

**The Multinational Coordinated
Arabidopsis thaliana Genome
Research Project**

Progress Report: Year Six

Multinational Coordinated *Arabidopsis thaliana* Genome Research Project

Progress Report:
Year Six

May, 1997

The Multinational Science Steering Committee:

Committee Chair: **David Meinke**, Oklahoma State University
Michael Bevan, John Innes Centre, Norwich, United Kingdom
Michel Caboche, Lab. Biol. Cellulaire, INRA, Versailles, France
Joseph Ecker, University of Pennsylvania, Philadelphia
Richard Flavell, John Innes Centre, Norwich, United Kingdom
Gerd Jurgens, University of Tübingen, Tübingen, Germany
Robert Last, Cornell University, Ithaca, New York
Jose Martinez Zapater, CIT-INIA, Madrid, Spain
Kiyotaka Okada, Kyoto University, Kyoto, Japan
David Smyth, Monash University, Clayton, Australia
Marc Van Montagu, University of Ghent, Belgium

The drawing on the opposite page is by Francisco Vergara-Silva, a commercial artist who is also a graduate student with Professor Elena Alvarez-Buylla at the Universidad Nacional Autonoma de Mexico and a visiting student in the laboratory of Dr. Elliot Meyerowitz at the California Institute of Technology. Dr. Meyerowitz and his coworkers, with partial support from the National Science Foundation, developed the ABC model for flower development, based on analysis of genes first isolated in *Arabidopsis*. Mr. Vergara-Silva's research is focused on the molecular mechanisms of flower evolution. Inspired by the central role of *Arabidopsis* as a model for understanding the biology and genetics of all plants, including the major crop species, Mr. Vergara-Silva has created an artistic rendering of *Arabidopsis* floral anatomy. *Arabidopsis* has not only served to enhance our understanding of and ability to engineer crop plants, it is also serving to create a better understanding of the biological relationship between plants and animals, all expressed in the design created by Mr. Vergara-Silva.

The Foundation provides awards for research and education in the sciences and engineering. The awardee is wholly responsible for the conduct of such research and preparation of the results for publication. The Foundation, therefore, does not assume responsibility for the research findings or their interpretation.

The Foundation welcomes proposals from all qualified scientists and engineers and strongly encourages women, minorities, and persons with disabilities to compete fully in any of the research and education related programs described here. In accordance with federal statutes, regulations, and NSF policies, no person on grounds of race, color, age, sex, national origin, or disability shall be excluded from participation in, be denied the benefits of, or be subject to discrimination under any program or activity receiving financial assistance from the National Science Foundation

Facilitation Awards for Scientists and Engineers with Disabilities (FASSED) provide funding for special assistance or equipment to enable persons with disabilities (investigators and other staff, including student research assistants) to work on NSF projects. See the program announcement or contact the program coordinator at (703) 306-1636.

The National Science Foundation has TDD (Telephonic Device for the Deaf) capability, which enables individuals with hearing impairment to communicate with the Foundation about NSF programs, employment, or general information. To access NSF TDD dial (703) 306-0090; for FIRS, 1-800-877-8339.

Table of Contents

Preface 5
Vision Statement 7
A New Long-Range Plan 7
A Summary of Progress 9
Recommendations 11
National and Transnational Projects 14
Memorandum of Understanding: 19

Preface

The “Multinational Coordinated *Arabidopsis thaliana* Genome Research Project” was established in 1990 to promote international cooperation in basic and applied research with *Arabidopsis*, a model plant species amenable to experimental manipulation in the laboratory. The primary objective of this project has been to understand the molecular basis of plant growth and development and to address fundamental questions in plant physiology, biochemistry, cell biology, and pathology. Initial plans were outlined in a publication (NSF #90-80) drafted six years ago by an ad hoc committee of nine scientists from the United States, Europe, Japan, and Australia. In recent years, this project has become a model for widespread participation and effective coordination of multinational research efforts in modern biology.

Arabidopsis thaliana, a small plant in the mustard family, was chosen for this large-scale research effort because it offers many advantages for detailed genetic and molecular studies. Among these features are its small size, short life cycle, small genome, ability to be transformed, availability of numerous mutations, and prolific seed production. By concentrating research efforts on a single model organism, detailed information on specific genes and cellular processes can be readily obtained and rapidly applied to a wide range of plants relevant to agriculture, health, energy, manufacturing, and the environment.

Each year since 1990, the scientific steering committee for the *Arabidopsis* Genome Project has prepared a progress report summarizing recent advances in *Arabidopsis* research. This is the sixth annual progress report published by the steering committee in conjunction with the U.S. National Science Foundation. Two years ago the report was a color brochure designed to explain the value and significance of *Arabidopsis* research to a wide audience. Last year the report presented a detailed overview of recent advances in research with *Arabidopsis*, along with technical information for use by members of the *Arabidopsis* community.

This year the report is designed to present an updated vision statement for the future that extends beyond the current focus of individual research laboratories and genome centers. Specifically, this report is designed to stimulate further advances in the use of *Arabidopsis* as a model system for the analysis of complex organisms. Multinational cooperation and communication continue to be an important feature of the *Arabidopsis* genome project. A brief overview of *Arabidopsis* research efforts in a number of participating countries is therefore included in this report. Additional information can be obtained through recent publications, electronic news groups and databases, and biological resource centers devoted to *Arabidopsis* research. As with any document that attempts to summarize the contributions of many individuals, this report may fail to include or misrepresent some significant achievements. The steering committee hopes that members of the *Arabidopsis* community will overlook such shortcomings and will communicate any concerns to committee members so that future reports will be as accurate as possible. We thank all members of the *Arabidopsis* community for their many contributions to the success of the initial phase of the Multinational Coordinated *Arabidopsis thaliana* Genome Research Project.

The Multinational Coordinated *Arabidopsis thaliana* Genome Research Project

Vision Statement

Remarkable progress has been made over the past six years towards meeting the goals outlined in “*A Long-Range Plan for the Multinational Coordinated Arabidopsis thaliana Genome Research Project*”. During this period of time, *Arabidopsis* research has grown exponentially and has impacted every major discipline of modern plant biology. As a result, the practice of using a model genetic system for basic and applied research has become widely accepted by plant biologists worldwide. This represents a fundamental change in the nature of plant research. In addition, recent advances in *Arabidopsis* research have further demonstrated to the scientific community that plants are useful models for addressing a wide range of fundamental biological questions.

Major advances in *Arabidopsis* research have extended from gene identification and molecular genomics to informatics and the establishment of essential community resources. Highlights of this coordinated research effort include:

Multinational framework established for the first complete analysis of a plant genome.

Supporting infrastructure of stock centers, databases, internet resources, and advanced technologies established.

Large-scale genomic sequencing program initiated through establishment of the *Arabidopsis* Genome Initiative.

Information from EST and genome sequencing efforts widely exploited by plant biologists. Impressive number of genes identified through mutagenesis and molecular cloning.

Significant advances reported in basic understanding of plant genetics, physiology, development, biochemistry, and plant response to pathogens and environmental signals.

The *Arabidopsis* community is poised to elucidate for the first time the structure of an entire plant genome, the sequence of every gene, and the functions of key regulatory elements involved in fundamental processes of development, metabolism, cell biology, and response to environmental signals. Our vision for the future is to build upon these dramatic accomplishments and promote the use *Arabidopsis* as a model system to understand not only plant structure and function but also more universal questions related to the nature and origin of biological complexity.

