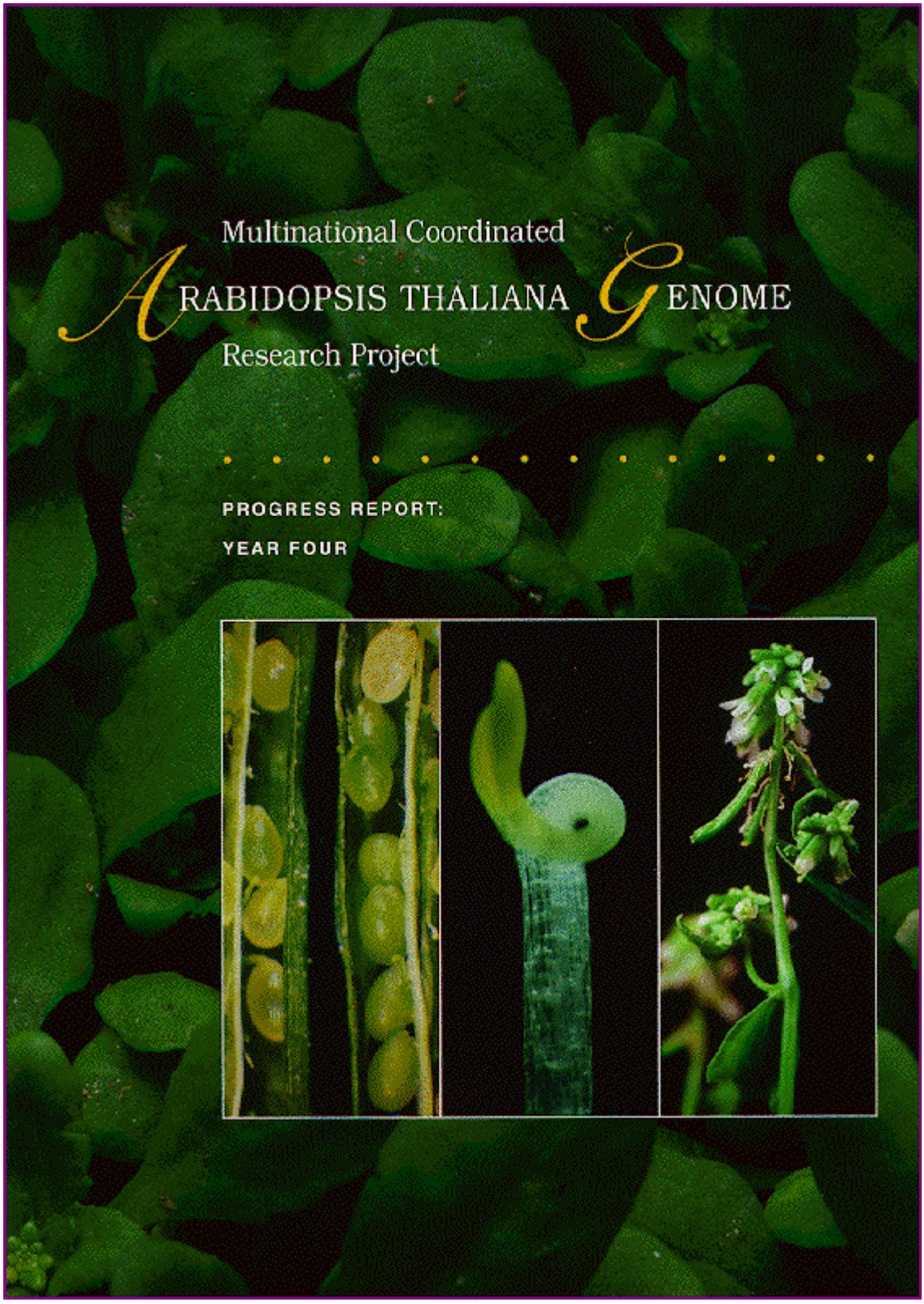
Multinational Coordinated *Arabidopsis Thaliana Genome* Research Project

Progress Report: Year Four



Cover photo credits

- Massed *Arabidopsis* plants, Randy Scholl, Ohio State University
- Young seedling, Joe Ecker, University of Pennsylvania
- Flowering plant, Randy Scholl, Ohio State University
- Seeds, David Meinke, Oklahoma State University

Contents

D. Acknowledgments

National Science Foundation

Preface

In 1990, an ad hoc committee composed of nine scientists from the United States, Europe, Japan, and Australia prepared a report called "Long-range Plan for the Multinational Coordinated *Arabidopsis thaliana* Genome Research Project." The publication (NSF 90-80) outlined a plan for international cooperation in studies of the model plant species, *Arabidopsis thaliana*. The project called for genetic and physiological experiments to identify, isolate, sequence, and understand genes; the establishment of worldwide electronic communication among laboratories; and the creation of databases so that new knowledge would be shared. The project plan also contained mechanisms for formal, annual progress reviews and the establishment of new goals.

This is the fourth annual report to summarize the achievements of the Multinational Coordinated *Arabidopsis thaliana* Genome Research Project. It includes traditional scientific reports as well as a collection of brief articles prepared for a more general audience. The information, graphics, and illustrations were provided by members of the Multinational Science Steering Committee and dozens of generous colleagues (see app. D)1.

It is the nature of any general progress report that represents the work of hundreds of scientists worldwide to fail to include or to misrepresent some significant achievements. We ask our colleagues to overlook such shortcomings and to communicate any concerns to committee members so that future reports will be as accurate as possible.

Our continuing goal is to show the scientific community -- and the public -- that knowledge of *Arabidopsis* can lead to a better understanding of all higher plants, and to an improved quality of life.

The Multinational Science Steering Committee
December 1994

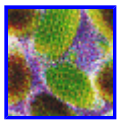
Executive Summary

The Project

The Multinational Coordinated *Arabidopsis thaliana* Genome Research Project is an international scientific collaboration which began in 1990. Its stated goal was -- and remains -- to understand, at the molecular level, the physiology, biochemistry, growth, and development of a flowering plant. The project meets this goal through study of a typical flowering plant, the mustard *Arabidopsis thaliana*. Because it is typical, information gained from study of *Arabidopsis* can be applied to other flowering plants, such as those grown for food and fiber. *Arabidopsis* was chosen for study because it has many advantages for laboratory work, including small size, small genome size, numerous available mutations, and prolific seed production. The project was started by an international group of research scientists who recognized the need to coordinate the various national programs focusing on *Arabidopsis* research.

Goals

The goals and objectives of the project include gaining new insights into the fundamental biological nature of plants, developing new methods for the study of plants, and developing community resources that aid continued scientific progress.



Each year since 1990, an international group of scientists -- the Multinational Science Steering Committee -- has written a detailed report describing the progress of the previous year and listing goals and objectives for the new year. These reports were published and distributed widely by the U.S. National Science Foundation. This is the fourth such annual progress report. Its aim is to increase public awareness of the success the project has had in gaining a better understanding of plants and in applying this new knowledge to important problems in industry and agriculture.

Overview and Prospects

The *Arabidopsis* genome project has made possible an extraordinary string of scientific achievements including, for example, the first molecular identification of a plant hormone receptor, the first molecular analysis of a blue-light receptor, the first production of biodegradable plastic in transgenic plants, the first control of flower development using transgenic plants, and the first complete analysis of the cell patterns of roots. The *Arabidopsis*

genome project is the world's leading source of new information on key aspects of plant growth, development, and metabolism. The excitement generated by rapid scientific progress has encouraged many students and experienced researchers from other fields to join the project. Consequently, the pace of discovery is accelerating.

The project's remarkable collaborative spirit and international character have made it a successful model for scientific cooperation. Participating scientists and scientific administrators come from Asia, Australia, Europe, the Middle East, and the Americas. Continued commitment to the goals of the *Arabidopsis* genome project will lead to a more profound understanding of plants. This knowledge will be used to create new varieties and entirely new types of plants for human use.

Summary of Recent Progress

The project's goals fall into two categories -- promotion of scientific discovery and creation of an infrastructure that supports continued scientific success.

In 1994, scientific progress included breakthroughs in the understanding of plant hormones, plant development, and the interaction of plants with their environment. In this last area, striking progress was made in understanding how plants perceive and respond to light and how they defend themselves against bacterial attack. This new knowledge has already led to practical applications in industry and agriculture. Also, dramatic progress was made in the large-scale analysis of the *Arabidopsis* genome. Large numbers of gene transcripts were sequenced, major improvements were made in the genetic and physical maps of the chromosomes, and a systematic genomic sequencing effort started up.

Examples of research advances include:

- Identification and cloning of genes needed for root and flower development
- Cloning of red- and blue-light receptors
- Identification and cloning of genes involved in response to the plant hormones ethylene and abscisic acid
- Identification and cloning of bacterial defense-response genes
- Genetic map increased to include 400 markers
- Identification of 11,500 expressed sequence tags (ESTs)
- Near-completion of the physical map of chromosome IV
- Start-up of European systematic genomic sequencing project

These achievements depended on the concomitant improvement of the scientific infrastructure. New advances in this area include:

- EST clones made available by stock centers
- New collection of transferred DNA tagged lines made available
- Inauguration of a World Wide Web-based newsletter
- Increased participation in electronic bulletin board
- Planning for development of a new electronic database for *Arabidopsis* research

New Goals for 1995

The project's major new goal is to begin an international collaboration to systematically sequence the entire nuclear genome of Arabidopsis. This goal is feasible because the *Arabidopsis* genome is smaller than that of any other flowering plant and only five times larger than that of yeast, whose entire genome will be completely sequenced by next year. Meeting this goal is critical, because progress in understanding how genes function -- individually or in groups -- is held back by the lack of information on regulatory sequences and on the proteins coded by *Arabidopsis* genes.

Other goals include continuing the large-scale partial sequencing of DNAs complementary to gene transcripts, continuing genetic and biochemical experiments to uncover the function of individual genes, and expanding the support organization which enables rapid scientific progress.

In summary, specific goals for 1995 include:

- Establishment of international, collaborative, large-scale genomic sequencing project
- Continued development of polymerase chain reaction-based marker and other technologies needed for large-scale genome sequencing
- Addition of 5,000 new ESTs to database
- Genetic mapping of 500 EST clones
- Continued research on the function of isolated and cloned genes
- Development of the next generation of public databases on *Arabidopsis*

The Role of *Arabidopsis* in Plant Science Research

Plants are vital to our existence. They provide the oxygen we breathe, the food we eat, the fibers for our clothes, the materials to build our homes, and the raw goods for our industries. A quarter of our medicinal drugs comes from plant species. The paper on which this report is printed is a plant product.

