

**A Long-range Plan for the
Multinational Coordinated
Arabidopsis thaliana Genome
Research Project**



The National Science Foundation has TDD (Telephonic Device for the Deaf) capability, which enables individuals with hearing impairment to communicate with the Division of Personnel and Management about NSF programs, employment, or general information. This number is (202) 357-7492.

Any opinions, findings, conclusions, or recommendations expressed in this report are those of the participants and do not necessarily reflect the views of the National Science Foundation.

Cover drawing courtesy of Dr. Chris Somerville, Michigan State University

PREFACE

This document, “a Long-range Plan for the Multinational Coordinated *Arabidopsis thaliana* Genome Research Project”, was developed by the international community of scientists based on discussions that took place at four workshops sponsored by the National Science Foundation as well as discussions with their fellow researchers. A draft document was distributed to all participants of the Fourth International Conference on *Arabidopsis* Research in Vienna in June, 1990. Many of the Conference participants met and discussed the document during the Vienna meeting, and their comments have been incorporated into this final document.

The following scientists either participated in some of the workshops and/or contributed directly to the development of this document.

- Dr. Frederick Ausubel, Harvard University, U.S.A.
 - Dr. Vicki L. Chandler, University of Oregon, U.S.A.
 - Dr. Ronald W. Davis, Stanford University, U.S.A.
 - Dr. Caroline Dean, Cambridge Laboratory at Norwich, U.K.
 - Dr. Elizabeth Dennis, CSIRO, Australia
 - Dr. Kenneth Feldman, DuPont Company, U.S.A.
 - Dr. Gerald Fink, Massachusetts Institute of Technology, U.S.A.
 - Dr. Richard B. Flavell, John Innes Institute, U.K.
 - Dr. Charles S. Gasser, Monsanto Company, U.S.A.
 - Dr. Howard M. Goodman, Harvard University, U.S.A.
 - Dr. Maarten Koornneef, Agr Univ of Wageningen, The Netherlands
 - Dr. Elliot Meyerowitz, California Institute of Technology, U.S.A.
 - Dr. Oliver E. Nelson, University of Wisconsin, U.S.A.
 - Dr. Kiyotaka Okada, National Institute of Basic Biology, Japan
 - Dr. Jim Peacock, CSIRO, Australia
 - Dr. Jozef Schell, Max Planck Institute, Fed Rep of Germany
 - Dr. Yoshiro Shimura, University of Kyoto, Japan
 - Dr. Chris Somerville, Michigan State University, U.S.A.
 - Dr. Marc van Montagu, University of Gent, Belgium
- and
- Participants of the Fourth International Conference on *Arabidopsis* Research,
June 2-5, 1990, Vienna, Austria

A Long-Range Plan for the Multinational Coordinated *Arabidopsis thaliana* Genome Research Project

Executive Summary

A long-range plan for a multinational coordinated *Arabidopsis thaliana* genome research project has been developed by the international community of scientists who are actively engaged in fundamental plant biology research. The project was developed based on the recognition that a profound understanding of plant biology is essential in order to meet the immediate and future challenges facing world agriculture and the global environment; that the use of an experimental model system is extremely effective in studying biology; that *Arabidopsis* is a useful model system for flowering plants; and that international coordination is necessary for rapid and efficient advances in *Arabidopsis* genome research.

Mission: The mission of the project is to identify all of the genes by using a functional biological approach leading to determination of the complete sequence of the *Arabidopsis* genome by the end of this century.

Scientific Goals: The ultimate goal of the project is to understand the physiology, biochemistry, and growth and developmental processes of a flowering plant at the molecular level, using *Arabidopsis* as an experimental model system.

Program Objectives: The international community of scientists has agreed to collaborate on the following objectives.

1. Identification and characterization of the structure, function, and regulation of *Arabidopsis* genes.
2. Development of technologies for plant genome studies.
3. Establishment of biological resource centers.
4. Development of an informatics program to facilitate exchange of research results.
5. Development of human resources.
6. Support of workshops and symposia.

International Collaboration: An important principle for implementing the proposed multinational cooperation is to guarantee free exchange of ideas and information through open communication and interactions at the scientists' level. It is understood that there will be no pooling of funds between cooperating countries and each entity will lend financial support to the activities of its own scientists. Multinational committees have been formed to coordinate activities related to biological resource centers and informatics. The Multinational *Arabidopsis* Science Steering Committee has been established to implement overall research coordination. This committee will annually review scientific progress and identify needs and new opportunities for the *Arabidopsis* research community. The committee will also act in an advisory capacity to various national funding agencies.