A New Long-Range Plan

We present here a plan for the future of the *Arabidopsis* Genome Research Project that should provide a model for other genome projects and stimulate additional discussion and scientific advances throughout the scientific community. This is clearly a multinational effort with many countries participating in *Arabidopsis* research and several regions of the world involved in sequencing the genome. With continued advances in *Arabidopsis* research, beyond the completion of genomic sequences and protein databases, we envision for the first time a path to understanding the fundamental principles of biological function and organization, which in plants may be quite different from animals. The challenge for the future is to provide continued documentation for the growing realization among biologists across many disciplines that plants have a tremendous potential for providing valuable information on fundamental principles of biological organization. With these objectives in mind, we have established the following goals for *Arabidopsis* research:

5 Year Goals:

Determine the nucleotide sequence of the entire nuclear genome of *Arabidopsis*.

This target date is two years ahead of the current timetable but no technical barriers exist to completing the sequence within five years. Early comple-

tion of this phase of the project will have a catalytic effect on research advances in many disciplines of biology.

Complete screens for most classes of informative mutations.

Analysis of existing mutations has already resulted in extraordinary advances in plant biology. Screens for additional mutations need to approach saturation within the next 5-10 years to complement rapid advances in molecular genomics.

Obtain insertional knockouts of every major class of genes.

This approach is needed to address the biological significance of genes identified through large-scale sequencing efforts. Significant progress has already been reported in the creation of insertional mutants using T-DNA sequences and transposable elements.

Continue the detailed characterization of individual metabolic, cellular, physiological, and developmental pathways.

These studies promise to provide the first complete picture of a flowering plant, from the molecular to organismal levels, to complement recent advances with other model eukaryotes.

Continue the widespread use of *Arabidopsis* as a model system to study basic principles of modern genetics.

Research with *Arabidopsis* holds great promise in understanding complex networks and levels of gene regulation, including epigenetics, and how genetic programming is orchestrated.

Establish improved computing systems that organize information on cellular processes involved in plant growth and development.

The challenge of managing information will continue to expand with further advances in *Arabidopsis* research.

Make information and advances obtained through the *Arabidopsis* Genome Project available to those working on other genome projects.

Emphasis must be placed on disseminating information and materials obtained through this genome

project to the broad community of biologists worldwide.

Maintain essential community resources and multinational coordination.

The *Arabidopsis* Genome Project should continue to be a model for multinational cooperation with respect to genomic sequencing efforts, informatics, and stock center maintenance.

10 Year Goals:

Determine the functions and locations of key gene products identified as a result of this genome project.

Emphasis will shift from gene identification to protein function and from isolated gene sequences to a comprehensive view of unique and redundant gene products.

Uncover the mechanisms by which complex networks of gene products become established and localized.

Regulatory networks must be characterized within a cellular context to understand plant growth and development and to compare results obtained for *Arabidopsis* with those reported for other eukaryotes.

Combine information on essential gene products with advances in plant physiology and biochemistry to establish a comprehensive picture of plant structure and function.

The function of signal molecules, metabolites, and specialized compounds throughout the plant life cycle should be established within the framework provided by the genome project.

Use *Arabidopsis* to resolve questions concerning evolutionary relationships among eukaryotic organisms and the evolution of common cellular and developmental pathways.

Sequencing the *Arabidopsis* genome will provide a unique opportunity to examine at a molecular level the origin and diversity of evolutionary pathways among eukaryotic organisms.

Maintain essential community resources and multinational cooperation.

Long-Term Goal:

Combine the reductionist view of cell and genome organization and function in *Arabidopsis* with the global view of plant development and evolution to arrive at an understanding of how a complex multicellular organism works and how this strategy compares with that employed in other major groups of organisms.

Initially we see the potential to understand in great detail the cellular basis of plant growth and development and resolve long-standing questions in plant physiology and biochemistry such as:

Mechanisms of hormone action;

Details of light perception and signal transduction pathways;

Regulation of cell division and differentiation;

Specialization of tissues, cells, and organelles;

Origin of specialized structures during plant evolution; and

Complete integration of cell and molecular biology with classical botany and agriculture.

The *Arabidopsis* Genome Project must continue to serve as a model for the application of knowledge gained from the detailed analysis of a single organism to important questions being addressed in many different groups of related organisms.

A Summary of Progress

Initial Project Goals:

The central mission of the *Arabidopsis* Genome Project, as outlined in the initial report, was to identify every gene in a model plant species and to determine the complete nucleotide sequence of the *Arabidopsis* genome by the end of the century. The ultimate goal is to understand the physiology, biochemistry, growth and development of a flowering plant at the molecular level. Six program objectives were identified by the international community of scientists who drafted the initial report.

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 was highlighted as being crucial for implementing the proposed research. Specific goals in these six areas were listed with target dates of 1, 2, 5, and 10 years.

Genome Analysis:

Significant advances have been reported in the isolation and characterization of informative mutations, cloning and sequencing of essential genes, and detailed analysis of gene products. More than 500 mapped genetic loci have been identified and analyzed in detail. YAC and BAC libraries with large inserts have been constructed, arranged in minimal contigs, and widely used for molecular genomics. Large numbers of valuable genes and cDNAs have been sequenced and deposited in GenBank. These advances have provided a wealth of information for the plant biology community. EST sequencing projects have also played an important role in the development of microchip technologies for rapid evaluation of entire genomes. Large-scale genomic sequencing has commenced with the establishment of the *Arabidopsis* Genome Initiative, a collection of respected genome centers worldwide that are coordinating research efforts to produce at least 20 Mb of finished sequence within 2 years and complete the remainder of the genome by the year 2004. At the current rate of approximately 200 sequenced genes every month, a tremendous amount of information is being made available without delay to the scientific community.

Technology Development:

Major breakthroughs include the development of high-efficiency transformation and insertional mutagenesis systems, the establishment of improved methods for pursuing forward and reverse genetics in *Arabidopsis*, the production of novel constructs for characterizing patterns of gene expression, and the applica-

tion of technical advances in animal and microbial genome projects to plant biology.

Biological Resource Centers:

The Nottingham *Arabidopsis* Stock Centre (NASC) at the University of Nottingham (UK) and the *Arabidopsis* Biological Resource Center (ABRC) at Ohio State University (USA) were established several years ago to maintain a wide range of biological materials (seed stocks, DNA clones, libraries, and related information) required for completion of the genome project. These stock centers have done an outstanding job of working cooperatively and efficiently to serve the diverse needs of the *Arabidopsis* community.

Informatics and Communication:

A coordinated network of databases and internet resources has been established to support advances in *Arabidopsis* research and facilitate rapid communication of technical information.

- The *Arabidopsis thaliana* Database (AtDB) serves as the primary source of information for research on *Arabidopsis* and provides numerous links to related WWW resources worldwide. This database was originally developed at Massachusetts General Hospital and is currently being upgraded and maintained at Stanford University.
- The *Arabidopsis* Information Management System (AIMS) was established in association with the ABRC at Ohio State to facilitate ordering of stocks and accessing information related to *Arabidopsis* research.
- The BIOSCI *Arabidopsis* Genome Electronic Conference was established as part of the BIOSCI/BIONET network and has emerged as a model for widespread information exchange worldwide.
- Internet resources devoted to *Arabidopsis* research can be readily accessed through:

AtDB: <http://genome-www.stanford.edu/Arabidopsis/>

ABRC: <http://aims.cps.msu.edu/aims>

NASC: <http://nasc.nott.ac.uk>

Meetings and Workshops:

The International Conference on *Arabidopsis* Research has become a popular forum for the exchange of information on *Arabidopsis* research. Each year this meeting attracts over 600 participants from more than 15 different countries to a conference site that alternates between North America and Europe. The Cold Spring Harbor Course on *Arabidopsis* and numerous transnational exchange programs and workshops have trained a new generation of plant biologists familiar with the genetics and molecular biology of *Arabidopsis*.