Yet, despite the important contributions of plants to our standard of living, far less is known about them than about mice, flies, or the bacteria that inhabit our intestines. We need to learn more about how plants grow and develop; how they produce useful chemicals; how they protect themselves from pests; and how they sense, respond to, and even alter our environments. One way to learn these things is through study of a plant's genes. The information that plants use to grow and develop, and to interact with their environment, is coded in their genomes. To fully understand plants, we need to read and interpret their genomic information.

In the 1980s, there was a growing awareness that significant investments in studies of many different plants, such as corn, oilseed rape, and soybean, were diluting efforts to fully understand the basic properties of all plants. Scientists began to realize that the goal of completely understanding plant physiology and development is so ambitious that it can best be accomplished by turning to a model plant species that many scientists then study. Fortunately, because all flowering plants are closely related, the complete sequencing of all the genes of a single, representative, plant species will yield much knowledge about all higher plants. Similarly, discovery of the functions of the proteins produced by a model species will offer much information about the roles of proteins in all higher plants.

During the last 8 to 10 years, *Arabidopsis thaliana* has become universally recognized as a model plant for such studies. Although it is a non-commercial member of the mustard family, it is favored among basic scientists because it develops, reproduces, and responds to stress and disease in much the same way as many crop plants. What's more, *Arabidopsis* is easy and inexpensive to grow, and produces many seeds; this allows extensive genetic experiments, often involving tens of thousands of plants. Also, *Arabidopsis* has a comparatively small genome, thereby simplifying and facilitating genetic analysis. Compared to other plants, it lacks the repeated, less-informative DNA sequences that complicate genome analysis.

Initially, there was much debate about whether an improved understanding of *Arabidopsis* would help in the breeding of commercial crops, and much controversy over decisions to devote limited resources to this non-commercial species. However, the many advances reported over the past few years offer clear evidence that this plant is not only a very important model species for basic research, but also extremely valuable for applied plant scientists and plant breeders. Publications on *Arabidopsis* in top-quality journals are increasing exponentially, following substantial increases in investment by many governments. In the United States, for example, the U.S. Department of Agriculture, the Department of Energy, the National Institutes of Health, and the National Science Foundation collectively supplied US$7.5 million in 1990 for *Arabidopsis* research and US$22 million in 1993. And the European Community has invested a significant portion of its biotechnology research resources to *Arabidopsis* genome research over the last 5 years. In fact, many of the world's leading laboratories in plant science have initiated programs using *Arabidopsis*, and many young plant scientists have chosen to start their careers using this species.

But how can discoveries with *Arabidopsis* contribute to the development of improved crops? Simply put, once a gene has been discovered in *Arabidopsis*, the equivalent gene may be found more easily in other plants. Thus, the function of many genes isolated from crop plants can be better understood via study of their

Arabidopsis homologues. So knowledge gained from *Arabidopsis* on the defense mechanisms against pathogens, for example, can be used directly to develop disease-resistant plants in other species.

Genetic comparisons between *Arabidopsis* and crop species are increasing, as shown by the large number of *Arabidopsis* publications cited for 1993 that also involved studies of crop plants such as soybean, rice, maize, wheat, barley, rye, pepper, tomato, potato, cotton, or sorghum. There is ample reason to believe that, in the coming years, *Arabidopsis* will serve more and more as a resource base for breeders of crop plants and as a model plant species that furthers the knowledge of plant scientists worldwide.

Highlighting Progress: *Arabidopsis* Spurs Basic and Applied Research

Arabidopsis has several features, such as rapid growth and small size, that make it an ideal experimental model for plant biology research. But such natural qualities alone are not enough to make it a popular organism for experimental work. Equally important to researchers is the rapid development of new tools which allow a detailed probing of the plant's genome. In the past few years, *Arabidopsis* researchers have produced a variety of such tools including synthetic DNA markers for mapping the genome, a collection of new mutants, specialized transformation techniques, and a large collection of partially sequenced complementary DNAs, which represent genes that are expressed.

Another important resource is the combined collection of genetic maps, which were developed from information generated in many collaborating laboratories. The maps are of critical importance for gene cloning and genetic analysis. Also significant is the establishment of databases, stock centers, and other research infrastructures which allow rapid dissemination of information and easy exchange of ideas and materials. Progress on development of these research tools was outlined in the third annual progress report for the Multinational Coordinated *Arabidopsis thaliana* Genome Research Project (NSF 93-173). Their further development is described in the next section of this report.

All of these research tools and resources allow scientists to dissect the *Arabidopsis* genome systematically. This has resulted in the identification of individual genes and their functions and -- more generally -- a better understanding of plant processes.

This section offers specific examples of how studies of *Arabidopsis* have greatly improved our understanding of disease resistance, root development, and other plant processes. The pace of *Arabidopsis* research has accelerated so much in the past year that it is not possible to summarize all recent, significant advances in this report. Thus, the following highlights are meant to be illustrative rather than comprehensive.

1. Scientific Highlights

Resistance to Microbial Pathogens

Because the study of how pathogens attack plants involves the study of two organisms, it is useful to manipulate both the pathogen and its host genetically in order to tease apart various aspects of plant-pathogen interactions. The basic strategy is to identify and study *Arabidopsis* mutants that show either enhanced or reduced resistance to a particular pathogen. A goal is to apply the knowledge obtained with

Arabidopsis to crop plants which may be genetically engineered to have increased resistance to important agricultural pathogens.

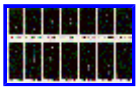
Plant breeders and plant pathologists have long known that certain varieties of crops are more resistant than others to particular viral, bacterial, or fungal pathogens. Indeed, breeding for disease resistance is a major goal of most plant-breeding programs. However, this is time consuming. Also, it involves many crosses - and may be only partially effective. Although individual genes which confer disease resistance have been identified by this process, until this year too few of them had been cloned to gain detailed insight into how these genes actually work at a molecular level. The molecular cloning this past year of an *Arabidopsis* resistance gene, called RPS2, has significantly added to our understanding of how this gene, and similar ones, work.

Unlike resistance genes in crop plants that were found largely in natural wild populations, the *Arabidopsis* RPS2 disease-resistance gene was found by mutagenizing *Arabidopsis* plants in the laboratory. Most *Arabidopsis* varieties resist infection to a particular strain of the agriculturally important bacterial pathogen known as Pseudomonas syringae. *Arabidopsis* mutant plants that were susceptible to infection by this pathogenic P. syringae strain were therefore sought in the labs of Brian Staskawicz (Berkeley, CA) and Frederick Ausubel (Boston, MA). The susceptible *Arabidopsis* mutants pointed to the RPS2 gene, which was subsequently cloned in the Staskawicz and Ausubel laboratories. RPS2 encodes a novel protein containing a motif made of 14 imperfect leucine-rich repeats which is involved in protein dimerization, and a motif that binds adenosine triphosphate. Significantly, this protein likely serves as a receptor for a specific molecular ligand from the pathogen. The structure of the RPS2 protein agrees with models proposed by plant pathologists; this suggests that resistance genes encode receptors for molecular signals from the pathogens. A signal transduction cascade leads to the activation of a variety of defense responses.

Remarkably, RPS2 is similar to the product of the RPP5 gene of *Arabidopsis*, which confers resistance to the fungal pathogen Peronospora parasitica. The RPP5 gene has recently been cloned by Jonathan Jones (Norwich, UK). Moreover, the two *Arabidopsis* resistance genes RPS2 and RPP5 are also similar to the N gene of tobacco, which confers resistance to tobacco mosaic virus; the L6 gene of flax, which confers resistance to the fungal rust pathogen Melamspora lini; and the Cf9 gene of tomato, which confers resistance to the fungal pathogen Cladosporium fulvum. The cloning of these resistance genes, all within the last 12 months, is a major accomplishment.

Root Development

The promise of the *Arabidopsis* root as a model for plant organ formation began to be realized this year. The attractive features of this root for developmental studies include its simple architecture, transparency, and continuous developmental program. Laboratories at the John Innes Centre (Norwich, UK), University of Utrecht, New York University, University of Michigan, and National Institute of Basic Biology (Japan) -- among others -- have made significant contributions. Noteworthy was the publication of a cell lineage map of the *Arabidopsis* embryonic root and root meristem by Ben Scheres' laboratory (Utrecht, The Netherlands): It was the first complete root fate map to be generated for any species.



A variety of *Arabidopsis* mutations that affect root development have also been described recently. They are being used to better understand how roots develop. For example, mutants have been identified in Philip Benfey's laboratory (New York, NY) in which the radial pattern of embryonic root meristem is disrupted. Also, mutants have been found in the laboratories of Scott Poethig (Philadelphia, PA); Keith Roberts (Norwich, UK); and John Schiefelbein (Ann Arbor, MI) in which the signaling process, which specifies the fate of epidermal cells to form root hairs, has been disrupted.

Flower Development

Floral growth begins with development from a vegetative meristem, which produces leaves, to an

inflorescence meristem which may branch to form several floral meristems, each of which develops into a separate flower. Then, each floral meristem undergoes developmental controls; these determine the formation of floral organs such as petals, sepals, and stamens.

During the past year, there have been three noteworthy advances in understanding flower development. First, interactions among the meristem identity genes, which control the fate of the meristems, have been clarified. Second, molecular confirmation has been obtained for the regulation of a number of meristem and organ identity genes by other genes. Third, regulation of floral homeotic genes (organ identity genes) by the meristem identity genes has been demonstrated. In the wake of these advances, there are now more than a dozen laboratories working in this area worldwide.

Light Signal Transduction

Genetic analysis of *Arabidopsis* has shown that light responses are not endpoints of a linear signal pathway. Instead, they are the result of the integration of a variety of input signals through a complex network of interacting signals. Two main classes of genes in the light signal transduction pathways have been identified. Photoreceptor genes encode either red-/far red- light receptors (phytochromes) or a putative blue-light receptor; signal transduction pathway genes encode proteins that carry signal from the photoreceptors.

