to genes in *Arabidopsis* or other target species. “I would argue,” Venter said, “that this comparative data is going to be far more valuable to have than to completely sequence any one species.”

Given the value of obtaining complete genomes for target species and the possibility of using ESTs to skim the cream off several dozen genomes for a relatively small investment, the researchers at the workshop thought the best approach to an agricultural genome project would be a multi-tiered one. At the top tier would be a few target species, such as *Arabidopsis* and rice or corn, whose genomes are sequenced completely. At the lowest tier would be dozens of

Thus the amount and type of mRNA in a cell is a direct measure of the pattern of gene activity in that cell, and researchers have taken advantage of this to create gene libraries based on the expression of genes in cells. Scientists isolate all the mRNA from a cell, make DNA strands that are complementary to each mRNA strand, and then make many copies of each of these complementary DNA, or cDNA strands. The resulting cDNA samples, each consisting of many copies of a single cDNA strand corresponding to one particular gene, make up a gene library, and scientists have assembled libraries with tens of thousands of these samples for certain plants and animals. Since each of these cDNA strands consists of tens of thousands of base-pairs—the chemical units that make up DNA—researchers generally determine only a partial sequence of a few hundred base-pairs, enough to identify the strand. These partially sequenced strands of cDNA are called expressed sequence tags, or ESTs.

The gene chips consist of thousands of these samples of cDNA from one organism—up to 10,000 samples in a square centimeter—on a glass slide. To determine gene activity, a researcher isolates the mRNA from the cells of interest, uses the mRNA to make complementary strands of cDNA that are labeled with fluorescent molecules, and then lets that cDNA react chemically with the cDNA samples on the gene chip. Because the labeled cDNA sticks to complementary bits of cDNA on the gene chip, a researcher can then scan over the gene chip and tell by the spots of fluorescence which cDNA from the library corresponds to genes that are active in the cell. Another use envisioned for gene chips is to detect polymorphisms in genomic DNA. By developing gene chips that can assay the allelic composition of hundreds of genes at a time it will be much more feasible for plant breeders to use marker-assisted breeding.

“This is the technology that will take over plant breeding at some point in the future,” Somerville said, but not until extensive amounts of DNA sequence information is available for the major agricultural species. And that, he said, should be one result of an agricultural genome project.

plants and animals whose genomes are mapped out in some detail and most of whose genes are identified by expressed sequence tags. In the middle would be nodal organisms with genomes delineated to an intermediate amount.

The project should include not just crop plants, but also livestock, crop trees, and even microorganisms, each group with its own target species. The microorganisms are often overlooked in discussions about an agricultural genome project, noted Jim Cook of USDA’s Agricultural Research Service at Washington State University, but they are of crucial importance to agriculture. Many are pathogens, of course, which farmers would like to rid their crops of, but some are beneficial, such as bacteria that live symbiotically with the roots of crop plants and help the plant absorb nutrients from the soil.

In addition to this multi-tiered approach, workshop participants had several other bits of scientific advice. One key to the success of an agricultural genome project will be keeping abreast of the most up-to-date mapping and sequencing

techniques. “There are new technologies emerging, the EST sequencing and the radiation hybrid maps, that make this a lot cheaper than it used to be,” said Neal Copeland at the National Cancer Institute in Frederick, Maryland. “Looking back at the mouse, if we knew about this technology we could have done it a lot cheaper and a lot faster.”

Another key will be coordination. Since only a few genomes will be sequenced completely, comparison of genetic information across species will play a major role in an agricultural genome project, and the maps will have to be made with that in mind. Much of the mapping done today does not match up from one organism to another, Bennetzen said. “They are maps that are specific to a given species. That is a big problem.” If the multi-tiered approach is to be a success, “mapping is going to need to be done in each species in a transferable mode.”

Finally, it is important to keep in mind that all of this—the ESTs and radiation hybrids, the sequences and maps—represents just the first phase of what will be a very long, though ultimately very profitable, process. “When you go back to think about it, this is the easy and cheap phase,” Copeland said. “We’re talking about a lot of dollars, but this is cheap compared to what it’s going to take to figure out what all these genes are doing.” After the mapping and sequencing are done, researchers will face the task of understanding the structure and function of the various genes, how they are controlled, and how they interact with one another—and also uncovering how the variations in genes from species to species alter these details. It will be these discoveries that make possible the coming revolution in agriculture. “What we are really doing here [with an agricultural genome project], Copeland said, “is generating the basic infrastructure for the next millennium.”

ORGANIZING THE PROJECT

Any genome project will involve two very different types of science. Generating ESTs, creating maps, sequencing DNA—these are mostly repetitive procedures which involve doing the same thing over and over again hundreds, thousands, sometimes millions of times. To a large degree, they can be automated. On the other hand, working with an individual gene to understand its function and how that function is linked to the gene’s structure is a different challenge with each new gene, demanding the attention of and intellectual contributions from individual researchers. These two types of science will be most effective under two very different types of organizational structure.

For the mapping and sequencing part of the project, the *Arabidopsis* genome project may be a good model. It is a very narrowly focused, carefully coordinated program funded by three agencies working in tandem: the National Science Foundation, the Department of Energy, and the Department of Agriculture. The project is being carried out at more than half a dozen laboratories organized into three

teams, and these teams are in turn coordinating their efforts with *Arabidopsis* genome programs in Europe and Japan.

This complex division of labor demands careful planning and constant communication, said Christopher Somerville, a co-investigator on one of the teams. “One of the successes of the *Arabidopsis* group,” he said, “has been the development of coordinating committees at both the national scale and an international scale to make sure that we do not engage in any wasteful duplication of effort and to resolve issues that range all of the way from nomenclature to allocating work to the different groups, organizing meetings and [running] databases.”