A Long-Range Plan for the Multinational Coordinated *Arabidopsis thaliana* Genome Research Project

I. Introduction

People depend on plants for food, energy, building materials, and numerous industrial products. Plants also play a vital global role in regulating levels of oxygen and carbon dioxide gases and in the recycling of carbon and nitrogen. The world population is expanding rapidly, industrial needs are increasing and diversifying, and global climate changes are being recorded that are likely to affect plant growth.

At the present time, there is, therefore, a pressing need to develop a deeper understanding of fundamental aspects of plant biochemistry, physiology, development and genetic mechanisms. Such knowledge will permit us to manipulate them genetically so as to help manage human and global needs. For this to occur, it is highly desirable to discover the characteristics and functions of all the genes that determine the physiological and developmental processes of a plant. Until recently this was an unrealistically ambitious objective. Now, however, following technical developments in molecular biology, there exists an exciting opportunity to achieve this goal.

The goal of understanding the genetic basis of the physiological and developmental processes of plants is sufficiently ambitious that it can only be accomplished by studying a plant species that has many advantages for the research and that will be studied by sufficient numbers of scientists as a model system. A large proportion of the scientific community has chosen *Arabidopsis thaliana* as its model flowering plant species, and work is already progressing well to establish a worldwide coordinated research base. With suitable funding, and the exploitation of established and newly developed techniques for modifying genes, it is projected that within a foreseeable future a significant proportion of all the genes can be identified and their functions explored by existing genetic methods. However, since many genes will not be amenable to conventional genetic analysis, it will be necessary to determine the DNA sequence of the entire genome as a prelude to the application of projected new methods for analyzing gene function.

The value, to plant biology and to biology in general, of characterizing the complete genome of a plant should not be underestimated. Knowledge of the structure of the genome in one plant will enable equivalent genes and processes to be dissected and modified in the important crop and tree species. Exploitation of a model species is undoubtedly the quickest way to discover most genes in economically important plant species.

Arabidopsis thaliana, a member of the mustard family, is distributed widely over Europe, Asia, North America and Africa, with a limited distribution reported in Australia. Although related to important crop species, *Arabidopsis* itself has no agronomic value. For some time, *Arabidopsis* has been recognized by classical geneticists as a good experimental organism due to its small size, short life cycle and the ease of obtaining a large population. However, it has only been in the last five years or so that *Arabidopsis* has gained its unprecedented popularity among plant biologists, particularly molecular and developmental biologists, as the experimental model system of choice. For example, a quick survey of publications cited in *Biological Abstracts* using *Arabidopsis* as a key word yielded 36 papers in 1984 and 216 papers in 1989. The number of participants at the *Arabidopsis* meeting in 1987 was around 200 representing about 40 laboratories. The number doubled to 400 at the 1989 meeting, representing about 80 laboratories from 14 countries. It is worth

noting that this rapid growth of the field arose without any formal effort to develop *Arabidopsis* as a model system for research in higher plant biology.

The emergence of *Arabidopsis* as a useful model system for flowering plants offers a tremendous opportunity for plant biology. When one considers scientific advances made through the study of model systems such as *E. coli*, *D. melanogaster*, and *C. elegans*, the advantage of having a model system for higher plants is obvious. The need for a convenient model system is particularly acute in the area of molecular biology, as our knowledge of gene structure, function, and regulation of higher plant systems is far behind that of bacterial or animal systems. Recent progress in the area of genome mapping and of sequencing of specific genes amply demonstrates that *Arabidopsis* will be as useful to plant biology, and possibly to higher eukaryotes, as *E. coli* has been to microbiology in particular and to biology in general.

The advantage of *Arabidopsis* genome research is recognized by researchers in many countries including the U.S., the United Kingdom and other members of the European Community, Australia, and Japan. In order to ensure scientific advances without unnecessary duplication of effort, full cooperation of all nations involved in *Arabidopsis* genome research is essential. This document has been developed with involvement of the multinational community of scientists. Hopefully, the *Arabidopsis* Genome Research Project will become a model for multinational collaboration in plant biology research.

II. Mission Statement

The mission of the Multinational *Arabidopsis* Genome Research Project is to identify all of the genes by any means, and to determine the complete sequence of the *Arabidopsis thaliana* genome before the year 2000.

III. Current Status of *Arabidopsis* Genome Research Projects in Several Countries

Major research initiatives that use *Arabidopsis* for the benefit of biotechnology, agriculture, and general biological research are currently underway in many countries. The following is a brief summary of the current status of those efforts.