Studies in the laboratories of Peter Quail (Albany, CA), Nick Harberd (Norwich, UK), and Joanne Chory (La Jolla, CA) show distinct and overlapping roles for two phytochromes in the control of light-regulated responses. The two phytochromes appear to be partly redundant, and absorption of light by either phytochrome A or B leads to plant responses. Mutants have shown which response is due to each phytochrome. Mutants have also helped uncover a role for the Pr form of phytochrome B in the control of germination and shoot gravitropism. Previously, it was thought that this phytochrome had no biological activity.

The chemical nature of a blue-light receptor in plants was previously unknown, but has been revealed by the cloning of the Arabidopsis HY4 gene in Anthony Cashmore's lab (Philadelphia, PA). The predicted protein is similar to photolyases, which are known to be flavin-binding proteins. These results suggest a mechanism by which blue-light photoreceptors trigger a physiological response in higher plants. In addition, a variety of mutations have been isolated that affect the entire morphogenetic program of young seedlings in the dark.

Hormone Signal Transduction

Investigators studying the action of ethylene, which is an important plant hormone, have identified mutations that alter sensitivity to exogenous ethylene. By constructing and characterizing double mutant combinations, Joseph Ecker (Philadelphia, PA) has found the order of action of various gene products. Using this approach, the ETR1 gene was shown to function before CTR1 in an ethylene response pathway. Both the ETR1 and CTR1 genes were recently cloned, leading to an ability to control the response of *Arabidopsis* to ethylene (Caren Chang at the California Institute of Technology and Joseph Ecker).

The past year also saw the cloning of a gene for response to abscisic acid, another important plant hormone. It was accomplished by the laboratories of Jerome Giraudat (Gif-sur-Yvette, France) and Erwin Grill (Zurich, Switzerland).

2. Commercial Benefits

As predicted, discoveries made with *Arabidopsis* are leading to improvements in commercial crops. For example, even though the flowers of *Arabidopsis* are very different from those of snapdragons, the same genes control flower development in both. And those genes that guide the synthesis of oils in *Arabidopsis* are closely related to those that produce oils in commercial oil crops. Indeed, this relation is being exploited to produce plants with more desirable, edible oils.

About one-third of the calories in our diets comes from soybean or other vegetable oils. However, most vegetable oils are not well-suited to food uses because they are highly polyunsaturated. For many uses, the polyunsaturated oils are chemically modified by catalytic hydrogenation; this causes many double bonds to isomerize from cis to trans. Although the nutritional consequences of these trans unsaturations are even now being debated in the medical community, there is widespread interest in developing a new method of producing less highly saturated oils. During the past several years, genes for most of the fatty acid desaturases have been cloned from *Arabidopsis*. These have been used, first, to identify the corresponding genes from soybean, canola, and several other crop species; and, second, to genetically engineer crop plants with reduced levels of polyunsaturation. Early results suggest that it may be possible to eliminate the need for catalytic hydrogenation and to fine tune the oil composition of some crops to better suit human nutritional needs.

Because of the rich base of genetic information about lipids, oils, and starch in *Arabidopsis*, this plant has also been the test organism for efforts to produce biodegradable plastics in crop plants. In the first series of experiments, several genes from the bacterium Alcaligenes eutrophus were introduced into *Arabidopsis* so that the gene products built up in the cytoplasm. This resulted in the accumulation of small amounts of polyhydroxybutyrate (PHB), a biodegradable plastic. Recently, the amount of PHB accumulated by *Arabidopsis* plants was increased about a hundredfold by transferring the three genes from A. eutrophus to transgenic *Arabidopsis* plants so that the gene products collected in the plastids. These plants accumulated as much as 20 percent of their dry weight as PHB. This level of accumulation is considered adequate to merit development as a possible commercial product. Based on these results, several large chemical companies have started active research programs to develop transgenic crops that produce PHB. It appears that the results obtained with *Arabidopsis* can also be obtained with several commercial oilseed crops: The first field trials are expected in 1995.

One of the most active areas of *Arabidopsis* research concerns the mechanisms by which plants sense and respond to small signal molecules or "phytohormones." The gas ethylene has been long known to affect plant growth and development: It is used to alter the ripening of fruits and vegetables and the aging of flowers. Because of this, there is broad interest in preventing plants from producing or responding to ethylene in certain situations. A new strategy for the possible manipulation of a plant's response to ethylene comes from the recent discovery of a gene in *Arabidopsis* which mediates the biological effects of this gas. It is thought that this gene encodes a receptor which binds ethylene and then triggers a cascade of biological responses. A mutant form of this gene has also been isolated from *Arabidopsis* which prevents the normal response to ethylene. In fact, it has been shown that introducing this altered, unresponsive gene into plants such as tomatoes prevents the plants from responding to ethylene. This could significantly slow down the rate of fruit-ripening or wilting of flowers, keeping them fresh longer.

Although still in the research stage, it is expected that many other recent discoveries in *Arabidopsis* will be rapidly applied to the improvement of plants. Most promising is the recent cloning of specific disease-resistance genes which, in turn, should lead to new mechanisms for engineered disease resistance. On another front, the progress in characterizing genes for flower development will permit the production of new horticultural varieties.

Interview with Daphne Preuss

"Ten years ago, I never imagined I would be working with plants," says Daphne Preuss, an *Arabidopsis* researcher and faculty member in the Department of Molecular Genetics and Cell Biology at the University of Chicago.

At that time, Preuss was studying biophysics. From there, she turned to genetics, choosing yeast as her experimental organism. But about 4 years ago, she made a radical move: She decided to continue her genetics research using *Arabidopsis*, instead of yeast.

Why? Preuss says the plant is easy to work with and offers solid opportunities for young researchers. "Much less is known about plants than animals or microbes, such as yeast, and *Arabidopsis* offers an excellent tool for learning about plant systems," she says. Also, a better understanding of plants could offer solutions to important practical problems, such as the need to better manage the environment and produce more reasonably priced drugs. "We should take more advantage of the little green factories known as plants to make pharmaceuticals easier and cheaper to produce," she says.

Preuss's own research involves using *Arabidopsis* to study plant reproduction. In particular, she is looking at what disrupts communication between the male and female gametes, leading to failures in fertilization and reproduction. She says *Arabidopsis* is ideal for such studies because it is easy to get large quantities of pollen, which make up the male gametes of plants, and to find mutants. Preuss notes that she would not be surprised to learn that the factors that hinder or promote fertilization in *Arabidopsis* are related to those that affect reproduction in other species.

"Not long ago, *Arabidopsis* was a fringe discipline," says Preuss. "But recently, it has attracted researchers studying various crops species, such as corn and soybean, as well as researchers from outside plant biology, from bacterial, yeast, and fruit fly genetics, for example. *Arabidopsis* research is becoming more productive because scientists representing diverse fields are bringing their expertise to the discipline. *Arabidopsis* research is in a growth phase."

Highlighting Progress

Technical Progress of the Past Year on Genome Analysis and Research Resources

Since the initial long-range research plan was developed in 1990, knowledge about the *Arabidopsis* genome has increased to the point where a large-scale, systematic sequencing of the entire 100 megabase (Mb) genome is a realistic goal. This section summarizes progress made during the past year on research on the *Arabidopsis* genome. Ongoing efforts in building research tools and resources are also described.

1. Genome Analysis

Large-scale DNA Sequencing

Last year's report outlined two projects -- the complementary DNA (cDNA) sequencing projects -- which were designed to systematically sequence expressed genes of *Arabidopsis*. The cDNA projects are producing partial sequences called expressed sequence tags (ESTs). These ESTs are sequences unique to each cDNA, and serve as markers. During the past year, cDNA sequencing efforts have continued to produce useful information and EST markers, both of which are widely used by the *Arabidopsis* research community. In addition, a new effort -- the European Scientists Sequencing *Arabidopsis* (ESSA) Project -- is underway to start a systematic sequencing of the entire *Arabidopsis* genome.

European Scientists Sequencing *Arabidopsis* Project: Systematic sequencing of the *Arabidopsis* genome began last year, when a European Union network, funded by the European Community, was set up to sequence -- on a pilot scale -- 2 megabases (Mb) of chromosome IV and 0.5 Mb of other regions of genomic DNA. The ESSA effort also encompasses partial sequencing of 3,000 novel cDNAs, development of a sequence informatics node, and preparation of sequence-ready libraries. The project started up in September 1993, and most labs began work by early 1994.

The largest contiguous region sequenced so far has been 16 kilobases (kb) surrounding the GAP- 

A gene on chromosome III. In addition to the GAP-A gene, four novel open reading frames and a retrotransposon have been identified, as well as a peculiar AT-rich tract. The density of genes in this area, together with the availability of a means to identify open reading frames, is promising. A density of one gene every 4 to 5 kb is expected; this is close to what was previously predicted. Based on a genome size of 100 Mb, this suggests a total of 20,000 to 25,000 genes. Data are still arriving on the large regions of chromosome IV, as the participating labs have had to learn new methods of large-scale sequencing. Regions of overlap show a high degree of accuracy in the independently sequenced areas. By early 1995, the first year's quota of 350 kb of chromosome IV should be done. An analysis of this region will provide new information on gene density, clustering of gene families, the composition of intergenic DNA, and the sequences of novel plant genes.

The major limiting step in systematic genome sequencing is the provision of sequence-ready libraries: Present cosmid (i.e., small DNA fragments cloned from the genome) coverage accounts for only 80 percent of the regions to be sequenced in the next 2 years. The increased effort being put into yeast artificial chromosome (YAC -- i.e., cloned DNA corresponding to a large fragment of the genome) coverage means that YACs must be the main source of sequence substrates. Consequently, new methods for deriving random libraries from YACs are under study.

For more information, contact Michael Bevan, John Innes Centre, Colney Lane, Norwich, NR4 7UJ, UK; phone: 44-16-03-52571, ext. 2518/2520; fax: 44-16-03-505725; e-mail: bevan@bbsrc.ac.uk.

Large-scale cDNA Sequencing -- The French Program: This is the third year of the French cDNA sequencing program. Seven libraries have been used, representing tissue from etiolated seedlings, cell suspensions, green shoots and leaves, flower buds, immature siliques, dry seeds, and wounded leaves.