Advances During Previous Year:

Establishment of the *Arabidopsis* Genome Initiative was a highlight of the previous year. Representatives from each of the major centers participating in large-scale sequencing of the *Arabidopsis* genome met in August, 1996 at the National Science Foundation in Washington, D.C. to agree on the most rapid and efficient approach to completing this critical phase of the genome project. The document signed by participants at this meeting is presented in Appendix II, the Memorandum of Understanding. An overview of this document and the objectives of the *Arabidopsis* Genome Initiative was recently published (Bevan et al., *Plant Cell* 9: 476-478; 1997). Current participants in this initiative include a consortium of 18 laboratories in the European Union, three independent but complementary groups of sequencing centers in the United States, and major sequencing centers in France and Japan. These centers have already generated a considerable amount of sequence information that is being made available to the scientific community through internet linkages.

The Seventh International Conference on *Arabidopsis* Research, held at the University of Anglia, United Kingdom, in June 1996, documented once again the dramatic scientific advances that are being made with this model organism. A detailed review of this meeting was recently published (Somerville and Somerville, *Plant Cell* 8: 1917-1933; 1996). Readers are encouraged to consult this review to appreciate the impressive diversity of scientific advances described at the meeting. Topics of discussion included disease resistance, growth regulators, embryogenesis, vegetative development, flowering, environmental responses, genomics, epigenetics, molecular biology, new technologies, informatics, and common resources. An updated

summary of community standards for nomenclature, mapping, and genetic analysis in *Arabidopsis*, which are required to maintain consistency in the genome project, is scheduled for publication in *The Plant Journal* (August, 1997) and will be distributed to participants at the Eighth International Conference on *Arabidopsis* Research, to be held in Madison, Wisconsin in June 1997.

Continued funding of essential community resources and collaborative projects was another highlight of the previous year. Each of the participants in the *Arabidopsis* Genome Initiative received significant funding from national sources to support large-scale sequencing of the *Arabidopsis* genome. An interagency (NSF/USDA/DOE) grant is providing \$12.7 million over 3 years to support genomic sequencing efforts in the United States. The European Union is providing \$7.5 million over 2 years to support a consortium of sequencing laboratories in Europe; the Japanese Government is providing \$4.5 million per year to support *Arabidopsis* sequencing efforts at the Kazusa Institute; and additional funding for *Arabidopsis* research was included with the establishment of a new sequencing center in France. Continued funding was also announced for the *Arabidopsis* Biological Resource Center (USA) and the Nottingham *Arabidopsis* Stock Center (UK).

Practical Applications:

Considerable interest has been generated in recent years over practical applications of basic research with *Arabidopsis*. The large number of companies using *Arabidopsis* in their research programs clearly illustrates the anticipated benefits of working with a model plant system. Examples of practical applications include the discovery of hormone and signal receptors and transduction pathways and their potential modification to address agricultural and environmental problems; the genetic dissection of disease resistance and stress response pathways related to crop productivity; the identification of genes involved in determining plant architecture and progression through critical developmental pathways; and the detailed analysis of biochemical pathways that may be modified to produce desired products for manufacturing and human health. Particular attention has been paid in recent years to advances made with *Arabidopsis* in modifying levels of polyunsaturation in seed oils to benefit human nutrition, producing biodegradable plastics such as

polyhydroxybutyrate in crop plants, and reducing the time required for trees to flower through introduction of *Arabidopsis* genes involved in regulating development of the shoot meristem. These advances have been covered widely in the press and have served to inform the general public of pending advances in plant biotechnology. Studies on the genetic basis of variability in *Arabidopsis* have begun to bridge the gap between the molecular genetics of model organisms and the breeding of important crops. *Arabidopsis* has also been employed in the classroom to demonstrate classical principles of genetics, ecology, and plant physiology to a new generation of students.

Recommendations

A vision for the future of the *Arabidopsis* Genome Project has already been presented in the introductory sections of this document. We present here additional recommendations, arranged under the same headings used in previous reports, for the widespread use of *Arabidopsis* as a model system to address basic questions of biological complexity and organization.

Genome Analysis:

Rapid completion of the genome sequencing initiative and widespread dissemination of the resulting data in an appropriate format remain top priorities for the *Arabidopsis* community. There are no apparent technical barriers to completing this project within 5 years. The question is primarily one of funding. Although we recognize the need to support continued analysis of different organisms in order to provide an accurate vision of biological diversity, there is no doubt that sequencing the entire *Arabidopsis* genome will result in a dramatic and fundamental change in the nature of plant research. Plant biologists have already begun to make extensive use of data generated during the initial phase of the *Arabidopsis* Genome Initiative. Completion of this project will further demonstrate the value of concentrating limited resources on a single model system and will significantly reduce the immediate need to complete large-scale genome projects on related angiosperms because many common genes will already be identified. There will also be an increasing number of cases where genes identified in *Arabidopsis* provide insights into the origin and function of related genes present in animal systems, including humans. Examples of this trend have already begun to emerge from laboratories

studying a number of essential cellular functions in *Arabidopsis*.

Sequencing the *Arabidopsis* genome will provide a unique opportunity to address fundamental questions of genome organization and evolution. There should be immediate practical applications to some of this work, as in the rapid identification of large numbers of genes with agricultural importance in related *Brassica* species. The completed database of 20,000 sequenced *Arabidopsis* genes should also greatly simplify efforts to identify related genes in more divergent angiosperms and to focus on those genes that are responsible for major differences in cellular and developmental pathways characteristic of distinct groups of flowering plants. Comparative studies with sequenced microbial and invertebrate genomes will also become possible and should result in a more complete picture of the function and evolution of eukaryotic organisms.

As the sequencing initiative approaches completion, the emphasis of individual research programs will continue to shift from the elucidation of gene structure to the analysis of gene function and localization of gene products. Knockout mutations and reverse genetics will likely play an important role in determining the biological significance of sequenced genes. Interpreting phenotypes of these knockouts will require continued progress in the traditional isolation and characterization of many different types of mutations. Some of these screens will need to approach saturation over the next 5 years to complement rapid advances in molecular genomics. Emphasis must also continue to be placed on following established guidelines for mutant analysis and gene nomenclature in order to maintain accuracy in expanding databases.

Continued dissection of individual metabolic, cellular, physiological, and developmental pathways will be required to place gene products identified through the sequencing project within the context of a living organism, a dramatic change made possible through large-scale sequencing efforts genes. This will enable plant biologists to establish a global view of cell organization, complexity, and response to perturbation that should become a model for other organisms, in part because plants will be that responses to isolated signals, mutations, and experimental treatments may soon be monitored at the level of the entire genome and not simply in relation to a select number of markers because

plants are particularly responsive to many signals that impact all organisms. Further advances in cell and molecular biology should make it possible in the long term to determine in great detail the location and function of a wide range of signal molecules, metabolites, macromolecules, and specialized compounds within differentiated plant cells and throughout the plant life cycle.