1) UNITED STATES: Ongoing research on *Arabidopsis* in the United States is substantial. For example, the NSF, NIH, USDA, and DOE already support general *Arabidopsis* research at the level of approximately \$6 M per year. The National Science Foundation (NSF) has taken the lead in developing the *Arabidopsis* Genome Research Project as a research initiative for the U.S. and has requested an increase of \$5 million for this effort in its 1991 budget. The initiative will help accelerate the already active field of *Arabidopsis* research to move at a faster pace and help encourage more scientists take advantage of *Arabidopsis* as an experimental model system for higher plant biology. In addition to the NSF, the NIH Center for Human Genome Research includes *Arabidopsis* as a model system for genome research. The Department of Energy (DOE) is also an important supporter of *Arabidopsis* research. The U.S. Department of Agriculture (USDA) has initiated a program on plant genome research and as part of that initiative, supports *Arabidopsis* genome studies. The NSF, NIH, DOE, and USDA have agreed to work together and coordinate activities relevant to *Arabidopsis* genome research.

2) UNITED KINGDOM: The Agricultural and Food Research Council (AFRC) of U.K. has established a major research initiative in plant molecular biology. AFRC provided 3.5 million Pounds (\$6.1 million) in 1989/1990, with a projected increase to 4.5 million Pounds (\$7.9 million) in 1990/1991 and to 6 million Pounds (\$10.5 million) in later years. AFRC is spending a third of this fund to support a coordinated program on the molecular biology of plant development using *Arabidopsis* as an experimental model system. In support of the coordinated program, a library of overlapping clones covering the entire *Arabidopsis* genome is being developed at the Cambridge Laboratory, J.I. Center for Plant Science Research. This effort is being coordinated with similar efforts that have been on-going for some time in the U.S.

3) EUROPEAN COMMUNITIES: The European Commission has launched in May, 1990, a transnational *Arabidopsis* genome research project involving 29 laboratories which will be coordinated together, and directly supported with a budget of 3 million ECU for the initial phase (1991-92). The participants have been selected by peer review after a call for proposals closed in December 1989. This is one of the few large-scale "targeted" projects (so-called T-projects) which the Commission established within the European Biotechnology Program, BRIDGE. The coordinator of this project is M. Bevan (Norwich) and the associated laboratories have organized their work on three levels; the resource center and physical mapping, the development of new methods for gene identification, and the search for genes and functions for selected developmental pathways. An additional budget is likely to be adopted for the second phase (1993-94) to maintain the group already created, and to possibly expand its activities in a wider international context.

4) AUSTRALIA: The Australian participation in the program is centered in Canberra at the Australian National University (ANU) and at CSIRO, Division of Plant Industry. The Department of Industry, Technology and Commerce (DITAC) of the Australian Government has provided a grant of A\$150,000 to begin this effort. CSIRO and ANU have a number of research projects (about A\$ 1 M) using *Arabidopsis*.

5) JAPAN: As part of the Human Genome program, the Ministry of Education, Science and Culture and the Science and Technology Agency are considering the inclusion of *Arabidopsis* and rice as model systems.

6) OTHER NATIONS: Brazil, Canada, Israel, Korea, Mexico, and a number of Western European (e.g., Austria, Finland, Sweden, and Switzerland) and Eastern European (e.g., Czechoslovakia, Poland and USSR) nations are known to have active *Arabidopsis* laboratories.

IV. Program Goals

The ultimate goal of the *Arabidopsis* Genome Research Project is to understand the physiology, biochemistry, and growth and developmental processes of a generic flowering plant at the molecular level, using *Arabidopsis* as an experimental model system. This work is expected to complement the research going on in other plant species. As an aid to achieve this ultimate objective, the immediate goal of the project is the identification and characterization of the structure, function, and regulation of *Arabidopsis* genes. An emphasis will be on the completion of the development of physical maps including the intercollation of the genetic and physical maps, the sequencing of specific regions of chromosomes that are of biological interest, the development of technologies for plant genome studies, training of young scientists, and the establishment of mechanisms for the collection and dissemination of biological resources and data resulting from the *Arabidopsis* Genome Research Project. Given the rapid advances in sequencing technologies and the lack of repetitive sequences

in the *Arabidopsis* genome, it is likely that substantial sequencing of the *Arabidopsis* genome will be accomplished without a systematic effort to simply sequence the chromosomes from one end to the other.

A close collaboration with the other genome projects, such as the Human Genome Project, is important to avoid duplication of effort and to take advantage of any technological and scientific advances made by other efforts. Equally important is the continued encouragement of multinational cooperation at the level of bench scientists within the *Arabidopsis* research community. Further development of the spirit of multinational cooperation will be one of the more important, yet intangible, goals of the *Arabidopsis* Genome Research Project.