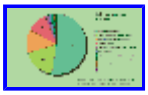
The project's major change in 1994 was that its initial support from CNRS (Centre National de la Recherche Scientifique) would no longer be available for EST work. Funding has subsequently been taken over by the ESSA program, although GREG (Groupement de Recherches et d'Etudes sur les Genomes) supports sequencing of full length cDNAs. ESSA only provides for sequencing of new clones. Because of the redundancy within and between libraries and efforts by the American consortium, new genes are found less frequently. As a result, different groups have set up screening procedures to increase chances of finding new genes.

With the availability of the new YAC library, more efforts will focus on mapping cDNAs on the *Arabidopsis* chromosomes. About 70 cDNAs have already been mapped by the Institut National de Recherche Agronomique (INRA) group in Versailles. All together, the French teams have now released about 4,100 ESTs -- to the European Molecular Biology Laboratory database and the Database of Expressed Sequence Tags (dbEST) -- out of the 6,000 that have been sequenced. Also, most of these clones have been sent to the Ohio State University *Arabidopsis* Biological Resource Center (ABRC) for distribution. This represents a deposit of 2,500 ESTs for 1994, of which 1,700 correspond to 650 new genes with 5' and 3' tags. So far, ABRC has distributed 460 clones to 170 people, in addition to those directly distributed by French groups. Finally, four teams are participating in the ESSA genomic sequencing effort: They have determined 74 kb around four different loci.

For more information, contact Michel Delseny, URA 565 CNRS, University of Perpignan, Perpignan 66860, France; phone: 33-68-662119; fax: 33-68-668499; e-mail: Delseny@univ-perp.fr.

Large-scale cDNA Sequencing -- The Michigan State University (MSU) Program: The goal of the MSU *Arabidopsis* cDNA sequencing project is to produce 36,000 ESTs in order to identify more than 80 percent of the genes expressed by this organism. The cDNA library being used, PRL2, is composed of cDNAs generated from equal quantities of four pools of messenger RNA (mRNA). These four mRNA sources were 7-day-old etiolated seedlings; roots grown in tissue culture; rosettes from plants (staged weekly), half with a 24-hour light cycle and half on a 16-hour light/8-hour dark cycle; and aerial tissue (stems, flowers, and siliques) from the staged plants. The Ziplox vector was used for directional insertion of the oligo-dT primed

mRNA. Until normalization of this library is achieved, it is screened to eliminate the most redundant clones.



The project was initiated in 1992 with funding from the U.S. Department of Energy and the State of Michigan; in late 1993, the National Science Foundation (NSF) granted project funding for 3 years. Since February 1994, technical personnel, working with two ABI 373A automated fluorescent sequencers and an ABI catalyst 800 molecular biology workstation, have produced 7,400 quality sequences, or an average of 925 sequences per month. The average edited sequence is more than 350 b in length.

The project's biocomputing group, based in Minneapolis, Minnesota, edits and analyzes the sequences. The group has programs that analyze the quality of the sequence and trim off vector and 3Õ low-quality regions. The edited sequence is then formatted and deposited in the dbEST at the National Center for Biotechnology Information (NCBI). Data from the project, including fully tabulated BLASTX and BLASTN analyses on the clones, are available through the World Wide Web (WWW; see below for contact information).

Comparing several ESTs to previously sequenced clones indicates that over half the cDNA clones are essentially full length, i.e., they encode the translational start site. More than one-third of the ESTs have significant homology to known genes. About 10 percent of the ESTs that show similarity to genes have not yet been identified in any plant species.

The biological materials from this project include cDNA clones and the PRL2 library. More than 1,500 cDNA clones and 125 aliquots of the PRL2 library have been sent to laboratories worldwide.

For information about the MSU project, contact Thomas Newman, MSU-DOE Plant Research Laboratory, Michigan State University, East Lansing, MI 48824-1312, USA; phone: 517-353-0854; fax: 517-353-9168; e-mail: 22313tcn@msu.edu. Project data are available at http://www.ncbi.nlm.nih.gov and http://lenti.med.umn.edu; boolean keyword searches can be performed on the data using XMOSAIC. The biological materials from this project can be obtained from ABRC.

Genome Mapping

Maps provide valuable reference points for research in molecular genetics. Recently, great efforts have been made to expand two types of maps for genetic information in *Arabidopsis* Ñ a genetic map, which plots the estimated arrangement of genes on each chromosome; and a physical map, which determines the actual distances between markers on a chromosome in terms of kilobases. Completion of the high-density physical map and integration of the genetic and physical maps are two goals of the Multinational Coordinated *Arabidopsis thaliana* Genome Research Project.

Mapping Mutants: Significant advances were made in 1994 in mapping the chromosomal locations of mutant genes. The most recent map of genes identified by mutation, which was compiled by Maarten Koornneef (Wageningen, The Netherlands) and David Meinke (Stillwater, OK), includes more than 280 visible markers distributed over five chromosomes. This represents more than twice the number of mutant genes included on the genetic map just 2 years ago. Embryo-defective mutants, isolated and characterized by David Meinke and colleagues, represent the largest collection of new visible markers added to the genetic map. With recent advances in mapping procedures - particularly the wide distribution of cleaved amplified polymorphism sequence markers introduced by Frederick Ausubel (Boston, MA) and simple sequence length polymorphism markers developed by Joseph Ecker (Philadelphia, PA), it is likely that another 50 to 75 mutant genes of widely different types will be added to the genetic map this year.

Physical Maps: The number of DNA-based markers mapped continued to increase throughout 1994; the total is currently about 379. (See app. B for the latest map.) As new markers became available, they were hybridized to the YAC libraries to increase YAC coverage of each chromosome.

The biggest advances in the linking of YAC contigs were achieved once the YAC library --

prepared in a collaborative effort based in France by the Center for the Study of Human Polymorphisms (CEPH), INRA, and CNRS (a collaboration known as CIC) -- was distributed in May 1994. This library consists of 1,152 clones with an average insert size of 450 kb; it contains very few chimeric YAC clones.

The library is being used by the Joseph Ecker (Philadelphia, PA) and Howard Goodman (Boston, MA) labs to add to the physical maps of chromosomes I, II, and III; and by the Caroline Dean lab (Norwich, UK), for chromosomes IV and V. The combination of the CIC library, increased marker coverage, and a few successful walking experiments has resulted in chromosome IV being covered by just 10 contigs (R. Schmidt, J. West, and C. Dean, all of Norwich). Chromosome V is currently covered by 35 to 40 contigs (R. Schmidt, K. Love, Z. Lenehan, and C. Dean, all of Norwich), after having used over 100 markers on four YAC libraries. Chromosome II (H. Goodman lab) is covered by about 40 contigs, for a total distance of about 17 Mb. The largest contig on chromosome II -- constructed in part using YAC end probes -- is about 3.2 Mb, covering about 5.7 centimorgans (cM). A similar level of coverage has been achieved for chromosome I (J. Ecker lab).

Cosmid contigs from the Howard Goodman laboratory and EST clones are also being integrated into the chromosome II, IV, and V YAC contigs. Efforts are continuing toward generating a 1.5 Mb cosmid contig (I. Bancroft, K. Love, and C. Dean, all of Norwich; C. Cobbett, Melbourne; and H. Goodman), covering the region of chromosome IV which is being sequenced as part of the European Community's ESSA program. Currently, nine cosmid contigs, covering 730 kb, have been restriction mapped and distributed to participating laboratories. Joining these contigs will require new strategies, since sequences within the gaps are repetitive or underrepresented in the cosmid libraries being screened.

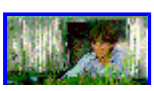
Gene Identification

Significant advances have been made in the cloning and molecular characterization of genes originally identified by mutation, and the isolation of new mutants with informative phenotypes. Two conclusions can be drawn from this work:

- Mutagenesis in *Arabidopsis* has not yet reached saturation. Despite the large size and impressive diversity of existing mutant collections, many genes with important functions remain to be identified.
- With continued advances in chromosome walking and insertional mutagenesis programs, many additional mutant genes should be cloned over the next several years. It should then be possible to determine what metabolic and regulatory functions are disrupted in the many mutants characterized by the *Arabidopsis* community.

Since mid-1993, impressive progress has been made in cloning genes identified by mutation. Published examples include genes involved in hormone perception and response (ABI1, AXR1); flowering (LD, PI, AP2, TSL, MS2, CAL, and SUP); vegetative development (GL2, FEY, PFL, and PAC); basic metabolism (FAD2); resistance to plant pathogens (RPS2); perception of light (HY4); essential cellular functions (EMB30); and transduction of environmental and developmental signals (DET1, FUS6, and COP9). Just 10 years ago, this type of progress would have been considered impossible.

Some of the mutant genes are related in sequence to important regulatory genes known from different organisms. Others represent novel sequences which may provide new insights into eukaryotic cell function. Included in this collection are several genes cloned by chromosome walking, one gene cloned by transposon tagging, and a large number of genes cloned by transferred DNA (T-DNA) insertional mutagenesis. Continued availability of a large collection of T-DNA tagged lines, and ongoing advances with transposon tagging and chromosome walking, should lead to further growth in the number of mutant genes cloned in 1995.



Considerable progress has also been made in expanding existing collections of mutants. For example, Chris Somerville (Stanford, CA) and his colleagues reported the identification of

about 40 novel mutants with altered cell wall polysaccharide composition, which were isolated by screening 5,000 mutagenized lines by gas chromatography of sugar derivatives. These mutants should permit a new approach to the difficult problems associated with understanding cell wall biosynthesis and function.

fied by screening for mutants with phenotypes similar to well-established mutants. A good example is the identification by Ruth Finkelstein (Santa Barbara, CA) of several additional ABI loci involved in mediating responses to abscisic acid. Another example is the discovery of additional members of the leafy cotyledon class of mutants by Peter McCourt (Toronto, Canada); Helmut Baumlein (Gatersleben, Germany); John Harada (Davis, CA); and David Meinke (Stillwater, OK). These few examples underscore the conclusion that further screening of mutagenized populations is likely to yield additional examples of new genes with functions similar to known genes.

More detailed analysis of existing mutants has also resulted in interesting overlaps between phenotypic classes. For example, studies by John Schiefelbein (Ann Arbor, MI) and colleagues have recently shown that the ttg mutant -- known for many years as being defective only in trichome formation and seed coat pigmentation -- also exhibits interesting defects in the spatial distribution of root hairs. Recent efforts in other laboratories have shown that several fusca mutants, first identified by Andreas Muller (Gatersleben, Germany) based on inappropriate accumulation of anthocyanins during embryogenesis, are identical to several de-etiolated and constitutive photomorphogenic mutants. Joanne Chory (La Jolla, CA) and Xing-Wang Deng (New Haven, CT) showed that these latter mutants exhibit interesting defects in photomorphogenesis following germination.