Technology Development:

One goal of the *Arabidopsis* Genome Initiative is to utilize, when appropriate, emerging technologies for rapid analysis of genomic sequences. These advances may reduce the time and cost required to complete the genome project but should be weighed against the need to maintain the quality of data being generated. Other technologies that utilize microchips and related materials to analyze genome organization and expression on a large scale have the potential to make a dramatic impact on research in plant biology and find numerous applications in *Arabidopsis* research. With these methods it should be possible to monitor the responses of large numbers of genes to experimental and genetic modifications. Development of improved methods for creating large numbers of insertional mutants and knockouts of cloned genes will be needed to determine the biological significance of sequences identified during the course of this project. Methods that allow rapid and sensitive detection of gene products in living cells will become increasingly important as advances in molecular genetics and cell biology are combined to form an integrated picture of plant structure and function. The application of *Arabidopsis* research to problems in agriculture, health, manufacturing, energy, and the environment holds tremendous potential but may require the establishment of improved methods for widespread production and utilization of specialized chemical and biological reagents. Participation of private industry in technology development should be encouraged and will help to demonstrate the long-term practical benefits of intensive research on *Arabidopsis*. Emphasis must nevertheless be placed on making certain that this participation does not interfere with the open and collaborative spirit that characterizes the *Arabidopsis* community.

Biological Resource Centers:

Maintenance of resource centers that service the community by providing biological materials ranging from seed stocks to isolated clones and molecular mark-

ers remains a top priority for both research and teaching institutions worldwide. The present arrangement of complementary stock centers located in the United States and Europe has served the community well and should be maintained. The addition of modest service charges at both centers represents a reasonable attempt to limit operating costs and ensure long-term support from a combination of sources. Completion of the *Arabidopsis* Genome Initiative will create additional challenges with respect to management and distribution of materials. Effective communication between stock center managers, database curators, advisory committees, and the scientific community will be required to maximize long-term benefits of this genome project.

Informatics and Communication:

The present level of electronic integration and transnational communication of information related to *Arabidopsis* research far exceeds that outlined in the original project. The electronic *Arabidopsis* news group established 10 years ago has become a model for scientific communication and plays an important role in fostering multinational cooperation and rapid exchange of technical information. The rapid growth of internet resources devoted to *Arabidopsis* research has made it possible to analyze complex data with minimal effort. The *Arabidopsis thaliana* database (AtDB) has become a central resource for accessing information related to *Arabidopsis* research. Continued funding of this critical resource is needed to coordinate information management related to the genome project and related efforts in hundreds of laboratories worldwide. Improved methods for monitoring and updating information may be required as the amount and complexity of technical data continue to expand.

Meetings and Workshops:

The annual *Arabidopsis* meeting will continue to play an important role in highlighting recent advances

in *Arabidopsis* research. The current arrangement of alternating meeting sites between the United States and Europe has served the community well and should be continued. Other sites in Asia and North America may need to be selected on an occasional basis to provide opportunities for individuals from different regions to attend at minimal cost. Combining this meeting with others in related disciplines is also to be encouraged, as for example the 1998 *Arabidopsis* Conference in Madison to be held in conjunction with the Annual Meeting of the American Society of Plant Physiologists. Regional courses and workshops such as those offered at Cold Spring Harbor and many other sites worldwide should continue because they serve a valuable training function for scientists entering the *Arabidopsis* field. Additional meetings and workshops that combine scientists working on different model systems may be needed in the future to stimulate more comparative studies on biological organization and complexity in model organisms.

Multinational Cooperation:

This genome project is indeed a multinational effort with over 35 countries participating in *Arabidopsis* research and several regions of the world involved in sequencing the genome. Further evidence of the multinational nature of this research community is provided in Appendix I, which is devoted to a summary of national and transnational projects. Rapid development of internet resources devoted to *Arabidopsis* research has allowed even more extensive exchange of data and strengthening of multinational collaborations. We encourage continued advances in the multinational nature of this project so that it may continue to serve as a model for other genome projects and facilitate the solution to regional problems related to science and society.

Appendix I

National and Transnational Projects

Arabidopsis research continues to blossom throughout the world. Recent stock center transactions included requests from Argentina, Australia, Belgium, Brazil, Canada, Chile, China, Colombia, Denmark, Finland, France, Germany, Hong Kong, Hungary, Iceland, Israel, Italy, Japan, Korea, Luxembourg, Malaysia, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, Russia, Singapore, South Africa, Spain, Sweden, Taiwan, Turkey, Ukraine, the United Kingdom, and the United States. International coordination is an important component of the *Arabidopsis* Genome Project. Examples are given here of recent advances in regional centers of *Arabidopsis* research.

AUSTRALIA

Research programs continue into fundamental aspects of plant biology, including flowering (morphogenesis and fertilization), cell biology (regulation of cell shape, cellulose biosynthesis, cell division and photosynthesis) and interactions with environmental factors (heavy metals, phosphate, UV light, and pathogens). A program at the CSIRO Division of Plant Industry, funded by the Department of Industry, Science and Technology, has generated a 2 Mb YAC contig around MS1 on chromosome 5 (Chapple et al., Aust. J. Plant Physiol. 23: 453-465, 1996). The Australian Government has taken a novel approach to funding genome studies (see Science 275, 25-26, 1997). An Australian Genome Research Facility has been established for DNA sequencing, gene mapping, and mutation detection. Funding (\$7.9 million USD) has been provided for robots and automated DNA sequencing equipment. Salaries and other costs will be met from user fees. DNA sequencing will be performed at the University of Queensland in Brisbane. Mapping and mutation detection with novel chemical methods will be performed at the Walter and Eliza Hall Institute for Medical Research in Melbourne. Major users of this facility will likely have medical and agricultural interests. The Australian National Genomic Information Service at the University of Sydney pro-

vides rapid access to all sequence databases via a dedicated WWW site.

Contact Person: David Smyth, Monash University
E-mail Address: david.smyth@sci.monash.edu.au

BELGIUM

This has been an important time for biotechnology in Belgium and specifically in Flanders. Last year the government decided to establish the "Flanders Interuniversity Institute for Biotechnology". Core activities of this institute are formed by four well-known research groups. One of these is the Lab of Plant Genetics in Ghent, directed by Marc Van Montagu. The Flemish government also supports several other projects that aim at further development of *Arabidopsis* as a model plant. In addition, the Belgian government supports collaborations among Belgian universities. For example, the laboratory at Ghent interacts with those in Antwerp, Brussels and Liège in a national program on plant growth and development. This program includes the study of leaf morphogenesis, flowering, cell division, and hormone regulation. A very promising result of last year's program was the establishment of a protein database of *Arabidopsis* using two-dimensional electrophoresis and partial amino acid sequence analysis. This is also the topic of new European project for which the lab in Ghent is the informatics coordinator (<http://sphinx.rug.ac.be:8080/>).

Contact Person: Marc Van Montagu, University of Ghent
E-mail Address: aruyt@gengenp.rug.ac.be

CANADA

Significant progress continues to be reported in selected laboratories located throughout the country using *Arabidopsis* as a model system. Projects focus on a wide range of topics such as seed maturation, triacylglyceride biosynthesis, phosphate transport, herbicide resistance, flower development, plant pathogenesis, meiosis, and vegetative development. Several labo-

ratories and companies with interests in *Brassica* species have also benefited from research with *Arabidopsis*.