Neal Copeland of the National Cancer Institute, who has worked on the mouse genome, echoed the importance of such communication. “The genetic map [for the mouse] was done in five or six large mapping centers throughout the world. Early on, we recognized the need to interface all this data being generated by these different groups. So what we did was modeled after the Human Genome Project. We set up chromosome committees. In the mouse we now have a chromosome committee for each chromosome. We have an international meeting once a year, usually one year in Europe, one year in the United States, where everybody gets together, all the chromosome committees and anybody else who is interested, and they go through and they revise the maps, and then they get published. They used to get published in print form, but now basically they are being published in electronic form.”

Of course, an agricultural genome project will be much broader and bigger than the *Arabidopsis* or the mouse program, so communication and coordination will be that much more important—and that much more difficult to maintain. Not only will researchers working on the same species need to keep in touch and integrate their efforts, but scientists studying different species, different orders, perhaps even different phyla and different kingdoms, will also need to develop lines of communication and to harmonize their mapping techniques, database formats, and so on. A further complication arises with the need to coordinate research done in the private sector with that performed in the public sector (see Box 5: Public and Private Genomes).

Economies of scale also become more important with a large genome project, Copeland noted. In the mouse genome, for example, most of the microsatellites—a type of gene marker consisting of short sequences of base-pairs repeated many times—were mapped at one location, the MIT Genome Center. “I think we wouldn’t be nearly as far along in the development of the microsatellite map had it not been done in one central facility,” Copeland said. Similarly, ESTs for the mouse are generated at just one place, Washington University. “This is not a mouse group,” he said, “but the reason they are doing it is, again, economy of scale. These people are experts in the field, and it wouldn’t make any sense for a mouse lab that didn’t have this technology to do this. So in thinking about doing EST maps for the agricultural species, it makes sense to be doing them in facilities that have the technology in hand and that can do them more cheaply.”

BOX 5 PUBLIC AND PRIVATE GENOMES

Because of the importance of genome information, there is often a tension between the public and private sectors over access to it. Private companies, seeing competitive advantage in having genome data that their competitors do not, may seek to keep some of that data to themselves. On the other hand, scientists in universities and government labs often wish to get their hands on as much genome information as possible in order to maximize the quality of their research. Participants in the workshop described two cases in which this tension has been apparent, even in the early days of agricultural genome programs.

In Europe, Christopher Somerville said, commercial considerations have delayed the dissemination of *Arabidopsis* genome data that was paid for partly with public funds. The seventeen labs in the consortium, he said, "have completed about two megabases [two million base-pairs] of DNA but have imposed a long delay on the release of that sequence because the Europeans decided to organize their genomic sequencing in collaboration with industry." As part of the consortium agreement, industry partners were given the right to the first look at the *Arabidopsis* sequence as it was finished, Somerville explained, and "for whatever reason, that sequence has not yet been released."

Meanwhile, in the United States, the vast majority of expressed sequence tags, or ESTs, for the corn genome are in private hands. According to Tony Cavalieri of Pioneer Hi-Bred International, his company has today about 80,000 ESTs from corn, representing an estimated 45,000 different genes. And although the company has shared some of that information with some public-sector scientists, he said, it has no plans to release the data and let its competitors take advantage of research paid for by Pioneer. "That information is important," he explained. "It allows you competitive advantage" by helping researchers find the entire gene and patent it.

In general, the workshop participants were not happy with such situations. "I use a lot of Pioneer seed, and I am thankful for the work that Pioneer is doing in

Such considerations as economies of scale and the desirability of concentrating expertise and the most advanced technology in a few places may argue for a very different type of organization than biologists are used to, said Bob Cook-Deegan from the Institute of Medicine. "I don't think that the [Human] Genome Project is the right model for the technologically intensive part [of an agricultural genome program]. The right policy model might actually be how the Department of Defense, particularly DARPA, has handled chip design, interactive computing, and information processing techniques. They channel a lot of resources into a few centers, rather than NSF's or NIH's more distributed system."

But whether the Human Genome Project or the highly concentrated programs that DARPA runs serve as the model, many participants believed that one tier of an agricultural genome project should be a directed, technologically intensive effort aimed at mapping and sequencing as much of as many different

genome work," said Wallie Hardie, National Corn Growers Association, "but I am really not very comfortable with just one company doing this work in genome mapping. I would like to see the power of this technology in the public sector so a lot of the companies could get these differentiated products to me quicker." Neal Copeland of the National Cancer Institute argued that basic genome information, such as ESTs, maps, and sequences, should be thought of as scientific tools and, as such, should be made widely available in order to keep science moving forward as quickly as possible. "These building blocks," he said, "are going to drive all of the science that goes on in the next decade or the next millennium, and this all needs to be in the public domain." Even Pioneer's Cavalieri agreed: "Our preference would be that the EST information was available, that the tools for figuring out function were available, and that we would compete on the basis of product development using the genes that were known and available to the world at large." Pioneer decided to go after the ESTs itself, he said, only when it seemed that nothing was being done in the public sector and the company became worried that its competitors would move ahead of them in this area of basic science.

Thus most researchers at the meeting called for a government project that would make agricultural genome information publicly available. When work is done in collaboration with industry, the basic information—ESTs, maps, and so on—should be put in the public domain, although companies may be given intellectual property rights to other discoveries, such as gene function.¹ It may even be useful in some cases for the government to buy data that the public sector currently owns. "One way of saving a lot of time and work and money," suggested Brian Larkins of the University of Arizona, "would be simply to purchase Pioneer's EST database and make it available." That might be possible, said Cavalieri of Pioneer. "I do think there would be some willingness to talk about what the private and public initiatives could be to solve some of these questions."