V. Scientific Objectives

1. Genome Analysis

The genome sequence will be approached in a series of steps, each of which will provide essential information for the sequencing and the sequence analysis, but each of which will also provide, long before the sequence is available, vital information of importance for ongoing research. Each of these steps may be carried out at a particular place, or center, but each may be done more felicitously and with less expenditure by groups of researchers functioning as "centers without walls".

- a) An ongoing objective, which must be a continuing emphasis for the course of the genome project, is the establishment and use of methods for defining genes and their functions at the organismal level: that is, the production of mutants and the analysis of their phenotypes. Knowledge of individual genes affecting significant processes in plant development, physiology, and growth is a prerequisite for interpretation of molecular information on the structure and organization of the genome.

A major goal is saturation mutagenesis to define as many genes as possible. This should include the development of facile methods for transposon tagging.

- b) A concurrent objective, and the first purely molecular objective, is to link up the existing collection of yeast artificial chromosome (or other large-fragment-containing) clones into an ordered set representing the entire genome. While this is a substantial task, it is possible with current technology: if the existing 200 clones representing mapped RFLPs are hybridized one at a time to a YAC library, we will have 200 YACs whose genetic map position will be known, representing more than half of the genome. Regions of the genome where, by chance, RFLP markers are densely arrayed will be entirely linked up almost immediately. If numbers of clones and their genetic map positions are available through the database, and copies of a gridded array of YAC clones are available through the clone center, any laboratory could map genes and isolate genetic loci with little effort, even when the physical map is only partly done.

Once the YAC library has been ordered, new probes can easily be mapped onto the genome by hybridization to the YAC filters. Until that time recombinant inbred lines, currently being developed in several laboratories, will constitute a permanent supply of material for RFLP mapping. These can be widely distributed and will generate linkage data which can be compared directly.

- c) After the physical linkage of YAC clones is established, it may be valuable to establish restriction maps for several infrequent-cutting enzymes. This can be done by hybridizing mapped YAC clones to gel blots of genomic digests. Large-fragment pulsed-field gel analysis will be required, but is within the limits of current technology. Such a map will serve as a check for accuracy of the ordered YAC map, and will

give the ability to rapidly identify chromosome rearrangements, which could lead to balancer stocks for classical genetic work.

- d) A further objective may be the production of a series of cDNA libraries representing all RNAs specifically present in a number of different tissues and cell types. Again, this is within current technology; such an effort will use existing subtractive library methods. These libraries would themselves be available through an *Arabidopsis* resource center, and an effort would be made to obtain the sequence of all of the cDNA clones in each library (after a cross-hybridization analysis, to avoid sequencing more than one clone of each cross-hybridizing class). This would provide immediate access to cell-type specific clones, and would give a molecular definition of each cell type and developmental stage, thus providing RNA markers by which the details of mutant phenotypes would be established. The sequence of the clones would enable interpretation of genomic sequences, by indicating the positions of genes and introns.
- e) The last objective presently envisioned would be to use ordered clones and restriction maps to sequence the entire genome. This sequence would be easily related to genetic map position, because each clone sequenced would derive from a mapped YAC clone. Further, comparison with the cDNA sequence database would identify the genes within the genomic sequence, and would reveal the cell types of expression of each genomic region.

1-YEAR GOALS:

- Institute projects to identify mutable genes with key phenotypes in physiology and development.
- Establish YAC libraries with the longest possible clones and the fewest chimeric clones as practicable.

2-YEAR GOALS:

- Link up ordered collection of YAC libraries.

5-YEAR GOALS:

- Complete chemical and insertional mutagenesis to saturation for developmental and physiological phenotypes.
- Begin cloning and analysis of identified genes.
- Initiate a concerted effort to fill in the gaps in DNA sequences of the genome. (It is expected that new technologies will have been developed by this time to facilitate this effort.)

10-YEAR GOALS:

- Complete genome sequence and continue analyzing gene functions and interactions of sequenced genes.

2. Technology Development

The *Arabidopsis* Genome Research Project offers unique opportunities to contribute technological advances for genome and plant research. Special opportunities may arise because of the differing characteristics of the plant genome in comparison to that of the human and animal genomes. Because of the relative ease in creating cytogenetic stocks, such as chromosomal substitution and duplication lines, as well as the ability to carry out breeding at will, plants are better suited for certain kinds of genetic studies, such as gene evolution and quantitatively inherited traits. Unique needs may also arise in an area such as the regeneration of whole plants from transformed cells. The *Arabidopsis* Genome Research Project will support research for the development of innovative and novel methods and techniques that will advance plant genome research.