2. Biological Resource Centers

The *Arabidopsis* resource centers were established in 1991 to preserve and distribute biological materials supporting the *Arabidopsis* genome research project. These centers also disseminate genome-related information to the large *Arabidopsis* research community. The three stock centers -- the *Arabidopsis* Biological Resource Center (ABRC) at Ohio State University in Columbus, Ohio; the Nottingham *Arabidopsis* Stock Centre (NASC) at the University of Nottingham, United Kingdom; and the European DNA Resource Center at the Max Planck Institute for Genetic Research in Cologne, Germany -- share these duties. The dramatic expansion of *Arabidopsis* research over the last 5 years is demonstrated by the number of stocks the centers hold and distribute, and the worldwide dispersion of researchers they service.

The resource centers have seeds and clones that are useful for research, especially exploration of the genome. For example, the large seed collections of Albert Kranz (Frankfurt, Germany) and George Redei (Columbia, MO) -- accumulated through their long careers -- have been incorporated into the stock centers. Together, these comprise 1,000 stocks, including many important mutant and wild-type lines collected from all over the world. The mapping collection and mutants of Maarten Koornneef (Wageningen, The Netherlands) are also available. In addition, new mutants affecting development and metabolism are being generated in many *Arabidopsis* laboratories, and are being shared through the three centers. About 300 such lines are now distributed, and 70 new donations -- received in response to a recent campaign -- were recently made available.

T-DNA lines and transposable element-transformed lines are useful tools for cloning genes with identifiable phenotypes. More than 5,000 T-DNA lines are being distributed by the centers; 1,600 new lines have been received from Kenneth Feldmann (Tucson, AZ). Moreover, about 200 transposon lines are held. The stock centers also have about 100 promoter trap lines, a new resource for isolating genes identified through their expression pattern. More donations of these useful stocks are expected soon.

Recombinant inbred populations have become very useful for genetic mapping. Two recombinant inbred populations, consisting of a total of 450 lines, are held by the stock centers. Trisomic stocks, lines transformed with specific genes, and representatives of related species are also held.

The DNA resources distributed by the stock centers include about 300 restriction fragment length polymorphism (RFLP) mapping clones, four YAC libraries, 50 individual clones, 6,000 ESTs, cDNA libraries, genomic libraries, and filters generated from the YAC libraries suitable for probing. In response to a recent campaign by the Ohio center for donations, a number of cloned genes, new RFLP clones, and cDNA and genomic libraries were deposited. Also newly received are the *Arabidopsis* RFLP Marker Set donated by T. Schaeffner (Munich, Germany); a cosmid library containing T-DNA transforming sequences from Kenneth Feldmann; and a hybrid library from John Walker (Columbia, MO). Also, Robert Whittier (Tsukuba, Japan) has agreed to donate his Pl bacteriophage library.

Stocks from the resource centers are distributed worldwide. The number of stocks sent have increased significantly in the last 3 years, going from 15,000 total seed stock distributed in 1992 by ABRC and NASC combined, to about 45,000 seed stocks distributed in 1994. As for DNA, l,000 clones and six YAC libraries were sent in 1991; just 2 years later, about 3,100 clones and 166 libraries were sent. Distribution of ESTs was started in late 1993: About 1,300 ESTs have been sent out since. Increasing numbers of mutants and new batches of T-DNA lines will be donated to the centers in the near future. Also, clones representing the complete physical map and new ESTs are expected.

Ordering and Contact Information

ABRC: Randy Scholl, *Arabidopsis* Biological Resource Center, Ohio State University, Columbus, OH 43210, USA; phone: 614-292-9371; fax: 614-292-0603; e-mail: arabidopsis+@osu.edu; *Arabidopsis* Information Management Systems (AIMS) WWW server URL http://genesys.cps.msu.edu:3333/.

NASC: Mary Anderson, Nottingham *Arabidopsis* Stock Centre, Department of Life Science, Nottingham University, Nottingham NG7 2RD, UK; phone: 44-1159-791216; fax: 44-1159-513251; e-mail: arabidopsis@nottingham.ac.uk; NASC WWW server URL http://nasc.nott.ac.uk.

NASC stock information is distributed through a hard copy seed list, an *Arabidopsis thaliana* database (AAtDB), AIMS, and the AAtDB Research Companion gopher server.

European DNA Resource Center: Jeff Dangl, European DNA Resource Center, Max Planck Institute, Carl von Linne-Weg 10, Cologne D-50829, Germany; phone: 49-221-5062-630; fax: 49-221-5062-613; e-mail: dangl@vax.mpiz-koeln.mpg.dpb.de.

Note that service at this stock center was discontinued as of December 31, 1994.

3. Informatics

Databases

An *Arabidopsis thaliana* Database: AAtDB continues to be a key resource for sharing genetic map information. Its data are presented in graphic, tabular, and text formats. Information in AAtDB is provided by the *Arabidopsis* community, either directly from investigators or from publicly available collections and databases.

- Features: AAtDB is in its sixth data release and includes:
- Genetic markers, including RFLP, random amplified polymorphic DNA (RAPD), and "classical" markers
- Recombinant inbred chromosomal maps; the integrated map; a visible marker map, including many embryo defective loci; and RAPD maps
- Primary F2 and recombinant inbred population recombination data and two-point data
- Stock information, including phenotype descriptions from the Green Book

- Both DNA resources and germplasm resources
- Over 5,800 *Arabidopsis* DNA sequences from GenBank, including recent sequence comparisons
- Bibliographic information from Agricola and Medline
- Contact information for *Arabidopsis* researchers, including postal and e-mail addresses, hone and fax numbers, publications, research interests, and research associates
- Scanned images of RFLP autoradiograms, photographs of mutant plants, and restriction enzyme digests of RFLP probes

To use the system: Distributed versions of AAtDB need a UNIX workstation running the X-windows display or a Macintosh workstation. AAtDB is available over the Internet without charge via anonymous file transfer protocol (FTP) from weeds.mgh.harvard.edu in the AAtDB directory.

The UNIX-based ACeDB software and its C source code files are available via anonymous FTP from ncbi.nlm.nih.gov. A WWW version is available through the server at the National Agricultural Library, http://probe.nalusda.gov:8300.

AAtDB Research Companion: The AAtDB Research Companion provides Internet access to AAtDB. The Companion is a computer source (weeds.mgh.harvard.edu) that serves information to Internet users through WWW, gopher, Telnet, and FTP. A link is also provided to the gopher server, which offers all information in AAtDB in text. In addition, the gopher server includes:

- FTP archives containing files to install the workstation version of AAtDB
- Images of stocks, gels, and hybridizations contained within AAtDB
- Genetic maps and tables
- Archives of the BioSci *Arabidopsis* Genome Electronic Conference
- *Arabidopsis*: Complete Guide, a collection of molecular biology protocols
- *Arabidopsis* Information Service, a searchable index of the 27 volumes
- *Arabidopsis* cDNA sequences from the NCBI dbEST library

For additional information, contact John Morris, Curator, AAtDB project, Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02113, USA; fax: 617-726-6893; e-mail: john.morris@frodo.mgh.harvard.edu.

***Arabidopsis* Information Management System:** AIMS is an on-line database system running on a central machine at Michigan State University. It was originally developed to support data management, including stock ordering and inventory, at Ohio State University's ABRC. The database is implemented on top of a commercially available Sybase system. All AIMS graphics features are offered in an object-oriented fashion using X-windows. Its nongraphics features can also be accessed on a microcomputer or VT100-type terminal. Mosaic interface to AIMS is also available; this provides access to AIMS data and stock ordering.

Features: AIMS manages both data and programs such as MapMaker; it includes a general mechanism to input private data and manage output. Data and features now available or to be added soon include:

- Full information on seed stocks
- Information on all cloned genes available at ABRC
- Information on RFLP and RAPD stocks, including crosses showing the polymorphism and enzymes used
- Information on YAC libraries available at ABRC
- Genetic mapping data
- Sequence and homology search results for all EST cDNA clones
- Color pictures of phenotypes for many of the newer stocks
- Images of gel banding patterns for RFLP stocks
- Images of comparative hybridization results for RFLP clones
- Sequence data and homology results for clones

- References on *Arabidopsis*
- Personnel data, including the ABRC mailing list
- Raw recombination data for linkage experiments

Stocks can be ordered through on-line AIMS and the Mosaic AIMS interface. AIMS keeps track of all on-line and e-mail stock orders, which can be accessed.

To use AIMS: Telnet to this Internet address: genesys.cps.msu.edu (Telnet or X-windows) or http://genesys.cps.msu.edu:3333/ (Mosaic URL).

For more information, contact Sakti Pramanik, Computer Science Department, Michigan State University, East Lansing, MI 48824, USA; phone: 517-353-3177; fax: 517-432-1061; e-mail: pramanik@cps.msu.edu.

Database of Expressed Sequence Tags: Maintained at the National Center for Biotechnology Information as part of the National Library of Medicine at the National Institutes of Health, Bethesda, Maryland, dbEST carries detailed descriptions of sequences, including putative homology assignments using the Basic Local Alignment Research Tool (BLAST) set of programs. Also, the database has information on contributors, available genetic map locations, and instructions on where to obtain physical DNA clones. The latest release of dbEST as of this writing is version 2.43, which carries over 67,000 entries. *Arabidopsis* is the third largest group of entries.

Arabidopsis EST entries from NCBI are also held on the AAtDB Research Companion, gopher server at Massachusetts General Hospital, Boston, Massachusetts. The EST report files are WAIS indexed to allow rapid searching using keywords.

For more information about dbEST, contact Mark Boguski, National Center for Biotechnology Information, Building 38A, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA; phone: 301-496-1475; fax: 301-480-9241; e-mail: info@ncbi.nlm.nih.gov; WWW URL: http://www.ncbi.nlm.nih.gov.