¹Refer to the NRC proceedings Intellectual Property Rights and Research Tools in Molecular Biology (1997) and proceedings of NRC forum on Intellectual Property Rights and Plant Biotechnology (1997).

agricultural genomes as possible, as quickly as is feasible. Much of it should be carried out in places specializing in a particular technique or techniques, rather than having expertise and interest in a particular species. Large portions of the work will not be intellectually challenging, but they will be technically demanding.

The second tier will be quite different. The maps and sequences are of little use without details about what the various genes do and how they do it, and this information cannot be produced in the same assembly-line fashion that the maps and sequences can. Instead, the best approach is to fund individual investigators who are pursuing their own research interests.

To date, much of the individual-investigator funding for agricultural genome research has come from the USDA's National Research Initiative (NRI). Last year, said Ron Phillips, NRI's chief scientist, the National Research Initiative

awarded 72 grants totaling \$9 million in the plant genome area. “Since [the plant genome program] started in 1991,” he said, “there has been money put toward about 50 different species.” NRI also funds research into animal genomes. But whether the support comes from NRI or as part of a broader agricultural genome project, workshop participants agreed that investigator-initiated research will be the most effective way to pursue such specialized investigations as how a gene’s structure is related to its function.

Finally, many researchers at the workshop stated that things are changing so quickly in the area of genomic research that any genome plan should be flexible enough to take advantage of advances and new understandings as they come along. “I wonder if it wouldn’t make sense to suggest that you emulate the five-year planning process that the Human Genome Project undertook in 1989, 1990,” said Daniel Drell, a biologist with the Department of Energy. “It basically involves getting affected scientists, those who can contribute, to work up a draft five-year plan and then circulate it and see what reaction it elicits.” And then, of course, Copeland said, keep in mind that the five-year plans will certainly be obsolete long before the end of five years. “In going back and looking at the 1991, 1995 projections for what would happen in the mouse and the human genome projects, what we have learned is that all these projections have been way underestimated. We have gone much beyond where anybody could have even dreamed that we would be right now.”

APPENDIX

A

Program and Discussion Questions

Remarkable progress has been made in the last few years to produce genomic maps of organisms important to agriculture such as *Arabidopsis thaliana* (mustard), *Cochliomyia homin* (screwworm), *Caenorhabditis elegans* (nematode), *Oryza sativa* (rice), *Zea mays* (maize), and *Sus domesticus* (swine). Identification of homologous chromosomes across species has raised questions as to the need to coordinate mapping activities. Increased cooperation among animal, microbial, and plant geneticists could reduce redundancy and increase genetic knowledge shared among researchers.

The National Research Council's Board on Agriculture in collaboration with the Board on Biology will hold another in a series of forums under the Forum on Biotechnology, to identify issues and opportunities in designing a USDA initiated agricultural genome project. Scientific experts from a variety of genome mapping programs will assemble for an open exchange of views in a neutral setting. Researchers from government, academe, and industry will share their perspectives on current programs and opportunities and issues for a more integrated genome initiative .

FORUM QUESTIONS

1. What is the status of some of the current genome mapping programs ?
2. What have been the most important issues for these mapping programs?
3. What are the opportunities and issues in designing an agricultural genome mapping program?

FORMAT

NRC will have a roundtable discussion to promote an open exchange of views among NAS members, federal agency administrators, industrial scientists, and university researchers. Three panels from USDA, other federal agencies, and private industry will share their experiences on *Arabidopsis*, maize, swine, rice, and human genome mapping projects. A summary report of the forum will be prepared for publication by the National Academy Press.

APPENDIX
B

Forum Agenda

- 7:30 am Light Breakfast (Lecture Room)
- 8:00 Welcome
Dale Bauman, Cornell University,
Chair, Board on Agriculture
Michael Clegg,
University of California, Riverside
Chair, Board on Biology
- 8:10 Toward a USDA Food Genome Project
Ronald Phillips, Chief Scientist,
National Research Initiative
Competitive Grants Program
- 8:30 Panel One: USDA Programs
Arabidopsis: Chris Somerville,
Carnegie Institute of Washington
Maize: Brian Larkins, University of Arizona
Livestock: Colin Scanes,
Iowa State University
Livestock: Dan Laster,
U.S. Meat Animal Research Center
- 9:30 Q&A

- 10:00 Break
- 10:15 Panel Two: Commercial Agricultural
Enterprises
Wallie Hardie,
 National Corn Growers Assoc.
Tony Cavalieri, Pioneer Hi-Bred
Chuck Schroder (unconfirmed),
 National Cattlemen’s Beef Assoc.
Sam Buttram, Dekalb Swine Breeders, Inc.
- 11:00 Q&A
- 11:30 Working Lunch
- 12:30 Panel Three: Non-Agricultural Genome Programs
Mary Clutter,
 Biological Sciences Directorate,
 National Science Foundation
Ari Patrinos, Office of Health and
 Environmental Research,
 U.S. Department of Energy
Neal Copeland, Frederick Cancer Research
 and Development Center,
 National Cancer Institute
- 1:30 Q&A
- 1:45 Wrap-Up
- 2:00 Adjourn

APPENDIX

C

Participant Biographies

FORUM CHAIRS

Michael T. Clegg is Dean of the College of Natural and Agricultural Sciences at the University of California, Riverside. He is the leading student of the evolution of complex genetic systems. He is recognized internationally for his contributions to understanding the genetic and ecological basis for adaptive evolutionary changes within populations and at higher taxonomic levels. Clegg's current research interests include: population genetics of plants; plant molecular evolution; plant phylogeny; and genetic conservation in agriculture. He received his Ph.D. degree in genetics from the University of California at Davis. Clegg is member of the National Academy of Sciences and chairman of the Board on Biology.