Contact Person: Bertrand Lemieux, York
University
E-mail Address: fs300500@sol.yorku.ca

CHINA

Approximately ten laboratories in mainland China are now working with *Arabidopsis* in fields that include signal transduction pathways involving the plant hormones abscisic acid (ABA) and indoleacetic acid (IAA), floral development, vegetative development and the analysis of multiple rosettes and abnormal leaf shapes, biosynthetic pathways of tryptophan/IAA, and response to stresses such as salt and cold treatment. Genetic approaches are being taken in most of these projects, which are mainly supported by the National Natural Science Foundation of China. Various mutants have been isolated from mutagenized seeds generated with chemical mutagens or transposon insertion and some have been characterized and mapped. Mutants isolated from transposon tagged lines may help to clone the genes involved. Great effort is also being made to produce thousands of lines with insertion of the Ds transposon.

Contact Person: Jiayang Li, Chinese Academy of
Sciences
E-mail Address: jyli@ss10.igtp.ac.cn

EUROPEAN UNION

A network of 18 EU sequencing labs recently completed the first stage of its *Arabidopsis* genome effort with a large-scale pilot project involving 2 Mb on chromosome 4 and three small regions on chromosomes 1 and 3. A suite of computer programs was used to identify genes and characterize other features throughout this genomic region. Approximately 360 genes were identified based on sequence analysis within the 1.9 Mb FCA contig. The density of one gene every 5 Kb is consistent with current estimates of 20,000 total genes. Detailed analysis of this genomic region has provided the first glimpse of the structure and function of a plant genome. Comparisons with twenty genes previously sequenced within this region revealed some problems with missed exons and inaccurate splice sites that need to be addressed in future projects. The annotated FCA sequence will be submitted to EMBL in June, 1997. A protein structure database and graphical display of pre-

dicted genes will also be available from the MIPS Web site. The next stage of sequencing in the EU, aimed at sequencing over 5 Mb of chromosome 4, should be completed within a year. Plans are being made for additional sequencing efforts, contingent upon additional funding, in conjunction with the *Arabidopsis* Genome Initiative. Progress continues in other collaborative projects, described in previous reports, dealing with insertional mutagenesis and the isolation of genes of agronomic importance from *Arabidopsis*.

Contact Person: Michael Bevan, John Innes
Centre
E-mail Address: bevan@bbsrc.ac.uk

FRANCE

Considerable progress was reported in large-scale collaborative projects involving genome analysis and insertional mutagenesis. The GREG program supported advances in EST mapping and construction of YAC contigs. The ESSA program generated 300 Kb of genomic sequence in the laboratories of M. Delseny (Perpignan), M. Kreis (Orsay), and R. Mache (Grenoble). Additional sequence will be contributed over the next two years. A national sequencing Program (TGS) headed by J. Weissenbach was initiated with a goal of 20 Mb per year. Part of this effort will be devoted to the *Arabidopsis* Genome Initiative. Over 28,000 T-DNA lines have been generated at INRA (Versailles), of which 11,000 have been distributed to the Nottingham stock center. A coordinated effort (GDR) of CNRS and INRA has been funded to exploit this T-DNA collection for mutant identification, function searches, and reverse genetics. Several hundred mutants are currently being studied with defects in different aspects of cell biology, metabolism, development, and response to pathogens. Advances in genome programs have led to a dramatic increase in the number of projects utilizing *Arabidopsis* as a model system. More than 50 such projects currently are being headed by a senior scientist.

Contact Person: Michel Caboche, INRA,
Versailles
E-mail Address: caboche@versailles.inra.fr

GERMANY

Arabidopsis research is now well-established in Germany, both at universities and national institutes for

plant research. The mid-term of the 6-year funded national *Arabidopsis* research program was celebrated with a meeting that gathered about 60 participants, including speakers from the international *Arabidopsis* community, near Tubingen in September, 1996. Topics covered included genome research, gametophyte development, embryogenesis, seed maturation, postembryonic development, responses controlled by light and hormones, cell biology, and nutrient transport. The program has expanded in scope and considerable progress has been made towards the molecular analysis of genes identified by mutation. Part of this success can be attributed to the increasing exchange of information and techniques between member laboratories. To further this exchange, a satellite workshop was held for members of the research program.

Contact Person: Gerd Jurgens, University of Tubingen

E-mail Address: geju@fserv1.mpib-tuebingen.mpg.de

ITALY

Research with *Arabidopsis* in Italy has increased rapidly over the past year to include at least 15 different laboratories. Research topics include response to pathogens, protein kinases, HD-ZIP transcription factors, myb and polyamine genes, photosynthesis mutants, cold stress, nodulin genes, auxin and ethylene physiology, root development, and tropisms in microgravitational conditions (ESA space projects). Even though a national program to support *Arabidopsis* research has not yet begun, financial support is coming from different sources, e.g. the National Research Council, the Ministry of Agricultural Resources, and the European IV Frame programs. Research groups are located both in universities and in National Institutes (Consiglio Nazionale delle Ricerche, ENEA, Istituto Nazionale della Nutrizione). A group called ARABITALIA was started at the 7th International *Arabidopsis* Meeting in Norwich in June 1996. This year's meeting is scheduled for the end of September at Macerata (Italy), in conjunction with the 41st symposium of the Italian Society of Agricultural Genetics. At this time a booklet will be distributed with information about Italian research teams working on *Arabidopsis*. Additional projects and collaborative efforts are expected to be initiated on this occasion.

Contact Person: Fernando Migliaccio, IBEV Institute / Consiglio Nazionale delle Ricerche Monterotondo (Rome)

E-mail Address: miglia@nserv.icmat.mlib.cnr.it

JAPAN

Arabidopsis research is well-established in Japan. Nearly 40 laboratories in universities, national institutes, and private companies use *Arabidopsis* as a major experimental plant. During the last year, many laboratories used transgenic plants to examine the expression of isolated genes and search for tagged mutants. The scientific activities of *Arabidopsis* researchers in Japan were reported at several regional workshops and symposia: the annual meetings of the Botanical Society of Japan, The Genetics Society of Japan, the Japanese Society of Molecular Biology, and the Japanese Society of Plant Physiologists. The 7th Workshop on *Arabidopsis* Studies was held at Hokkaido University, Sapporo, August, 1996. This meeting was organized by Satoshi Naito, Kotoro Yamamoto, and Yoshibumi Komeda, and included about 120 participants and 33 oral presentations. Talks covered mutant analysis, transformation, transcriptional regulation, and gene cloning. The Japanese *Arabidopsis* communication network, nazuna-net, organized in January 1995, includes approximately 335 members (April, 1997) and is being widely used for information exchange (contact: Dr. Takayuki Kohchi: kouchi@bs.aist-nara.ac.jp). A large-scale genome sequencing project has been started at Kazusa DNA Research Institute in coordination with the *Arabidopsis* Genome Initiative (contact: Dr. Satoshi Tabata: tabata@not1.kazusa.or.jp). The Sendai Seed Stock Center (SASSC) has been operated since 1993 by Dr. Nobuharu Goto (n-goto@ipc.miyakyo-u.ac.jp).

Contact Person: Kiyotaka Okada, Kyoto University

E-mail Address: kiyo@ok-lab.bot.kyoto-u.ac.jp

NETHERLANDS

The number of *Arabidopsis* groups in the Netherlands has stabilized and includes established investigators in Wageningen (de Vries, Koornneef, Pereira), Utrecht (Smeekens, Scheres), and Leiden (Hooykaas). Other groups are using *Arabidopsis* on a smaller scale. Numerous collaborations involving the exchange of scientific materials and expertise are continuing as out-

lined in previous reports. The annual ARANED meeting takes place each winter and allows discussion of various organizational and scientific issues.