Major coordinated efforts, such as the Human Genome project, will play a large part in the development of gene-sequencing and other genome-project-related technologies. However, as the discovery of restriction enzymes and of the polymerase chain reaction would indicate, individual scientists are also likely to discover and develop better and faster methods. While it will be important for *Arabidopsis* workers to maintain a close interaction with the human genome research community in order to take advantage of any technological advances applicable to the plant genome research, it is equally important to recognize that the *Arabidopsis* community will undoubtedly need to develop technologies of its own.

5-YEAR GOALS:

- Support innovative technologies that directly address characteristics unique to plant genomes.
- Support application and modification of technologies developed for non-plant genome systems to *Arabidopsis* genome research.

3. Biological Resource Center

The establishment of repositories of biological research materials is a high priority activity for the *Arabidopsis* research community. Having a central location to deposit and distribute seeds, cloned DNA fragments, libraries, etc. will make generally and immediately available the results of research, and will serve as central organizing points for the genetic and molecular studies that will provide for and make use of the results of the genome sequencing efforts. Such resource centers should satisfy the following requirements:

- A center should function as a point of collection, maintenance, cataloging, and distribution of seeds, including mutant strains and ecotype collections. Activities would include checking the mutant strains, making multiple mutant stocks for gene mapping, and providing strains to researchers on request.
- A center should store, maintain, catalogue, and distribute cloned DNA fragments, cloned genes, genomic fragments, and cDNA and genomic libraries. When a physical map is completed, contiguous genomic regions should be stored as clone subcollections, to be made available on request.
- A center must function as a source of information for the biological collection it houses. Complete information on the stored items including the origin and biological characteristics must be made available, and a compilation of such information should be published widely. In its role as an information clearinghouse, the center should maintain an up-to-date electronic information bulletin board concerning the availability of DNA clones, DNA libraries, mutants, marker lines, lists of approved grant projects, meeting announcements, *Arabidopsis* bibliography, and a directory of Who's Who in *Arabidopsis* research. Publication of a newsletter should be considered for those researchers who have no access to the electronic bulletin board.

The *Arabidopsis* research community recommends the creation of at least two such biological resource centers. Duplication is advisable for security reasons, if for nothing else. Whether there should be separate centers for seeds and DNA clones is an issue that needs to be decided by the organizers of resource centers with close consultation with the scientific community. All centers that house collections of biological resources for *Arabidopsis* research should coordinate their activities and develop a formal mechanism to maintain close collaboration on an on-going basis. For this purpose, an official liaison should be appointed to each center from the others.

1-YEAR GOALS:

- Appoint a multinational liaison committee to establish a common mechanism of communication among resource centers.

- Establish at least two biological resource centers, with advisory committees to maintain communication between the centers and the *Arabidopsis* research community.
- Operate functional electronic information bulletin board and/or publish newsletters.

5-10 YEAR GOALS:

- Maintain long-term support for the established resource centers.

4. Informatics: Data Collection, Management, Distribution, and Analysis:

The direct information product of the *Arabidopsis* Genome Research Project will be collections of data identifying and characterizing the structure, function, and regulation of the genes of *Arabidopsis thaliana*. These data sets will include many genome maps and DNA sequences. Since the data are expected to be of sufficient quantity and complexity that it would be impractical to maintain and use them without the use of computers, the development of appropriate computer tools and information systems will be essential for the ultimate success of the *Arabidopsis* Genome Research Project. Equally important will be the development of mechanisms by which the multinational community of scientists can share the data and make use of the database for their research.

Because *Arabidopsis* provides a useful model for flowering plants, it is anticipated that the information products resulting from the *Arabidopsis* Genome Research Project will be of interest to a wide-ranging community of users and researchers in fields as diverse as applied work in agriculture or biotechnology and fundamental studies in plant physiology, genetics, or development. The discoveries of the *Arabidopsis* Genome Research Project will also be of comparative interest to all workers on other genome projects, including the human genome initiative. Since the findings of the *Arabidopsis* Genome Research Project will be of use to all of these communities, the needs of these communities must be determined so that the *Arabidopsis* information resources can be designed to provide maximum utility to all users.

At present, it is not clear whether the most useful information product of the *Arabidopsis* Genome Research Project will be a single large database or a set of distributed, networked databases. It is also unclear how genome databases will be structured in the future and whether existing database technologies can be adapted to meet the overall, long-term needs of the *Arabidopsis* Genome Research Project. However, it is certain that the *Arabidopsis* genome databases will need to be comprehensive and up-to-date. And, it will be imperative that multiple databases be effectively linked with one another and that researchers have ready, electronic access to all relevant databases.