Dale Bauman's primary research areas include regulation of nutrient use in bovine lactation, growth, and pregnancy and bovine mammary gland biology. Bauman and his colleagues crystallized the concept of homeorhesis, the process of long-term regulation of nutrient use during a particular physiological state such as lactation. His concepts of metabolic regulation are widely accepted and applied to many aspects of developmental biology. He received his undergraduate and masters degrees from Michigan State University and his Ph.D. degree in nutrition-biochemistry from the University of Illinois. Prior to his appointment at Cornell University, he was an associate professor at the department of dairy science at the University of Illinois. Dr. Bauman is a member of the National Academy of Sciences and chairman of the Board on Agriculture.

PARTICIPANTS

John M. Antle is a professor in the department of agricultural economics and economics at Montana State University. His research areas and teaching fields include environmental and natural resource issues in agriculture, econometric analysis of agricultural production, international economics, and economic development. He served as senior economist on the President's Council of Economic Advisers, where he was responsible for agricultural, trade, and environmental policy. He received his undergraduate degree at Albion College and his masters and Ph.D. degrees in economics from the University of Chicago. Antle currently serves on the National Academy of Sciences' Board on Agriculture.

Sandra S. Batie is Elton R. Smith Professor in Food and Agricultural Policy at Michigan State University. Her research includes work in agro-environmental policy, soil conservation policy, water quality policy, research methodology, and rural development policy. Prior to her current position, she was professor of agricultural economics at Virginia Polytechnic Institute and State University and has had two sabbaticals: one at the Conservation Foundation, and the other at the National Governors' Association. Batie received her undergraduate degree in economics from the University of Washington and her masters and Ph.D. degrees in agricultural economics from Oregon State University. She chaired the Committee on Long-Range Soil and Water Conservation and currently serves on the Board on Agriculture.

Jeff Bennetzen is professor in Purdue University's department of biological sciences, director and co-founder of the International Grass Genome Initiative, and director of the Purdue Genetics Program. His research interests include plant genome organization, function, and evolution; particularly as determined by comparative genome mapping. He is also interested in the hyper-evolution of plant disease resistance genes and compensatory changes in the pathogen. Bennetzen received his undergraduate degree in biology from the University of California at San Diego, and his Ph.D. degree in biochemistry and genetics from the University of Washington in Seattle.

Hans Bohnert was educated at the University of Heidelberg in Germany and received his Ph.D. degree in physiological chemistry. He is professor in the departments of biochemistry, molecular and cellular biology, and plant sciences at the University of Arizona. His laboratory research focuses on the molecular genetics and physiology of environmental stress responses in plants with particular emphasis on the gene expression and metabolic changes of plants challenged by high salinity ("salt stress"). Metabolic engineering is used to transfer genes for entire stress-protective pathways into stress-sensitive plants.

Leonard S. Bull is associate vice provost at North Carolina State University, Raleigh, for International Programs. He received his undergraduate and masters degrees in dairy science and dairy cattle nutrition from Oklahoma State University and his Ph.D. degree in animal nutrition from Cornell University. Following postdoctoral study in medical physiology at the University of Virginia, he held faculty positions at University of Maryland, University of Kentucky, and University of Maine. Bull's research interests include energy metabolism, digestive physiology, and energy intake regulation relative to body composition in animals. He currently serves on the Board on Agriculture.

Sam Buttram is manager of genetic research for Kansas at Dekalb Swine Breeders, Inc., and is responsible for the testing, selection, and genetic improvement of Dekalb's Genetic Nucleus Lines. He is also responsible for scheduling the production of great-grandparent and grandparent lines at Dekalb's company-owned locations and is involved in the testing and development of new lines. Buttram's educational background includes his undergraduate degree in agriculture from Western Kentucky University, his masters degree in animal nutrition from Texas Tech University, and his Ph.D. degree in animal breeding and statistics from Iowa State University.

Anthony J. Cavalieri is vice president of Pioneer Hi-Bred International, Inc. Prior positions include director of the department of research specialists, a group providing support to Pioneer's plant breeders in pathology; director of technology support which included the department of research specialists, technology access, regulatory affairs, and the data management department; and director of trait and technology development which includes Pioneer's biotechnology effort as well as related sciences. He received his Ph.D. degree from the University of South Carolina in biology with major emphasis in plant physiology and completed his postdoctoral research on soybean water relations at the University of Illinois.

Mary Elizabeth Clutter is assistant director of biological sciences at the National Science Foundation. Her research interests include the function of polytene chromosomes in plant embryos. Prior to her career at the National Science Foundation, Clutter worked at Yale University in biological research. She received her undergraduate degree from Allegheny College and her Ph.D. degree in botany from the University of Pittsburgh.

Robert James Cook is research leader of the Root Disease and Biological Control Research Unit with the U.S. Department of Agriculture, Agricultural Research Service at Washington State University. He is a leader in the field of biocontrol of plant diseases. He pioneered the use of bacteria that grow on root surfaces and controlled fungal diseases not amenable to control by other meth-

ods. By genetic analysis of the effective bacterial strains, he determined that phenazine-type antibiotics are involved in biocontrol of root infecting pathogens. Cook is member of the National Academy of Sciences. At North Dakota State University he obtained both his undergraduate and his masters degrees; he continued his studies at the University of California, Berkeley where he received his Ph.D. degree.