Contact Person: Maarten Koornneef, Wageningen Agricultural University

E-mail Address:
maarten.koornneef@botgen.el.wau.nl

NEW ZEALAND

Research with *Arabidopsis* in New Zealand is centered in the Auckland area at the university and at Hort+Research. The Foundation for Research Science and Technology (FRST) and the new Marsden Fund, dedicated to supporting innovative fundamental research, fund the work. Research projects are aimed at characterizing genes that influence reproductive development. *Arabidopsis* is also being used as the recipient of several foreign genes. These experiments aim to test the function of heterologous homeobox and MADS box genes in development, or the ability of selected genes to confer aluminium tolerance on plants. A T-DNA tagged population has been generated for mutant screening, and a project is underway to establish viral vectors for use in *Arabidopsis*.

Contact Person: Jo Putterill, University of Auckland

E-mail Address: j.putterill@auckland.ac.nz

REPUBLIC OF KOREA

Approximately 15 laboratories in Korea are currently involved in *Arabidopsis* research. Subjects under study include mutational and molecular analyses of leaf senescence, biochemical and mutational analysis of light signal transduction pathways, regulation of wound-inducible genes, role of drought-induced protein kinases, protoplast culture, genes induced by heavy metals, salts, and other environmental stimuli, anther development, and gene expression in guard cells. In addition, more than 20 laboratories are involved in *Brassica* research in Korea and some of their efforts are related to *Arabidopsis* research. These studies are funded primarily by individual grants from various government and private funding agencies.

Contact Person: Hong Gil Nam, POSTECH

E-mail Address: nam@vision.postech.ac.kr

SPAIN

The number of Spanish research laboratories adopting *Arabidopsis* as their experimental system keeps growing in Spain. The national agency (CICYT) supports an increasing number of research projects dealing with *Arabidopsis* in its Biotechnology Programme and in the Programme for the Promotion of Basic Knowledge. The European Union is another major source of funding for the Spanish groups. The Biotechnology Programme also funds the establishment of a Network of laboratories working with *Arabidopsis* and some of their activities. One such activity has been the production of 10,000 T-DNA transgenic lines that are currently available to Spanish groups and will be soon deposited in the International *Arabidopsis* Resource Centres. In a recent meeting of the *Arabidopsis* network held last January (1997) in Barcelona, networkers proposed new joint activities such as the exploitation of these lines in the search for T-DNA insertions in specific DNA sequences, the identification of natural variation in Spanish populations, and the generation of additional insertion mutagenized populations. The meeting also served to attract the interest of Spanish companies towards the benefits derived from *Arabidopsis* research.

Contact Person: Jose Martinez Zapater, CIT-INIA (Madrid)

E-mail Address: zapater@inia.es

UNITED KINGDOM

The UK *Arabidopsis* community benefited enormously last year by hosting the 7th International Conference on *Arabidopsis* Research in Norwich. During this time many new collaborations were established which helped consolidate and expand the *Arabidopsis* research base in the UK. A major funding program, Plant Molecular Biology II, funded by the BBSRC (Biotechnology and Biological Sciences Research Council of the UK) including 22 grants for work on *Arabidopsis*, comes to an end in June 1997. Grant holders organized a final meeting in February to report progress and discuss possibilities for future funding in relation to the current restructuring of the BBSRC. Many grant holders have secured funds from the European Commission or other BBSRC initiatives to continue their research. Currently, the BBSRC provides GBP 1.2 million plus 15 studentships for work on *Arabidopsis*. The Nottingham *Arabidopsis* Stock Centre (NASC) received

a further five years BBSRC funding from April 1997. This together with funds raised from user fees will secure the immediate future and enable NASC, in collaboration with ABRC, to meet the needs of the research program. The NASC continues to expand its information resources and now curates the Lister and Dean RI map as well as offering a mapping service to the international community. The centre is well used, with the most popular lines being the transgenic populations which can be used for mutant analysis, targeted tagging and reverse genetics. The John Innes Centre in Norwich hosted an eleven day practical EMBO (European Molecular Biology Organization) Course, “*Arabidopsis* as an Experimental Organism” and the Genetical Society of Great Britain will be hosting a prestigious one day meeting in London in November 1997: “*Arabidopsis thaliana*: big ideas from a small plant”.

Contact Person: Caroline Dean, John Innes Centre

E-mail Address: arabidopsis@bbsrc.ac.uk

UNITED STATES

Arabidopsis research remains firmly established in the United States, with active programs distributed

throughout a wide range of academic, government, and corporate laboratories. The diversity of this research effort should be evident at the upcoming International Conference on *Arabidopsis* Research to be held in Madison in June, 1997. Federal grant support for research with *Arabidopsis* remains strong with funding provided by a number of different agencies. One significant advance was the establishment of the *Arabidopsis* Genome Initiative and the awarding of \$12.7 million from NSF/DOE/USDA to support three respected sequencing centers within the United States to play a major role in completing the first sequence of a plant genome. Federal funds continued to support a wide range of research projects headed by individual investigators while at the same time maintaining the essential *Arabidopsis* Biological Resource Center (ABRC) at Ohio State University and the *Arabidopsis thaliana* database (AtDB) at Stanford University. The North American *Arabidopsis* Steering Committee continues to function in coordinating research efforts throughout the region, with Pam Green (email: 22313pjg@msu.edu) and Rob Last (rll3@cornell.edu) serving as co-chairs for the coming year.

Contact Person: David Meinke

E-mail Address: meinke@osuunx.ucc.okstate.edu

Appendix II

Memorandum of Understanding: Multinational Effort to Sequence the *Arabidopsis* Genome

On August 20-21, 1996, representatives of six research groups committed to sequencing the *Arabidopsis* genome met in Arlington, VA to discuss strategies for facilitating international cooperation in completing the genome project. All six groups have secured major funding to pursue large-scale genomic sequencing of *Arabidopsis*, and the EU and Japanese groups have been engaged in large-scale sequencing for some time. The primary objectives of this meeting were to establish *Arabidopsis* as a model for international coordination of sequencing efforts and to develop guidelines for rapid and efficient completion of the sequencing project by the year 2004. Representatives from Japan, France, the EU, and the USA were present at the meeting. A complete list of participants and observers is attached as Appendix I.

A remarkable degree of consensus was reached by the end of the meeting on the general strategy for the *Arabidopsis* sequencing project. All parties agreed to follow several practices that were seen as facilitating international cooperation. This document was drafted to serve as a modus operandi for the participating groups until such time as it is modified by mutual agreement of representatives of the participating groups. All signatories to this document have agreed to the following:

1. The *Arabidopsis* Genome Initiative (AGI) is intended to be an inclusive international collaboration. Any group that intends to engage in the sequencing of hundreds of kilobases of contiguous *Arabidopsis* genomic DNA will be invited to participate as a coequal collaborator in the AGI and will be expected to follow the guidelines outlined in this document.

2. A coordinating committee with representation from each of the participating groups was formed. This committee will be responsible for making all decisions

For current e-mail and World Wide Web addresses, please refer to Bevan, M., et al. 1997 Plant Cell 9:476-478

that affect the overall goals and operations of the AGI. In particular, it is anticipated that the AGI coordinating committee will be a planning and brokering system for establishing efficient ways of completing the genome. The committee will coordinate apportioning regions of the genome to the various groups in such a way as to minimize needless duplication of effort while maximizing progress toward complete sequencing of the genome. The committee will also be responsible for keeping the *Arabidopsis* community informed of continuing advances in the sequencing project.