Given the extent and the diversity of the informatics and communication needs of the *Arabidopsis* Genome Research Project, it is important that at least one *Arabidopsis* informatics resource center be created and charged with serving the informatics and communication needs of the *Arabidopsis* research community. It is imperative that any such informatics resource center establish close working relationships with all other *Arabidopsis* informatics or biological resource centers. To facilitate its role as "publisher" of *Arabidopsis* research results, it is highly desirable that an informatics resource center establish a close working relationship with one or more professional societies.

In addition to database development, it will be vital to develop new methods and tools for the analysis and interpretation of genome maps and DNA sequences and other data produced by the *Arabidopsis* Genome Research Project. Success in these areas of genome informatics will require the development of a coordinated national and multinational program to make the information and analysis tools from this project readily available to the widest possible range of scientists in the most useful, timely, and cost-effective manner.

Currently, it is possible to describe the informatics and communication goals of the *Arabidopsis* Genome Research Project only in broad terms. Therefore, considerable refinement of the goals will be necessary as this program develops and as informatics technology improves with time. To this end, it is recommended that a multi-national committee (the *Arabidopsis* Data and Communications Committee) be established to provide long-term guidance for the informatics requirements of the *Arabidopsis* Genome Research Project.

The responsibilities of the *Arabidopsis* Data and Communications Committee should include:

- identification of the user needs and the researcher needs for the informatics components of the *Arabidopsis* Genome Research Project,
- identification of the electronic communication needs of the *Arabidopsis* Genome Research Project,
- identification of the information products that should result from the *Arabidopsis* Genome Research Project,
- establishment of priorities for both technical objectives and policy areas,
- serve as liaison with the informatics components of other genome projects, such as the Joint Informatics Task Force (JITF) of the human genome initiative and the Data Advisory Committee for the Plant Genome Project including their multinational components,
- provide a focus for multinational activities related to *Arabidopsis*-genome informatics, and
- drafting and maintaining a long-term plan (the *Arabidopsis* Informatics Plan) for managing the informatics components of the *Arabidopsis* Genome Research Project.

Specific issues to be addressed should include:

- genome database structures, management, and services,
- development of algorithms, software, and standards,
- development and maintenance of standardized nomenclatures,
- use of electronic networks for collection and distribution of genome information,
- training and education of informatics personnel,
- coordination of *Arabidopsis* genome informatics activities among laboratories and agencies, and
- coordination of *Arabidopsis* genome informatics activities with other genome projects.

The challenge will be not only to design information resources that can meet the growing needs for access and for increasingly sophisticated search capabilities, but also to keep up with the voluminous amount of information that will be produced at ever faster rates.

A number of research efforts are currently in progress to improve database design, software for database access, and data-entry procedures for genomic research. Many of these are associated with the human genome project. For example, the National Center for Biotechnology Information (NCBI) was recently established at the National Library of Medicine in the U.S. to create automated systems for knowledge about molecular biology, biochemistry, and genetics; and to pursue research in biological information handling, particularly

with respect to human molecular biology. Thus, the mission of the NCBI supports, directly, that of the Human Genome Initiative and, indirectly, that of all other genome initiatives. Consequently, the *Arabidopsis* Data and Communications Committee should interact closely with the NCBI to facilitate a rapid exchange of ideas and to minimize duplicated effort.

1-YEAR GOALS:

- Convene a multinational *Arabidopsis* Data and Communications Committee.
- Produce a first draft of an *Arabidopsis* Informatics Plan.
- Plan, design, and establish at least one center for *Arabidopsis* Information Resources.

5-YEAR GOALS:

- Develop effective software and database designs to support large-scale mapping and sequencing projects.
- Create database tools that provide easy access to up-to-date physical mapping, genetic mapping, chromosome mapping, and sequencing information and allow ready comparison of the data in these several data sets.
- Develop algorithms and analytical tools that can be used in the interpretation of genomic information.

5. Human Resource Development

An acute need exists for scientists well-trained in modern plant biology. The *Arabidopsis* Genome Research Project will create additional pressure for more trained personnel and at the same time will create additional opportunities for training. The objective here should be to train scientists who are well-versed in plant biology and are thoroughly familiar with cutting edge technologies applicable to genome research. As the *Arabidopsis* Genome Research project takes a biological approach, the project will provide the opportunity to train such scientists. In addition, opportunities for cross-disciplinary training will be provided for producing scientists who are conversant both with biology and computational sciences. *Arabidopsis* has already attracted scientists trained in non-plant biology disciplines into plant biology research. Therefore, it has a proven record of exerting positive influence on plant biology by broadening the human resource base. By contributing broader scientific perspectives as well, it has raised the visibility of plant science.