Robert Cook-Deegan's primary research areas include genetics, history of the genome project, bioethics, neuroscience, cancer, mental health, dementia, and science and health policy. He is senior program officer with the Institute of Medicine's National Cancer Policy Board at the National Academy of Sciences. Cook-Deegan received his undergraduate degree from Harvard College and his masters degree from the University of Colorado.

Neal G. Copeland is director of the Mammalian Genetic Laboratory and head of the Molecular Genetics of Oncogenesis Section of the ABL-Research Program at the NCI-Frederick Cancer Research and Development Center. Prior to his current position he worked at the University of Cincinnati College of Medicine and the Jackson Laboratory in Bar Harbor Maine. Copeland received his undergraduate degree in biology and his Ph.D. degree in biochemistry from the University of Utah.

William DeLauder is president of Delaware State University. Prior to assuming the office of president, DeLauder served as associate professor of chemistry, acting chair and then full professor and chair of the chemistry department. DeLauder received his undergraduate degree from Morgan State College and his Ph.D. degree in physical chemistry from Wayne State University. As a postdoctoral fellow, he conducted research in physical biochemistry at the Centre de Biophysique Moleculaire du C.N.R.S. in France. His research on the physical properties of macromolecular systems and on the fluorescence properties of proteins has been published in leading scientific journals. In 1990 he was appointed to the National Advisory Council of the National Institute of General Medical Sciences of the National Institutes of Health. He currently serves on the Board on Agriculture.

Daniel Drell is biologist at the U.S. Department of Energy's Human Genome Program in the Office of Health and Environmental Research. His major responsibilities have included the DOE Microbial Genome Program, Bioremediation and Its Societal Implications and Concerns (BASIC) of Natural and Accelerated Bioremediation Research (NABIR) Program, and the Bioinformatics of Human Genome Program. Drell received his undergraduate degree Magna cum Laude from the department of biology at Harvard College. He continued his studies at

the University of Alberta, Edmonton where he received his Ph.D. degree from the department of immunology.

Kellye Eversole is president of Eversole Associates. She is consultant to the National Corn Growers Association where she has been primarily responsible for federal governmental relations for the National Corn Genome Initiative, mapping and sequencing the maize genome. She has been involved in the development of research legislation and policy matter affecting the agricultural industry, including the 1981, 1985, 1990, and 1996 farm bills.

Nina Fedoroff is interested in molecular biology, genetics, gene regulation, and transposable elements. She has worked on RNA replicase of the bacteriophage ϕ 2, the structure of nuclear RNAs, the primary structure of the 5s ribosomal RNA gene cluster of *Xenopus laevis*, the genetics, structure and regulation of maize transposable elements, and the use of maize transposons in gene discovery. Building upon the work of Barbara McClintock, Fedoroff elucidated the sequence of some of these elements, demonstrated their utility of gene cloning and was instrumental in converting the study of plant transposable elements into one accessible by molecular techniques. Fedoroff works in the Biotechnology Institute at Pennsylvania State University. She is member of the National Academy of Sciences. She received her undergraduate degree from Syracuse University and her Ph.D. degree in molecular biology from Rockefeller University.

Essex Finney, Jr. is recently retired associate administrator of the Agricultural Research Service. He is a member of the National Academy of Engineering and serves on the Board on Agriculture. His primary research interests are in development of rapid nondestructive methods for measuring quality characteristics of food and other agricultural products. He served as senior policy analyst at the Office of Science and Technology Policy, Executive Office of the President; associate director of the North Atlantic Area, ARS/USDA; and director, Beltsville Agricultural Research Center, ARS/USDA. Finney received his undergraduate degree in agricultural engineering from Virginia Polytechnic Institute and State University, his masters degree from Pennsylvania State University, and his Ph.D. degree in agricultural engineering from Michigan State University.

Cornelia Flora is director of the North Central Regional Center for Rural Development and professor of sociology at Iowa State University. Her research interests include rural America and global restructuring; agriculture and communities; science and sustainability; and rural economic development through self-development strategies. Prior to Flora's appointment at Iowa State University, she was professor and head of the department of sociology at Virginia Polytechnic Institute and State University, University Distinguished Professor at Kansas State University, and program advisor for agricultural development at the Ford Foun-

dation. She received her undergraduate degree at the University of California, Berkeley and her masters and Ph.D. degrees in rural sociology at Cornell University. Flora currently serves on the Board on Agriculture.

David John Galas is president, CEO, and chief scientific officer at Darwin Molecular Corporation. His research interests include molecular genetics of transposition, and the mechanisms and consequences of these recombination processes; and molecular interactions of DNA with proteins, and their consequences in gene control and recombination. Galas secured all of his degrees in physics from the University of California; studying at Berkeley for his undergraduate degree, and at Livermore for his masters and Ph.D. degrees. He currently serves on the Board on Biology.

Jim Gibb is vice president of the National Cattlemen's Beef Association's Center for Quality and chairman of the Beef Improvement Federation (BIF) Systems Committee. He also serves on the executive committee of the International HACCP Alliance. Gibb received his undergraduate degree from the University of Illinois and continued his studies at Colorado State University where he completed his Ph.D. degree. Honors he has received include induction into the American Polled Hereford Association Hall of Merit for Education and Research, and BIF's continuing service award.

Major M. Goodman has applied both traditional and modern methods to the study of evolution and genetic diversity in cultivated plants. For example, he developed a variance ratio method for measuring evolutionary divergence in maize and also demonstrated for the first time that a dehydrogenase enzyme is essential for normal embryogenesis in plants. He is professor in the department of crop science at North Carolina State University. Goodman studied genetics and received his masters and his Ph.D. degrees from North Carolina State University. He completed his undergraduate work at Iowa State University. Goodman is member of the National Academy of Sciences.