Communication

BioSci *Arabidopsis* Genome Electronic Conference: The *Arabidopsis* newsgroup, part of the BioSci/BIONET electronic newsgroup network, is an international collaboration via computer networks. Although BioSci runs over 60 newsgroups on various biology topics, the *Arabidopsis* group is the best example of Internet use by a community of international scientists working toward a common goal.

The *Arabidopsis* newsgroup is distributed worldwide through both USENET news, under the name bionet.genome.arabidopsis, and through e-mail. If USENET news access is not available on a personal computer, e-mail subscriptions can be requested by contacting one of the following addresses, based upon location:

```
Subscription address
Location

biosci@daresbury.ac.uk       Europe, Africa and Central
Asia

biosci@net.bio.net           Americas and the Pacific Rim
```

As of November 1994, there were 878 e-mail subscribers, up 21 percent from the previous year. A complete archive of all *Arabidopsis* postings is maintained for anonymous FTP and gopher retrieval on the Internet computer net.bio.net in the directory pub/biosci/arabidopsis. WWW users can connect using the URL gopher://net.bio.net/ and look in the *Arabidopsis* folder.

Arabidopsis postings are also indexed in the general biosci.src WAIS source on the computer net.bio.net. WAIS software indexes all text in every BioSci newsgroup posting and allows Internet users to search for any text string and then retrieve messages bearing the specified text. The WAIS indexes can be queried using either the gopher software or a WWW browser such as Mosaic, as instructed above. In either case, the option to pick is listed as "Search Bionet USENET Articles." The WAIS source can also be queried by e-mail. For WAISMAIL instructions, send the word "help" in the body of an e-mail message (leave the subject line blank) to waismail@net.bio.net. For help using the archives, contact biosci-help@net.bio.net.

***Weeds World*, The International Electronic *Arabidopsis* Newsletter:** The Multinational Science Steering Committee endorsed, at its meeting in Amsterdam in June 1994, the production of an electronic newsletter. Shortly thereafter, *Weeds World* was launched. It is a popular forum for the exchange of information and is used much like *the Worm Breeders' Gazette* which serves the *Caenorhabditis elegans* community. *Weeds World* is published three times a year and distributed through WWW, indexed and carried on the AAtDB Research Comp the *Plant Molecular Biology Reporter*, courtesy of the International Society of Plant Molecular Biology. There will be no regular hard copies made of the newsletter.

The newsletter is produced by Mary Anderson (Nottingham, UK); Sam Cartinhour (Beltsville, MD); and Randy Scholl (Columbus, OH). John Morris (Boston, MA) archives and indexes it on the AAtDB Research Companion. The first edition was published in November 1994.

The URL addresses are:

Nottingham:
http://nasc.nott.ac.uk:8300/home.html

Beltsville:
http://probe.nalusda.gov:8300/ww/home.html

Boston:
http://weeds.mgh.harvard.edu

Books

Several recent books are devoted to various aspects of the biology of *Arabidopsis*:

John Bowman, ed. Arabidopsis: *An Atlas of Morphology and Development*. New York: Springer-Verlag, 1994. 450 pp. ISBN 0-387-94089.

Csaba Koncz, Nam-Hai Chua, and Jeff Schell, eds. *Methods in* Arabidopsis *Research*. Singapore: World Scientific, 1992. 482 pp. ISBN 98l-02-904-5 (hardback); 98l-02-905-3 (paper).

Elliot Meyerowitz and Chris Somerville, eds. Arabidopsis. Cold Spring Harbor, NY: Cold Spring Harbor Press, 1994. 1,300 pp. ISBN 0-87969-428-9.

4. Workshops and Symposia

The number of scientific gatherings involving *Arabidopsis* researchers continues to grow. A summary of those held in 1994 follows.

- **The Cold Spring Harbor Course on *Arabidopsis* Molecular Genetics** was held July 4-24, 1994, and attracted 40 scientists -- 16 students and 24 instructors -- from six nations. The course is supported in part by a 3-year NSF award. The instructors included Joanne Chory (La Jolla, CA); Joseph Ecker

(Philadelphia, PA); and Athanasios Theologis (Berkeley, CA). The program's assistants were Steffer Abel (Berkeley, CA); Patrick Dunn (Philadelphia, PA); Lara Soowal (La Jolla, CA); and Abby Telfer (Philadelphia, PA). This course provided an intense overview of current topics and techniques in *Arabidopsis* biology, with an emphasis on molecular genetics. It also introduced approaches used in yeast that have the potential to advance *Arabidopsis* molecular genetics. It was designed for scientists with experience in molecular techniques who are working or wish to work with *Arabidopsis*. The course consisted of a rigorous lecture series, a hands-on laboratory, and informal discussions. Speakers provided both an in-depth discussion of their work and an overview of their specialty.

The lecture topics included all aspects of plant biology, focusing on molecular and genetic analysis of plant growth, development, and physiology using *Arabidopsis* as a model experimental system. Where appropriate, lecturers touched upon many other experimental systems, including nonplant systems such as the nematode C. elegans. The laboratory sessions covered *Arabidopsis* genetics and development, transient gene expression assays in protoplasts, complementation of yeast mutants for cloning *Arabidopsis* genes, two-hybrid system in yeast, transformation by Agrobacterium, in-situ detection of RNA, biochemical analysis of transcription factors, pulsed-field gel electrophoresis, and analysis of YACs containing the *Arabidopsis* genome and techniques for polymerase chain reaction-based mapping of mutations.

- **The Banbury Center Conference on the *Arabidopsis* Genome --** organized by Joseph Ecker (Philadelphia, PA); Michael Bevan (Norwich, UK); and Robert Martienssen (Cold Spring Harbor, NY) -- was held March 20-23, 1994, at the Cold Spring Harbor Laboratory, Long Island, New York. It attracted about 40 invited participants, and was funded by the Cold Spring Harbor Laboratory Corporate Sponsor Program. The meeting reviewed the current status of the multinational *Arabidopsis* genome project, evaluated what is required to complete sequencing of the 100 Mb *Arabidopsis* genome, and examined what can be done on a national basis to achieve this goal. Participants included scientists working on *Arabidopsis* and genome projects involving other organisms, e.g., mice, Drosophila, C. elegans, yeast, and humans. The meeting included discussions on large-scale genome projects, physical mapping and markers, gene identification, sequencing and informatics, and policy.

  The final session featured a roundtable discussion which yielded consensus on several key issues. Among investigators involved in the physical mapping of the *Arabidopsis* genome, it was agreed that better coordination of individual chromosome mapping efforts was necessary. To achieve this goal, these groups agreed to exchange all YAC libraries, some cosmid contigs, and all associated anchoring/sequence-tagged site mapping data. Participants agreed that a high priority for U.S. databases should be development of a "translator" for data exchange with AAtDB. Several issues were left unresolved regarding genomic sequencing of *Arabidopsis*. It was suggested that the North American *Arabidopsis* Steering Committee (NAASC) meet to focus on genomic sequencing, and that input from the *Arabidopsis* community be solicited before specific recommendations be formulated and submitted to U.S. federal granting agencies.

- **The Workshop on *Arabidopsis* Genome Sequencing** was held at the National Science Foundation, in Arlington, Virginia, June 8-9, 1994. Organized by NAASC as a follow-up to the Banbury Center Conference, it aimed to assess the feasibility and desirability of a federally funded *Arabidopsis* genome sequencing project. Sixteen invited participants, including elected members of NAASC, were present. The workshop report was presented to -- and subsequently endorsed by -- the Multinational Science Steering Committee at its meeting in late June 1994. This report was published on the *Arabidopsis* electronic bulletin board and is presented in appendix A.

- **The XV International Botanical Congress** was held in Yokohama, Japan, August 28-September 3, 1993. *Arabidopsis* research was often highlighted, illustrating the wide use of this plant as a model system for the study of flowering plants. The 18 sessions covered topics such as genetic control of

developmental processes, light-activated transduction pathways, tissue-specific gene expression in transgenic plants, molecular and genetic approaches to phytochrome function, metabolism and function of plant lipids, regulation of gene expression by plant hormones, cell wall ultrastructure and formation, physiological analysis of shoot growth, genome studies by DNA molecular markers, molecular dissection of ethylene biosynthesis, dynamic aspects of leaf senescence, molecular mechanisms of environmental stress, tropism, plant growth and development under microgravity, regulation of gene expression, cis-elements and trans-acting factors, embryo and seed development, molecular and physiological approaches to flower initiation, and biogenesis and differentiation of organelles.

- **The Fourth International Congress of Plant Molecular Biology** was held in Amsterdam, The Netherlands, June 19-24, 1994, and attended by more than 3,000 plant molecular biologists from 60 countries. It was organized by the International Society for Plant Molecular Biology and covered the entire field of plant molecular biology. *Arabidopsis* research, however, played a dominant role in the lectures, discussions, and more than 2,000 poster presentations. Of the 12 invited plenary lectures, five were on *Arabidopsis* research, including the opening talk by Michael Bevan (Norwich, UK) titled "Systematic Sequencing of the *Arabidopsis* Genome." Also, there were demonstratio ncluding summary, Jeff Schell (Cologne, Germany) singled out progress on genome research, noting that the international cooperation exhibited by the *Arabidopsis* research community is a model for scientific collaboration.

- **The Second International Conference on the Plant Genome (Plant Genome II) --** organized by Stephen Heller and Jerome Miksche of the U.S. Department of Agriculture (Beltsville, MD); Michael Gale of the John Innes Centre (Norwich, UK); and Susan McCouch of the International Rice Research Institute (The Philippines) -- was held January 1994 in San Diego. The conference brought together researchers interested in a wide variety of species, and much attention was focused on the similarities between the genomes of different plants. Increasingly, results obtained from one organism can be applied to others, simplifying much research. Progress on analysis of the *Arabidopsis* genome was discussed. For example, Caroline Dean and Howard Goodman reported on their collaboration to integrate their YAC and cosmid physical maps -- progress which has been made in spite of technical problems with existing YAC libraries. The creation of a new YAC library, reported by Dave Bouchez (Versailles, France), should help. Pieter Vogel, from the laboratory of Marc Zabeau of Keygene (Wageningen, The Netherlands), described the use of amplified fragment length polymorphisms; this greatly increases the efficiency of identifying and mapping molecular markers. Michel Delseny (Perpignan, France) summarized the French EST project, which has "tagged" several thousand cDNAs and is being deposited in public databases. Finally, Nazeem Ali from Kenneth Feldmann's laboratory (Tucson, AZ) discussed a way for transposon-mediated mutagenesis which allows specific chromosomal regions to be targeted. In addition, a plenary lecture by Chris Somerville (Stanford, CA) described progress toward large-scale sequencing of *Arabidopsis* cDNAs. Michael Unseld (Berlin, Germany) reported on sequencing of the mitochondrial genome. All meeting abstracts are on-line and searchable at http://probe.nalusda.gov:8000/plant/index.html.