Members of the committee for 1996-97 are Mike Bevan (Chair; EU consortium), Satoshi Tabata (Kazusa DNA Research Institute), Joe Ecker (Stanford-University of Pennsylvania-Plant Gene Expression Consortium {the SPP consortium}), Dick McCombie (Cold Spring Harbor-Washington University-Applied Biosystems Consortium {CSH-WU-ABI}), Steve Rounsley (The Institute for Genomic Research {TIGR}), Francis Quertier (French Genome Center) and David Meinke (Multinational *Arabidopsis* Steering Committee). Each member of the committee will be responsible for arranging a temporary or permanent replacement from the represented group when appropriate. New members will be invited to join the committee based on a nomination from one member of the committee and an affirmative vote by a majority. It is anticipated that the committee will maintain regular communication and will meet annually. Mike Cherry (Curator of AtDB) will develop an e-mail server to facilitate correspondence between members of the committee.

3. The six research groups are expected to complete different amounts of finished sequence because they have different capabilities and levels of funding devoted to this project. In order to prevent duplication of effort, it was considered useful to have the various groups initiate sequencing in different well-defined regions of the genome. It was agreed that each group should begin by nucleating sites over a contiguous region of a size that could be completed with the funding available. It was recognized that it may not be possible to define such a region with high accuracy because of variation in the ratio of genetic distance to physical distance. The goal

in this respect should be to avoid situations where one group obtains scattered regions of sequence that must eventually be finished (i.e., linked up) by other groups. Exceptions to this strategy are noted elsewhere in this document.

The SPP group will begin nucleating on chromosome 1. The EU group will nucleate the bottom arm of chromosome 4. The CSH-WU-ABI group will nucleate a 4 Mb region on the top arm of chromosome 4 and a 2 Mb region on the top arm of chromosome 5 (the latter in collaboration with the EU group and the Kazusa group). The TIGR group will nucleate chromosome 2. The Kazusa group will nucleate the lower part of chromosome 5. The region at the top of chromosome 5 of mutual interest to the EU, CSH-WU-ABI and Kazusa group will be sequenced collaboratively. The Kazusa group, which anticipates a monthly sequencing rate of approximately 500 Kb, expects to begin nucleating a region of chromosome 3 in 1997. The EU, TIGR, SPP and CSH-WU-ABI groups anticipate an average monthly rate of approximately 200, 220, 150 and 150 Kb per month, respectively. Thus, when all the groups are operating at full capacity, the average monthly rate for the entire AGI collaboration is expected to exceed 1.2 Mb per month. The philosophy of the AGI collaboration is that as the assigned regions near completion, the coordinating committee will designate new regions of unfinished sequence to the groups in proportion to their sequencing capabilities. For example, the French Genome Center is tentatively interested in sequencing BAC ends during the first year or two of operation but after that time it is anticipated that they will engage in sequencing a contiguous region of genomic DNA that will be decided at a later date.

Several of the participants had differing views about the relative merits of sequencing unique sequences versus regions of repetitive sequence such as centromeres and telomeres. On the one hand, it may be expected that the maximum number of coding sequences will be found by sequencing the regions of low copy number. On the other hand, it will be interesting to know the structure of the centromeric and telomeric regions. The majority view appeared to be that it was not necessary at this time to resolve this issue. However, the majority view was that renewals of existing grants should take into account the fact that some regions of sequence will be more difficult to complete than others and large stretches of contiguous sequence are more difficult to achieve

small scattered regions. Sequencing efficiency should be the sole criterion for choosing which clone to sequence. It was agreed by all parties that none of the groups should perform service sequencing for outside groups interested in particular clones. The reason for this is that the sequencing groups should not be seen to be favoring certain colleagues.

4. The most efficient strategy for sequencing the *Arabidopsis* genome is to shotgun sequence large clones such as BACS, YACS or inserts from P1 clones. Most of the groups have had preliminary experience with BACS and YACS and preferred BACS. The fact that most of the groups are currently satisfied with the available public BAC libraries will facilitate coordination and exchange of information. In particular, in order to minimize the requirement for additional physical mapping, it is desirable to obtain several hundred base pairs from the ends of a large number of BAC clones so that the minimum tiling path from a region of sequence to an overlapping clone can be determined by database analysis. The groups led by Craig Venter (TIGR) and Francis Quertier (French Genome Center) agreed to sequence the ends of approximately 14,000 BACS from public BAC libraries during the next two years and to make the information freely available to the community.

All of the groups will use public BAC, YAC or P1 libraries constructed from the Columbia ecotype that will be freely available to the world community. A suitable BAC library to begin with is the TAMU BAC library constructed by Choi et al (<http://probe.nalusda.gov:8000/otherdocs/ww/vol2/choi.html>) that is currently available at the Ohio Stock Center. The other BAC library was constructed by Thomas Altmann and collaborators (altmann@mpimp-golm.mpg.de) and is also publicly available (<http://ridb.rz-berlin.mpg.de>). A P1 library (the 'M library') developed by Bob Whittier and colleagues at Mitsui is also available at the Ohio Stock Center and a second library (the 'K library') is being tested at the Kazusa Institute.

5. The objective of the AGI is to obtain high accuracy sequence of the entire genome. There was general agreement that it was not possible to set a standard for exactly what high accuracy means or for mechanisms to enforce high accuracy. However, it was generally agreed that a minimal standard would be that > 97% of all sequence would be obtained on both strands or by

two chemistries. It was the opinion of the group that these criteria were of similar importance and that with most clones, about seven-fold redundancy of sequencing would be required for shotgun sequencing.

An unknown factor affecting the accuracy of the sequence concerns the fidelity of the BAC clones. Preliminary experience suggests that the BACS are generally faithful clones of the genome. However, it will be essential to verify the integrity of each BAC. A minimum criterion is that both ends of the BAC should map to the same region of the genome, typically to the same YAC. When 14,000 BAC ends are sequenced, it is expected that, on average, we will have 500 bp of sequence every 5 kb on average throughout the genome. The resulting library of end-sequenced BACS will represent a check on BAC integrity that will assist in revealing any major rearrangements, deletions or additions. No standard was agreed upon for BAC (or P1) integrity checking. However, most groups indicated that comparing fingerprints of tiled BACS would be the most appropriate criterion for integrity. After some discussion, it was agreed that a large-scale effort toward single-pass shotgun sequencing of the entire genome would not be worthwhile because the combination of available ESTs and the high output rate of the AGI collaboration would obviate much of the value of single-pass shotgun sequencing for gene discovery. However, it was also noted that the existing BAC libraries may not provide complete coverage of the genome and/or may contain small rearrangements or mutations; a shotgun library of the whole genome might provide clones to fill gaps and to verify the integrity of the BAC clones. In addition, it was pointed out that chromosome 3 will be sequenced later than other regions of the genome. The group endorsed a proposal by the SPP consortium to do a feasibility study involving shotgun sequencing of clones from chromosome 3. The value of this approach will be reassessed by the coordinating committee as the project proceeds.

6. All of the participating laboratories are committed to early data release via the internet. One approach discussed at the meeting involved daily release of preliminary sequence information (i.e., sequences that have been edited to remove vector and regions of high ambiguity and condensed into >1 kb contigs). *The C. elegans* sequencing groups follow this approach and the community has found it very useful. Two of the U.S. groups, the SPP consortium and the CSH-WU-ABI consortium

intend to release data in this way. Both groups anticipate release of finished, annotated sequence within six months of beginning to sequence a clone. The EU group does not consider it feasible, at the moment, to do daily releases because the consortium is composed of seventeen relatively small sequencing groups with varying levels of technical capabilities. The EU anticipates release of finished annotated sequence within one month of completion. The TIGR and Kazusa groups do not wish to release unfinished sequence because they believe that carefully edited sequence will be most useful to the community. Both groups promised release of information on a given clone to public databases within three to six months after sequencing began. The TIGR group will release finished, annotated sequence within three months of beginning to sequence a BAC. The Kazusa group estimates that they will release finished, annotated sequence within four to six months of beginning to sequence a clone. In all cases, the start date for sequencing a specific clone will be announced on linked WWW sites so that members of the community will know when to expect the finished sequence.