The exchange of students, as well as postdoctoral and senior researchers, under various training programs will be an excellent mechanism to ensure good communication and collaboration among scientists from various nations who are all engaged in *Arabidopsis* research.

The following mechanisms are considered to meet the training needs of the *Arabidopsis* Genome Research Project.

- a) Individual postdoctoral fellowships—These fellowships will provide support for both junior and senior post-doctoral scientists who wish to obtain further training in any scientific discipline relevant to plant genome research. The research area to be supported should be broad, and interdisciplinary endeavors should be particularly encouraged. Fellowship recipients should be allowed to carry out their research in the laboratory that best meets their specific training needs regardless of their nationality, the origin of the funds, or the location of the laboratory.
- b) Short-term exchange—The *Arabidopsis* community has identified a need to support short-term exchanges of personnel between laboratories to learn techniques, to compare data, to write papers, and other

activities that cannot be accomplished through telephone and other communication methods, and that require a short stay of a few weekends to a few months.

- c) Short courses—Hands-on courses devoted to *Arabidopsis* biology and research will provide an excellent opportunity to familiarize interested scientists with *Arabidopsis* as an experimental model organism. The format of currently existing short courses such as the one at the Cold Spring Harbor Laboratory and the EMBO or NATO courses could be adopted to meet the specific needs of the *Arabidopsis* research community.

ON-GOING GOALS:

- Support multinational *Arabidopsis* postdoctoral fellowships.
- Support short-term exchanges and short courses.

6. Workshops and Symposia

As in any scientific endeavor, research workshops and conferences are an important and integral part of the *Arabidopsis* Genome Research program. The *Arabidopsis* research community has been meeting periodically to discuss the latest scientific developments. The Fourth International Congress is scheduled to meet in Vienna, Austria, in June of 1990. There will be a symposium on *Arabidopsis* in Australia in September of 1990 supported by the Australian government. EMBO is planning a practical workshop on *Arabidopsis* in 1991. These general meetings to assess the state of science in *Arabidopsis* research are important and will be supported under the *Arabidopsis* Genome Research Project.

More focused workshops and symposia will also be supported when a group of scientists find needs for an intense discussion on a narrowly-focused subject, for example, a meeting to compare and assemble all available mapping information to build a consensus genetic/physical map for *Arabidopsis* or a meeting to discuss best strategies for the development of software for data handling.

An additional need is envisioned to support workshops to bring together scientists from diverse backgrounds with a common interest in *Arabidopsis* genome research. Examples of such workshop would be a meeting of biologists and informatics specialists, or a meeting of *Arabidopsis* genome researchers with human genome researchers. Another important example would be a meeting of *Arabidopsis* researchers with general plant geneticists to accelerate information transfer from a model system to crop plants.

ON-GOING GOAL:

- Support for workshops and symposia as needs are identified by the scientific community.

VI. Social and Legal Issues

1) Intellectual property rights

The proposed coordinated multinational project on *Arabidopsis* genome research will be successful only if all participants agree to share the results of their individual efforts.

All investigators participating in coordinated multinational *Arabidopsis* genome research efforts are expected to follow conventional practice in maintaining scientific integrity concerning the conduct of research and the dissemination of information and materials arising from the project. As a rule, the matter of intellectual property rights should be handled according to the legal convention of the country who provides funds for the individual research project in question. If any restrictions exist which potentially prevent free exchange of ideas, information and materials for research purposes, they should be identified and widely made known so that rights of individual investigators to intellectual property will not be inadvertently violated. In cases where an exchange of personnel is involved such as a visiting scientist or a postdoctoral research fellow, a clear understanding about the disposition of intellectual property rights should be made by all parties involved. When a research project is conducted under an official joint agreement between nations, such agreement should contain a clause on the intellectual property rights which is in accordance with the international agreement on intellectual property rights.

2) Release of biotechnology-derived plants into the environment

Should research supported under the *Arabidopsis* Genome Research Project lead to field testing of genetically-engineered plants, the investigators are expected to follow research guidelines and applicable regulations of the country where the field testing is being conducted.

VII. Multinational Collaboration

Scientists representing some of the countries cited in Section II met in conjunction with the *Arabidopsis* meeting in Bloomington, Indiana in October of 1989. It was agreed then that specific areas of cooperation needed to be rapidly implemented are the establishment of seed collections and data bases. It was also agreed that multinational committees should be named for both of these activities. Since the Indiana meeting, the United Kingdom has proceeded to set up a seed stock center at the University of Nottingham and a DNA clone center at the Cambridge Laboratory, J.I. Center for Plant Science Research in Norwich. The Japanese Ministry of Education, Science and Culture is planning to set up a seed stock center in Japan. The U.S. researchers are in the process of developing a plan for a resource center in the U.S.