- **The USA-Australia Bilateral Science Agreement Workshop on Plant Genome Technology**, organized by Sheila McCormick (Albany, CA) and T.J. Higgins (Canberra, Australia), was held near Cairns, Queensland, April 22-26, 1994. There were 26 participants from Australia and 18 from the United States. Despite the decidedly agricultural emphasis of the conference, eight of the 38 talks were on *Arabidopsis*.

- In Germany, an international *Arabidopsis* meeting, **the Deutsche Forschunggemeinschaft Meeting for *Arabidopsis* Schwerpunktprogram**, organized by Gerd Jurgens (Tubingen, Germany) and Jeff Dangl (Cologne, Germany), was held October 28-30, 1993, in Breisach, to celebrate the beginning of the National Research Association of Germany's *Arabidopsis* program. Talks covered hormone synthesis and action, signal transduction, cell development and differentiation, control of flowering, light regulation, pattern formation in the embryo, and vegetative development.

Status of National Programs

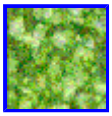
Australia

Australia's Department of Industry, Science, and Technology funds a physical mapping effort in the multinational *Arabidopsis* genome project. In this effort, genes involved in male fertility, P450s, cytokinin metabolism, and cellulose biosynthesis are being mapped and isolated. Random expressed sequence tags (ESTs) are also being mapped to the *Arabidopsis* genome, either to yeast artificial chromosomes (YACs) or by restriction fragment length polymorphism (RFLP). Additionally, *Arabidopsis* projects are ongoing at approximately six Australian institutions; these are focused on floral mutations, male sterility, root morphology, signal transduction on anaerobiosis, and sugar transport.

Contact: Elizabeth Dennis, CSIRO

Belgium

Belgium has several national programs that support *Arabidopsis* research, including some financed by the Flemish Community. One program aims specifically at the further development of *Arabidopsis* as a model plant: It includes mapping and sequencing projects and molecular biological studies of, for example, ethylene response and oxidative stress. In another project, the interaction of plant nematodes with *Arabidopsis* is under investigation. This effort includes (1) mapping and analysis of mutant *Arabidopsis* plants that are resistant to nematode infection and (2) establishment of a protein database of *Arabidopsis thaliana*, using two-dimensional electrophoresis and partial amino acid sequence analysis.



The Belgian Government promotes collaboration among universities. For example, the Laboratory of Genetics at Ghent works with other Belgian universities, such as those in Antwerp and Brussels, that also have a department of plant biotechnology. This particular project involves the study of regulatory genes involved in morphological development, and study -- on the microscopic level -- of interphase nucleus, using fluorescence in-situ hybridization.

Another Belgian project entails the evaluation of a novel genetic strategy -- selective restriction fragment amplification -- to identify markers closely associated with a mutation. This new technique was developed by Keygene, a Dutch company.

Contact: Marc Van Montagu, University of Ghent

Canada

The groups of Bertrand Lemieux at York University and Peter McCourt at the University of Toronto are collaborating with the group of Ronald W. Davis at Stanford University to map EST sites by polymerase chain reaction (PCR) amplification. These groups are using DNA isolated from YAC clones of a French collaboration - comprised of the Center for the Study of Human Polymorphisms (CEPH), the Institut National de Recherche Agronomique (INRA), and the Centre National de la Recherche Scientifique (a collaboration also known as CIC) - the YAC bank at the University of Pennsylvania, and the Grill-Somerville banks as template DNA. To date, primers for the amplification of 800 ESTs have been produced by Deval Lashkari (Stanford, CA); and DNA has been isolated from 3,456 *Arabidopsis* clones by Gus Lagos (York, Canada). The YAC clone pooling strategy of Balding and Torney (Los Alamos, NM) is being used; this allows complete mapping of EST sites without colony hybridizations. This 2-year project has a target of 2,000 EST markers.

Contact: Bertrand Lemieux, York University

European Community

The Biotechnology Research for Innovation, Development, and Growth in Europe (BRIDGE) project, "Molecular Identification of New Plant Genes," supported by the European Community (EC), ended in June 1994. Highlights of the program included the isolation of several genes that mediate environmental effects on flowering and the isolation of two genes involved in responses to abscisic acid. The transposon tagging development program, based on the transposons Ac and En, has provided two very useful systems. Using these, several thousand families containing potential mutants have been generated, as well as a series of lines with mapped transposon "launching pads" for targeted mutations. The physical mapping work has resulted in the assembly of five contigs spanning more than 16 megabases (Mb) of chromosome IV and a large area of chromosome V. The mapping has provided the basis for a large-scale sequencing project (described in the previous section).

The EC program, together with the United Kingdom's Biotechnology and Biological Sciences Research Council, also supported the Nottingham *Arabidopsis* Stock Centre (NASC) and the European DNA resource center at Cologne. (In fact, future plans for NASC may include its serving as another DNA resource center.) Some of BRIDGE's initial goals remain elusive: For

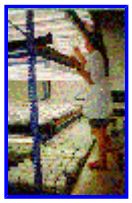
example, the lack of methods for gene replacement represents a serious impediment to progress, especially in determining the function of specific genes. On the other hand, the most enduring aspect of the BRIDGE project is the development of transnational teams, which continue to provide a significant strength to European biological research.

Future work in *Arabidopsis* is well-funded within the Fourth Framework Program, a 4-year plan that follows the EC's budgetary cycle. Work will continue on developing a gene replacement system; this work will be part of several efforts aimed at a large-scale search for gene function -- efforts strongly linked to the sequencing project. Research in such areas as the transition to flowering, plant growth and development, genetic and metabolic controls on carbon and nitrogen assimilation, signal perception, plant-pathogen interactions, and embryo and seedling development can all be advanced via studies of *Arabidopsis*. The EC is interested in sponsoring transnational research in these fields.

Contact: Michael Bevan, John Innes Centre, Norwich, UK

France

Fourteen French laboratories use *Arabidopsis* as a tool for advancing plant biology research in their studies of embryogenesis, abscisic acid signaling, dormancy, ion transport, root development, lipid metabolism, cell cycle, protein kinases, gametogenesis, plasmalemma proteins, nitrate and carbon metabolism, transposable elements, plant-pathogen interactions, and stomata. Three collective objectives in genome analysis are being pursued; these involve sequencing, mapping, and generation of transferred DNA (T-DNA) tagged lines.



- Sequencing: Four French labs are working on medium-scale genome sequencing under the EC-funded European Scientists Sequencing *Arabidopsis* (ESSA) Project. Additionally, eight labs have continued to pursue the complementary DNA (cDNA) program: Recent advances in this area are described in the previous section.
- Mapping: Two approaches are being developed to map transcripts identified by systematic cDNA sequencing. Five laboratories are using recombinant inbred lines generated at the John Innes Centre to map some of these transcripts by RFLP analysis. The YAC library, generated in conjunction with Daniel Cohen's group (CEPH, Paris), has been further characterized. The assignment of 70 sequences to YACs suggests a low level of chimerism. This library seems suitable for completing coverage of the *Arabidopsis* genome, and for mapping ESTs by PCR or filter hybridization. The French funding agency for genome analysis has agreed to fund the mapping of 500 ESTs by RFLP and YAC assignments.
- Generation of T-DNA lines: The infiltration method devised by Nicole Bechtold (INRA, Versailles) and colleagues has been used to generate a collection of new T-DNA lines in the lab of Georges Pelletier (INRA, Versailles), with a vector allowing the detection of genes by GUS promoter trapping. Twelve thousand lines have been generated so far, and will soon be released to the stock centers. Seven percent of these lines express GUS in different tissues. This collection is comparable, for several criteria, to the Feldmann collection at the University of Arizona.

Contact: Michel Caboche, INRA-Versailles

Germany

More than 30 laboratories in Germany are using *Arabidopsis* as their primary experimental organism. The research projects fall into three broad categories: developmental genetics, molecular physiology, and plant-microbe interactions. Groups working on developmental problems are funded under a 6-year special research program, "*Arabidopsis* as a Model for the Genetic Analysis of Plant Development"; this program was set up by the Deutsche Forschunggemeinschaft, Germany's national research association, in August 1993. More than 30 participants from both Germany and abroad attended the program's first meeting, which was held in

Breisach, October 28-30, 1993. (Highlights of this meeting are described in the previous section.) In 1994, the program convened a specialized workshop on pattern formation; this was held in Tubingen, October 20-21, and attracted 35 participants. Another meeting on hormones will be held in the near future.

Contact: Gerd Jurgens, University of Tubingen

Japan

Arabidopsis research is carried out at more than 20 laboratories at national, prefectural, and private universities; national institutes; and some companies. The research topics explored include mutational analyses of flower and leaf development, responses to physical and chemical stimuli in roots or stems, root morphology, virus infection, hormonal regulation, transcriptional regulation, and shoot and root regeneration from calli. Also under study are the isolation and characterization of genes involved in the heat-shock response, desiccation, lipid biosynthesis, signaling pathways, protein phosphorylation, and transcription. The molecular function of isolated genes is examined using transgenic plants. In addition, efforts focus on developing experimental techniques, including Agrobacterium-mediated transformation and a new PCR-based method for isolating the flanking region of inserted T-DNA from transformants

Several leading *Arabidopsis* studies were summarized in a special issue of the Japanese journal Shokubutsu Saibo Kogaku (Vol. 6, No. 3). The Japanese *Arabidopsis* seed stock center was established in 1993 at the Miyagi College of Education, Sendai. The center has seed collections donated by Albert Kranz and Nobuharu Goto; seeds can be ordered by mail from the center.

Arabidopsis study descriptions and findings were presented at various symposia, including the annual meetings of the Genetic Society of Japan, the Japanese Society of Molecular Biology, and the Japanese Society of Plant Physiologists. Also, the Fourth Workshop on *Arabidopsis* Studies, organized by Kiyotaka Okada and Yoshiro Shimura (National Institute of Basic Biology -- NIBB) was held at NIBB, December 3-4, 1993, with more than 90 attending. *Arabidopsis* databases and communication using e-mail were demonstrated. The Fifth Workshop was held November 25-26, 1994, at NIBB.