In summary, all of the groups agreed to establish linked WWW pages for posting complete lists of all clones that have been sequenced to date, along with the start dates of those clones that are still in progress, and the anticipated start dates for the next set of clones to be sequenced in the future. Each clone will therefore have a start date that will be widely advertised to the community. All of the groups anticipate that it will take less than six months to completely sequence and annotate a BAC, YAC or P1 clone and that they will deposit the complete annotated sequences in a public database (e.g., GenBank, EMBL, JDB). No sequence information will be withheld from the community for the sole purpose of benefiting selected individuals, groups, or private companies.

7. There was consensus that the value of the sequence obtained is proportional to the quality of annotation. Thus, each group will attempt to achieve a common standard of annotation. Each group will perform BLAST (or FASTA) searches to align ESTs and known genes and gene products to the genomic sequence. In addition, each group will use programs such as GRAIL and GeneFinder to identify ORFs. Annotation should be presented to the community in a format that can be readily accessed and understood by plant biologists worldwide.

It was agreed that all unassigned ORFs would be named according to the *C. elegans* system. A provisional agreement was reached that the following rules of nomenclature will apply: The first letter is the library name. T=TAMU BAC, F=IGF BAC, M=Mitsui P1 clone, K-Kazusa P1 clone, C=cosmid clone from Goodman library. The first letter is followed by the microtiter plate number, then the row and column numbers followed by a dot and the number of the ORF (numbered sequentially from one side of the clone to the other). Thus, a typical ORF might be called t23a11.12 (i.e., a TAMU BAC from plate 23, well a1, the 12th ORF from one end). It was agreed that zeros will not be included (i.e., t23a11.12 but not t23a11.012). It was also suggested that the names be all lowercase for consistency. Sometimes it will happen that after all the ORFs have been named, a new one will be found by some functional test or other criteria. In this case the two ORFs will be named with an extension to the name (e.g., t23a11.12.1 and t23a11.12.2). When two ORFs are found to belong to the same gene or an ORF is found not to be expressed, the name will be deleted. When one ORF spans two or more clones, the entire ORF will be given the name of the 5' region of the ORF.

It was recognized that annotation of a clone at the time of deposit in public databases will rapidly be rendered obsolete because of information about genes being discovered by the community at large. Thus, there will be an ongoing need for annotation of previously sequenced clones. Because most of the groups are funded to produce new sequence, it will be difficult for the groups producing sequence to also take responsibility for revising the annotation of previously completed sequence. There was broad agreement that the task of annotation revision should be institutionalized by assigning responsibility for revision to the curators of the *Arabidopsis* database (AtDB). The group expressed its strong enthusiasm and support for the continued funding of AtDB to make certain that essential informatics components of the *Arabidopsis* Genome Project are not overlooked. Mike Cherry agreed that it was a suitable responsibility for AtDB and agreed to accept the task to the extent that resources permit.

8. Because the U.S. groups associated with the *Arabidopsis* Genome Initiative will need to reapply for funding within 2.5 years, there was concern about the criteria that will be used to evaluate success. It was agreed that each of the groups will be evaluated based on their overall contribution to the AGI collaboration and that the criteria will not simply be dollars per kb.

9. It is considered essential to keep the entire community well informed of technical advances and practical applications of the genome project. Each group will mount a WWW page that will report the contribution of the group to the multinational sequencing effort. Each group will also work through Mike Cherry (AtDB) and the coordinating committee to make certain that community members receive the training required to make efficient use of the extensive sequence data that will be generated over the next several years. In addition, the coordinating committee will evaluate the feasibility of appointing a part-time public relations specialist to produce user-friendly documentation about the progress of the *Arabidopsis* Genome Initiative. These efforts should help to advertise the dramatic impact that sequencing the *Arabidopsis* genome will have on basic and applied research in plant biology.

Signed:

Mike Bevan, Ian Bancroft (EU consortium)
Satoshi Tabata, Kiyotaka Okada (Kazusa DNA
Research Institute)
Joe Ecker, Sakis Theologis, Nancy Federspiel (SPP
consortium)
Dick McCombie, Rob Martlensen, Rick Wilson,
Ellson Chen (CSH-WU-ABI)
Craig Venter, Steve Rounsley, Owen White, Chris
Somerville (TIGR)
Francis Quertier (French Genome Center)
David Meinke (Multinational *Arabidopsis* Steering
Committee),
September 14, 1996

List of Meeting Participants

Multinational Arabidopsis Steering Committee

David Meinke
Oklahoma State University
Stillwater, OK USA

EU Group

Mike Bevan
John Innes Centre
Norwich, UK

Ian Bancroft
John Innes Centre
Norwich, UK

Kazusa DNA Research Institute

Satoshi Tabata
Kazusa DNA Research Institute
Kisarazu, Chiba, Japan

Kiyotaka Okada
Kyoto University
Kyoto, Japan

SPP Consortium

Joe Ecker
University of Pennsylvania
Philadelphia, PA USA

Sakis Theologis
USDA Plant Gene Expression
Center
Albany, CA USA

Nancy Federspiel
Stanford University
Stanford, CA USA

CSH-WU-ABI Consortium

Rob Martienssen
Cold Spring Harbor Laboratory
Cold Spring Harbor, NY USA

Dick McCombie
Cold Spring Harbor Laboratory
Cold Spring Harbor, NY USA

Rick Wilson
Washington University
St. Louis, MO USA

Ellson Chen
ABD-Perkin Elmer
Foster City, CA USA

TIGR Group

Craig Venter
The Institute for Genomic Research
Rockville, MD USA

Steve Rounsley
The Institute for Genomic Research
Rockville, MD USA

Owen White
The Institute for Genomic Research
Rockville, MD USA

Chis Somerville
Carnegie Institution
Stanford, CA USA

French Genome Center

Francis Quetier
GENETHON
Evry Cedex, France

Observers

Michael Cherry
Stanford University
Stanford, CA USA

Greg Dilworth
U.S. Department of Energy
Germantown, MD USA

Machi Dilworth
National Science Foundation
Arlington, VA USA

Edward Kaleikau
CSREES/USDA
Washington, D.C. USA

DeLill Nasser
National Science Foundation
Arlington, VA USA

Henry Shands
ARS/USDA
Beltsville, MD USA

Jim Tavares
U.S. Department of Energy
Germantown, MD USA



The National Science Foundation promotes and advances scientific progress in the United States by competitively awarding grants for research and education in the sciences, mathematics and engineering.

To get the latest information about program deadlines, to download copies of NSF publications, and to access abstracts of awards, visit the NSF Web site at:

<http://www.nsf.gov>

- ☐ Location: 4201 Wilson Blvd.
Arlington, VA 22230

- ☐ For General Information (NSF Information Center): (703) 306-1234

- ☐ TDD (for the hearing-impaired): (703) 306-0090

- ☐ To Order Publications or Forms:
 - Send an e-mail to: pubs@nsf.gov
 - or telephone: (703) 306-1130

- ☐ To Locate NSF Employees: (703) 306-1234