Teresa Gruber received her undergraduate degree in agriculture from the University of Nebraska. She continued her studies to obtain her Ph.D. degree in plant breeding from the University of Minnesota, and her J.D. degree from Georgetown University Law Center. Gruber is Trade and Agriculture counsel, and legislative assistant to U.S. Senator J. Robert Kerrey from Nebraska. She is responsible for agriculture, agricultural appropriations, and trade issues for members of the Senate Agriculture Committee and Senate Finance Committee.

George R. Hallberg is associate director and chief of Environmental Research at the University of Iowa Hygienic Laboratory. He is adjunct professor at the University of Iowa and Iowa State University. He served as a committee member

for the Board on Agriculture's report *Soil and Water Quality: An Agenda for Agriculture* and currently serves on the Board on Agriculture. Other service includes the USEPA National Advisory Council for Environmental Policy and Technology. His research interests include environmental monitoring and assessment, agricultural-environmental impacts, integrated farm management, chemical/nutrient fate and transport, and health effects of environmental contaminants. He received his undergraduate degree from Augustana College and his Ph.D. degree in geology from the University of Iowa.

Wallie Hardie is president of the National Corn Growers Association (NCGA). He farms 2,300 acres near Fairmount, North Dakota, where he grows corn, soybeans, and sugar beets. Hardie is a founding member of the North Dakota Corn Growers Association. He holds undergraduate and masters degrees in agricultural economics from North Dakota State University. Prior to farming full time, Hardie worked as research specialist for the University of Minnesota Agricultural Experiment Station in St. Paul, Minnesota.

Edward Kahealani Kaleikau is division director of the plants division in the National Research Initiative Competitive Grants Program at the U.S. Department of Agriculture's Cooperative State Research, Education, and Extension Service. He is the USDA representative to the NSF/DOE/USDA *Arabidopsis* Sequencing Project and serves on numerous governmental committees on genetics and biotechnology. Kaleikau received his Ph.D. in plant genetics from Kansas State University and was a postdoctoral research fellow at Stanford University where he worked on plant mitochondrial gene expression.

Arthur Kelman is university distinguished scholar at North Carolina State University. Kelman's research areas include biochemical mechanisms of resistance in plants to bacterial diseases; nature and activity of pectic and proteolytic enzymes in plant tissues; the role of calcium nutrition in resistance of plants to phytopathogenic bacteria causing tissue maceration; the nature of wound healing processes in plants; and techniques for detection and identification of phytopathogenic bacteria. He has experience in science, technology, and public policy. He is member of the National Academy of Sciences and has served on numerous committees and boards. Kelman received his undergraduate degree in biology from the University of Rhode Island and his masters and Ph.D. degrees in plant pathology from North Carolina State University.

Hans J. Kende is professor at the Michigan State University—Department of Energy Plant Research Laboratory. He has made major discoveries about plant hormones; Kende showed that cytokinins are synthesized in roots and inhibit senescence in leaves, he isolated ACC synthase, and elucidated the role of ethylene and gibberellin in the growth of rice. Kende is recognized, world-wide, as a

leader in plant physiology. He is member of the National Academy of Sciences and received his Ph.D. degree in botany from the University of Zurich.

Gurdev S. Khush is principal plant breeder and head of the Division of Plant Breeding, Genetics, and Biochemistry at the International Rice Research Institute in the Philippines. Khush is the premier breeder of rice in the world today. His contributions in rice genetics and breeding have provided the scientific basis for the Green Revolution in Asia and other parts of the world. Khush is a National Academy of Sciences foreign associate member. He received his undergraduate degree from Punjab University in India and completed his Ph.D. degree in genetics at the University of California at Davis.

T. Kent Kirk is professor in the department of bacteriology at the University of Wisconsin, and adjunct professor of wood and paper science at North Carolina State University. He is past director of the Institute for Microbial and Biochemical Technology, Forest Products Laboratory, USDA Forest Service. Kirk's scientific expertise is in the area of the biochemistry of wood decay by fungi and the industrial application of fungi and enzymes. He received his undergraduate degree from Louisiana Polytechnic Institute, and masters and Ph.D. degrees from North Carolina State University. Kirk is member of the National Academy of Sciences and currently serves on the Board on Agriculture.

Michael J. Knapp is vice president for program development at the National Center for Genome Resources where he is responsible for creating the Genetics and Public Issues program, overseeing the company's external communications, managing government relations, and developing and implementing fund-raising plans. He received his undergraduate degree in economics and government from the College of William and Mary in Virginia.

Brian A. Larkins is Porterfield Professor of Plant Sciences in the department of plant sciences at the University of Arizona. His research has made outstanding contributions to understanding the biology of seed proteins. He characterized the zein seed protein genes of maize and illuminated the events leading from gene expression through to deposition of zein proteins into protein bodies of the maize kernel with significant implications for protein nutrition of humans worldwide. Larkins is member of the National Academy of Sciences. He received both his undergraduate degree and his Ph.D. degree in botany from the University of Nebraska.

Dan Laster is center director for the U.S. Meat Animal Research Center (MARC) at the U.S. Department of Agriculture's Clay Center in Nebraska. He obtained his undergraduate degree in agriculture at the University of Tennessee. He continued his studies at the University of Kentucky, where he received his masters

degree in animal nutrition, and at Oklahoma State University, completing his Ph.D. degree in animal breeding and physiology. Prior to his experience with the USDA, Laster held professorship, research, and managerial positions in universities and in the private sector.