It has become evident that a structured approach to coordinate overall collaboration of *Arabidopsis* research will be needed. At the initiative of NSF, the first meeting of the multinational steering committee was held in April, 1990 in Denver, Colorado, taking advantage of scientists gathering for a UCLA symposium in Colorado. The recommendations of the committee were (1) the members of the Denver meeting serve as an interim Multinational Science Steering Committee until a permanent Multinational Science Steering Committee be established at the Vienna meeting in June, 1990, and (2) to form two subcommittees, one on informatics and another on resource centers.

VIII. Implementation of the Multinational *Arabidopsis* Genome Research Project

The most important principle for implementing the proposed multinational cooperation is to guarantee free exchange of ideas and information through open communication and interactions at the scientist's level. A

basic premise of the proposed plan is that there will be no exchange of funds between cooperating countries and that each entity will lend financial support to the activities of its own scientists.

To coordinate the project, research scientists from each participating country or region should form individual national science steering committees. A Multinational Science Steering Committee will also be established, consisting of representatives selected by the various national science steering committees. Initially, the Multinational Science Steering Committee should consist of nine members: two from the U.K., two from the other E.C. communities, three from North America, one from Australia, and one from Japan. The standard term of each member of the Multinational Steering Committee is to be three years. However, to facilitate the orderly transition of members, initial appointees should be divided into three groups, with the groups serving for one, two, and three year terms, respectively.

Specific responsibilities of the Multinational Science Steering Committee include: (1) coordinate programmatic aspects of the *Arabidopsis* genome research, (2) communicate with the informatics and biological resource centers, (3) monitor and summarize progress of scientific activities of participating laboratories, (4) serve as a liaison to the broader plant biology community, (5) identify needs and opportunities of the *Arabidopsis* research community and communicate them to funding agencies of participating nations, and (6) periodically up-date this long-range plan.

APPENDIX

ARABIDOPSIS THALIANA AS A MODEL SPECIES

Arabidopsis thaliana owes its attractiveness to its small genome size, low level of repetitive DNA, short life cycle, and ease of manipulation in the laboratory, including the ability to be transformed. Added to this, *Arabidopsis* is a true flowering plant with the growth, developmental, nutritional, and light responses characteristic of higher plants. As background, a brief summary of information currently available on *Arabidopsis* is described below.

1) GENOME: The current (April, 1990) estimate of haploid nuclear genome size is 100,000 kilobase pairs, similar to the haploid nuclear genome size of *Caenorhabditis elegans*, and about seven times larger than that of the yeast *Saccharomyces cerevisiae*. This is many-fold smaller than the haploid nuclear genome size of any other plant used in research or in agriculture.

2) GENETIC MAPS: The classical genetic map of *Arabidopsis* contains about 90 mapped genes on five chromosome pairs, whose mutant phenotypes include developmental abnormalities (including homeotic mutants), enzymatic activity nulls, hormone non-producing and non-responding plants, and plants with inappropriate responses to the environment (including nutritional responses and responses to light). The classical genetic map is roughly integrated with two independent RFLP maps, which between them show the map locations of over 200 DNA fragments. The two RFLP maps can be mathematically integrated (they share 20 common mapped fragments). When the integration to provide a single 200-marker map is done, the average spacing between mapped locations will be around 350 kb over the entire genome.

3) GENE LIBRARIES: Many cDNA and genomic libraries exist and are broadly available, including genomic libraries in lambda, cosmid, and yeast artificial chromosome vectors. The current average size of *Arabidopsis* inserts in yeast artificial chromosome libraries is 150 kb.

4) CLONED GENES: A number (at least 60) of different specific genes and sequences of known function have been cloned from *Arabidopsis*. They represent genes involved in many aspects of plant growth and development, including amino acid biosynthesis, herbicide resistance, flower development, membrane proton transport, nutrition, photomorphogenesis, photosynthesis, seed protein synthesis, etc. There are several methods for cloning, including cross-hybridization with genes or probes from other species (or even kingdoms), differential cDNA screening, chromosome walking, and insertional mutagenesis (T-DNA tagging). An *Arabidopsis* telomere sequence is also known, leading to the possibility that (with an appropriate centromere) artificial chromosomes could be made.

5) MUTAGENESIS: One additional aspect of *Arabidopsis* of relevance is the ease with which large numbers of seeds can be collected and mutagenized, resulting in large populations of mutant plants for mutant screening and selection. The small size of the plant (between 10 to 40 cm tall at maturity requiring less than 1 cm² for growth of a plant), and its short generation time (five or six weeks) make possible large screens. Such screens have in the past resulted in many useful mutant strains.