C. S. Levings is Professor Emeritus at North Carolina State University. His research has opened up the field and made major contributions to higher plant mitochondrial genetics. He demonstrated only maternal mitochondrial inheritance, uniqueness of organization of higher plant mitochondrial DNA, mitochondrial DNA differences for male sterility, and plasmid-like DNAs which have the potential of acting as vectors in the genetic engineering of the mitochondrial genome. Levings is member of the National Academy of Sciences. He studied at the University of Illinois where he received his undergraduate degree in agricultural science and completed his masters and Ph.D. degrees in genetics and plant breeding.

Harley W. Moon is the F.K. Ramsey Chair in Veterinary Medicine at Iowa State University. He is member of the National Academy of Sciences and currently serves on the Board on Agriculture. Moon's research interests include vaccines for preventing *E. coli* infection in farm animals, livestock disease eradication, and prevention of edema disease in swine with genetically-modified vaccines. Prior to his current position, Moon was director at the ARS/USDA Plum Island Animal Disease Center, and professor in the department of veterinary pathology at The Ohio State University. He received his undergraduate, D.V.M., and Ph.D. degrees from the University of Minnesota.

William L. Ogren is Professor Emeritus at the University of Illinois and past research leader in the photosynthetic research unit of the Agricultural Research Service, U.S. Department of Agriculture. Ogren's research interests include biochemistry, physiology, and molecular genetics, and photorespiratory carbon metabolism. His past professional experiences include plant physiologist at ARS/USDA, and affiliate in the department of agronomy at the University of Illinois. Ogren received his undergraduate degree at the University of Wisconsin and his Ph.D. degree at Wayne State University. He is member of the National Academy of Sciences and serves on the Board on Agriculture.

Ronald L. Phillips is currently Regents' Professor in the department of agronomy and plant genetics at the University of Minnesota. His research includes initiating the regeneration of maize plants from tissue culture, investigating somaclonal variation, demonstrating reactivation of quiescent transposable elements in plants from tissue cultures, and documenting variable numbers of rRNA genes in maize inbreds and enormous amplification of DNA during endosperm development. He is member of the National Academy of Sciences. He studied at Purdue

University and received both his undergraduate and his masters degrees. Phillips continued his education and completed his Ph.D. work in genetics at the University of Minnesota.

Bernard Phinney is Professor Emeritus at the University of California, Los Angeles. He is member of the National Academy of Sciences and has served on many committees. His primary research interests revolve around plant hormones, especially the gibberellins (GAs) research on the biosynthesis and genetic control of the steps in the pathway; genetics analysis, cloning and gene expression of dwarf mutants (GA mutants) in maize; and identification and quantification of gibberellins in eukaryotes (plants). He received his undergraduate and his Ph.D. degrees in botany from the University of Minnesota.

Stephen D. Rounsley is assistant investigator in the department of eukaryotic genomics at The Institute for Genomic Research (TIGR). He finished his undergraduate studies in biological studies at the University of East Anglia in Norwich, United Kingdom. Rounsley continued his education at the University of California, San Diego, where he completed a thesis entitled "Diverse Roles for MADS box genes in the development of *Arabidopsis thaliana*," and received his Ph.D. degree.

Colin Guy Scanes is executive associate dean and associate director of the College of Agriculture and the Iowa Agricultural and Home Economics Experiment Station at Iowa State University. His educational background includes his undergraduate degree in biological chemistry and zoology and his doctorate of science degree from Hull University in the United Kingdom. Scanes completed his Ph.D. degree in zoology at Wales University. He has held past professorship and administrator positions at Rutgers University and the University of Leeds.

Peter A. Schad is vice president for bioinformatics and biotechnology at the National Center for Genome Resources; and adjunct professor at the University of New Mexico in Albuquerque and New Mexico State University in Las Cruces. Schad received his undergraduate degree in biology from Rutgers University in New Jersey. He obtained his masters degree in microbiology from The Catholic University of America in Washington, D.C. Schad continued his studies at Oregon Health Sciences University to receive his Ph.D. degree in microbiology and immunology.

Ronald Sederoff is the leader in the molecular genetics of forest trees. With colleagues, he showed that in *Sequoia sempervirens*, chloroplast and mitochondrial DNA are paternally inherited. His group was the first to transfer a gene into a conifer. His research provides the base supporting the genetic engineering of forest trees. The purpose of his current work is to better understand the biology

of forest trees and to use this information to accelerate breeding. In spite of the long generation times of these organisms, Sederoff and colleagues have developed methods for genomic mapping of individual trees and applied those methods to complex trait analysis, particularly growth and disease resistance. He is a National Academy of Sciences member and serves on the Board on Biology. Sederoff received his Ph.D. degree in zoology from the University of California in Los Angeles.

George E. Seidel, Jr. is a professor of physiology at Colorado State University. His areas of study include superovulation and embryo transfer, in vitro oocyte maturation and fertilization, cryopreservation of livestock embryos, and embryo microsurgery. He received his undergraduate degree from Pennsylvania State University, and his masters and Ph.D. degrees in reproductive physiology from Cornell University. Seidel is member of the National Academy of Sciences and currently serves on the Board on Agriculture.

Christopher R. Somerville is director of the Carnegie Institution of Washington. His research has applied the techniques of microbial genetics to the dissection of plant metabolic and developmental pathways in *Arabidopsis thaliana*. He resolved important problems in photosynthesis, plant lipid metabolism, and plant hormone response, and his success with this approach has established *Arabidopsis* as the quintessential organism in plant genetics research. Somerville is a member of the National Academy of Sciences. He completed his undergraduate, masters, and Ph.D. degrees in genetics at the University of Alberta.

Shauna Christine Somerville's research is primarily focused on the function of genes that confer resistance to powdery mildew in *Arabidopsis*. Her lab is engaged in map-based cloning of genes and the use of gene chips or gene microarrays to measure the expression of thousands of genes simultaneously. Her work had developed a micro array which monitors the expression of approximately 600 disease-related genes. Somerville contributed to the *Arabidopsis* EST project that has produced 35,000 partial gene sequences of *Arabidopsis* genes and she has been engaged in mapping large numbers of disease-related ESTs on to the *Arabidopsis* genome by anchoring to YAC clones. Somerville received her undergraduate degree in genetics and her masters in plant breeding from the University of Alberta. She received her Ph.D. degree in agronomy and plant physiology from the University of Illinois.

James Tavares is program manager with the Division of Energy Biosciences at the U.S. Department of Energy. Prior to joining DOE, he was with the Board on Agriculture of the National Research Council, National Academy of Sciences, where he worked on many reports addressing issues in agriculture, biotechnology, natural resources, and science research policy. Tavares also served as asso-

ciate executive director of the Board. He is a graduate of Brown University and earned his Ph.D. degree in biology from Yale University.

Shirley M. Tilghman is currently Howard A. Prior Professor of the life sciences and investigator at the Howard Hughes Medical Institute at Princeton University. Working in P. Leder's laboratory, Tilghman was the first to clone the beta-globin cDNA of mice and show that the gene had an intron. She identified the H19 gene in mice, an early example of parental imprinting, and showed how this gene and its regulatory elements initiate and maintain parental imprinting. Other research interests include using genetics to understand the role of the genes involved in the development of melanocytes, the pigment-producing cells in the mouse, and studying genes whose expression pattern is determined by whether the gene is inherited from mothers or fathers. Tilghman is a National Academy of Sciences foreign associate member and Institute of Medicine member. She received her Ph.D. degree from Temple University in biochemistry.

Koichiro Tsunewaki's research has provided an unparalleled understanding of the evolution of wheat and related species. He described informative and useful interactions of nearly 500 nuclear-cytoplasmic combinations produced through years of breeding, traced maternal lineage of wheats by analysis of chloroplast and mitochondrial DNAs, and discovered new male sterility, haploid, and other cytogenetic systems. He is professor in the department of bioscience at Fukui Prefectural University in Japan. Tsunewaki is foreign associate member of the National Academy of Sciences.

Barbara Valentino

Evolving Communications
Washington, D.C.

J. Craig Venter is the founder, president, and director of The Institute for Genomic Research (TIGR). Prior to the formation of TIGR, Venter worked at the National Institute of Neurological Disorders and Stroke at the National Institutes of Health. He developed a new strategy to generate expressed sequence tags (ESTs) in gene discovery which has revolutionized the biological sciences, particularly in the fields of genome research, molecular biology, and medical biology. Using this technology, the scientists at TIGR have accomplished important milestones in the structural, functional, and comparative analysis of genomes and gene products in viruses, eubacteria, pathogenic bacteria, archaea, and eukaryotes, both plant and animal, including humans. Venter completed his Ph.D. degree in physiology and pharmacology at the University of California, San Diego. He also received honorary doctor of science degrees from Monmouth College, West Long Branch in New Jersey, and from Emory & Henry College in Virginia.

Catherine E. Woteki is the Under Secretary for Food Safety at the U.S. Department of Agriculture. Her educational background includes her undergraduate degree in biology and chemistry at Mary Washington College in Virginia, and her graduate degree in human nutrition at Virginia Polytechnic Institute and State University. Prior to her commitment to the USDA, Woteki has held directorships at the Office of Science and Technology Policy and the Institute of Medicine, National Academy of Sciences.

James B. Wyngaarden is Professor Emeritus at Duke University. His expertise includes the control of purine synthesis; purine metabolism in normal and gouty man; metabolism of iodine and steroids; oxalate synthesis; inborn errors of metabolism and human genetics. He studied medicine and received his degree at the University of Michigan and has received numerous honorary doctorate of science degrees. He is a member of the National Academy of Sciences and Institute of Medicine. Wyngaarden's experience spans pharmaceutical corporate and industry administration; federal government agency administration; and hospital and medical center administration.

Roger E. Wyse is dean of the College of Agriculture and Life Sciences and director of the Wisconsin Agricultural Experiment Station at the University of Wisconsin, Madison. His prior positions include professorship, administrative and research responsibilities with Rutgers University, Utah State University, Michigan State University, and the U.S. Department of Agriculture. Wyse received his undergraduate degree in agronomy from The Ohio State University. He completed his masters degree in crop science, his Ph.D. degree in plant physiology, and his post doctoral study in biochemistry, all at the University of Michigan.

Bob Zimbleman is executive vice president of the American Society of Animal Science. His specialty is anchored in reproductive endocrinology, animal health, and science policy. Zimbleman received his undergraduate education at Colorado State University and completed his masters and Ph.D. degrees at the University of Wisconsin. His research experience includes improving fertility, cell division and anaphase movement of chromosomes, and biomedical ethics.

James J. Zuiches became dean of the College of Agriculture and Home Economics at Washington State University, following his work as program director for Food Systems and Rural Development at the W.K. Kellogg Foundation. He received his masters and Ph.D. degrees in sociology from the University of Wisconsin. Zuiches currently serves on the Board on Agriculture and was a member of the Committee on the Future of Colleges of Agriculture in the Land Grant University System.