Contact: Kiyotaka Okada, NIBB, Okazaki The Netherlands

The Netherlands

In January 1994, research groups working at the Universities of Amsterdam, Leiden, Utrecht, and Wageningen; and at the Dutch Center for Plant Breeding and Reproduction Research (CPRO-DLO) in Wageningen established an *Arabidopsis* network for The Netherlands. The group, called ARANED, is chaired by Maarten Koornneef (Genetics Department, Agricultural University, Wageningen); it will organize common research activities such as joint mutant screens and exchange of DNA from recombinant inbred lines. Also, the group will promote the exchange of information, materials, and experience through annual meetings. An ARANED research proposal to study embryogenesis and seed development was submitted to the Dutch National Science Foundation. Currently, *Arabidopsis* research in The Netherlands focuses on development, signal transduction -- especially related to hormones and light -- and plant-pathogen interactions.

Contact: Maarten Koornneef, Wageningen

Republic of Korea

Approximately 10 laboratories are currently involved in *Arabidopsis* research. Subjects under study include the mutational and molecular analysis of leaf senescence and photomorphogenesis at Pohang University of

Science and Technology (POSTECH); regulation of wound-inducible genes at Chonnam National University; function and regulation of drought-inducible and other protein kinase genes at Kyoungsang National University and POSTECH; biochemical study of photosignal transduction at Kyungbuk National University; protoplast culture at Chosun University; and T-DNA insertional mutagenesis at Kyungbuk National University and POSTECH.

A few other laboratories are also beginning to get involved in *Arabidopsis* studies. Additionally, there is a cDNA sequencing program on Brassica plants (Brassica napus and Brassica campestris), which may complement the *Arabidopsis* cDNA project. The current number of sequences obtained by the Plant Molecular Biology and Biotechnology Center in Jinju and the Agricultural Genetic Engineering Center in Suwon is nearing 3,500. All of these studies are primarily funded by the Ministry of Education, the Korean Science and Engineering Foundation, the Plant Molecular Biology and Biotechnology Center, the Department of Agriculture, and POSTECH.

Contact: Hong Gil Nam, POSTECH

Spain

The number of laboratories using *Arabidopsis* in Spain is growing. Up to 25 laboratories have expressed an interest in participating in the *Arabidopsis* Spanish network, and more than two-thirds of these already have projects which use this species. These projects are funded by national and European agencies. Most of the laboratories involved are at institutes of CSIC, the Spanish National Research Council; at universities in Madrid, Barcelona, and Valencia; and at INIA, the Spanish Institute for Agricultural Research in Madrid.

After a meeting held in 1993 in Valencia, efforts began which involved joint use of recombinant inbred lines and the production of additional T-DNA tagged lines. And, after the November 1994 meeting of the Spanish network in Madrid, efforts involving joint generation and screening of 10,000 T-DNA tagged lines commenced; this project constitutes a main objective for the next year.

Contact: Jose Martinez Zapater, Centro de Investigacion y Tecnologia, Madrid

United Kingdom

Overall, *Arabidopsis* is a well-established and widely used tool in the United Kingdom for innovative research in fundamental and applied science.

The United Kingdom's 4-year Plant Molecular Biology II (PMBII) Program -- which includes about 20 grants from the Biotechnology and Biological Sciences Research Council (BBSRC) to universities and research institutes for *Arabidopsis* research -- is now half complete. (The BBSRC was formed in April 1994 by a merger of the old funding councils, the Agriculture and Food Research Council and part of the Science and Engineering Research Council.) The mid-term meeting of all PMBII participants was held at the University of East Anglia, July 6-8, 1994. At this meeting, participants reported significant advances in the understanding of floral induction, plant hormones, plant development, and plant-pathogen interactions.

In addition to research sponsored under this coordinated program, BBSRC has made numerous other grants for *Arabidopsis* research. Beginning in 1994, further support for *Arabidopsis* genome research was made available as part of a new £2 million, 3-year BBSRC program on Plant and Animal Genome Analysis. Seven of the 14 funded proposals concerned *Arabidopsis*; these include an effort to strengthen the physical mapping effort, one to start site-selected transposon mutagenesis, and one to assess the colinearity of the *Arabidopsis* and Brassica genomes. Proposals for phase two of this program are now being considered.

BBSRC support for *Arabidopsis* researchers in the United Kingdom also includes funding for the Nottingham *Arabidopsis* Center. (Further details of the stock center's activities and

collaborations are provided in the previous section.) Also, EC-funded *Arabidopsis* research in the United Kingdom includes genome mapping and sequencing under the ESSA program.

Contact: Bernard Mulligan, University of Nottingham

United States

The North American *Arabidopsis* Science Steering Committee (NAASC) convened a workshop in June 1994 on *Arabidopsis* genome sequencing. The workshop report proposes establishing a program to sequence the entire *Arabidopsis* genome by the year 2004. The proposal was endorsed by the Multinational Science Steering Committee at its June meeting in Amsterdam. It was also widely circulated via electronic bulletin board, and has subsequently been used to present the views of the *Arabidopsis* community to various U.S. funding agencies.

In terms of funding *Arabidopsis* research in the United States, the Department of Agriculture, Department of Energy, National Institutes of Health, and National Science Foundation together provided a total of approximately US$22 million for fiscal year (FY) 1993 (i.e., the period Oct. 1, 1992, to Sept. 30, 1993) for research and related activities involving *Arabidopsis* as an experimental material. (These four agencies had signed in June 1990 an interagency agreement to collaborate on *Arabidopsis* research in the United States.) These funds supported approximately 180 individual research projects and about 35 individual post doctoral research fellowships, in addition to resource centers; tools and technology development projects; databases; conferences, workshops, courses, and graduate training programs; and equipment and facilities. The areas of research supported cover all aspects of plant biology from population biology and genetics to cell/developmental biology and physiology to biochemistry. Federal funding totaled approximately US$19 million in FY 1992, US$15 million in FY 1991, and US$7.5 million in FY 1990.

The *Arabidopsis* Biological Resource Center at Ohio State University has a mailing list of 1,225 U.S. individuals (out of a total of 2,102 on the list) in 47 States; this indicates the widespread use of *Arabidopsis* as an experimental system for plant biology research and teaching in the United States.

Contact: David Meinke, Oklahoma State University

Hybrid Development Flowers with *Arabidopsis*

Hybrid Development Flowers with Arabidopsis

Managing the reproductive lives of crop plants can be a major challenge. Because many plants, such as corn, benefit from hybrid vigor, growers try to avoid inbreeding , and try instead to have plants exchange pollen with others. However, many flowers are built so that the male anthers are so close to the female pistils that it is hard to avoid self-fertilization. Since the demand for hybrids is so great, many growers make exceptional efforts to avoid inbreeding. For example, for many years, the producers of hybrid seed corn detasseled millions of acres of corn plants by hand - a costly and time-consuming procedure.

Because hybrids are so desirable and the techniques for making them can be so difficult, researchers have long sought an easier way to breed hybrids. One strategy is to breed plants to have a trait known as cytoplasmic male sterility, where plants are unable to make pollen and, therefore, must be fertilized by pollen from other plants. This type of male sterility has the unusual characteristic of being determined by genes found on the mitochondria in the cytoplasm, instead of genes found in the nucleus.

George Pelletier, of the Institut National de Recherche Agronomique in Versailles, France, began his search for new ways of using cytoplasmic male sterility by studying plants such as tobacco and oilseed rape. A major achievement was the development of new techniques for manipulating the cells of such crop plants to gain a better understanding of the role of the genome in male sterility.

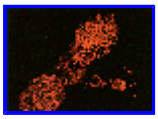
In 1992, however, Pelletier expanded his research to include work with *Arabidopsis. "Arabidopsis thaliana* appeared to us as an excellent material for studying specific functions of the male gametophytes [sex cells] due to the possibility of recovering mutants very easily [and] specific genes involved in pollen formation and plant reproduction," he says. "The information obtained from *A. thaliana* will be of high interest to crop improvement. It is already true for genes involved in lipid metabolism, and we believe it will be confirmed in plant development."

Concerning plant reproduction and the challenge of growing hybrids with greater ease, Pelletier adds, "We think that new systems of pollination control could be produced from promoters, and genes isolated from *A. thaliana*."

Such systems borrowed from *Arabidopsis* may accelerate the production of hybrids - and thereby enhance worldwide agricultural production.

Status of National Programs

Making Plastics (and Other Fine Things) from Plants



One of the major problems facing the developed world is the need to find renewable substitutes for declining stores of coal and oil, which supply energy and the raw materials for plastics, nylons, and other industrial products. Another major problem is to develop more healthful foods, ones that are broadly available and reasonably priced. In the past few years, *Arabidopsis* researchers have made several discoveries that should go a long way to meeting both of these challenges.

An important first step in designing better foods was made a few years ago when Chris Somerville and coworkers, then at Michigan State University in East Lansing and now at the Carnegie Institution of Washington at Stanford, California, found and cloned the *Arabidopsis* desaturase gene which codes for an enzyme that catalyzes the synthesis of polyunsaturated fatty acids. Dietary polyunsaturated acids have a role in lowering blood cholesterol and are needed for normal human growth.

Since the discovery of this gene in l992, scientists working with Somerville, including John Browse at Washington State University in Pullman and others at DuPont in Wilmington, Delaware, and at Monsanto in St. Louis, have isolated most of the eight different desaturase genes from *Arabidopsis* that control the polyunsaturation of plant oils. Plant breeders have then used these *Arabidopsis* genes to isolate the corresponding genes from crop species. What's more, they put copies of these genes into some plants, such as soybeans, canola, and flax, that typically make more saturated oils. This resulted in the production of nutritionally improved oils in these crops. These genetically transformed plants were field tested in the fall of l994, only 2 years after the key discoveries were made. "This gives an indication of how quickly industry can move to apply basic research," says Somerville. "We think we can tailor plant oils to specific nutritional needs." Somerville adds that "Most of the work that led up to this came from basic studies of *Arabidopsis*."

Related research focuses on developing custom-designed plants that will reduce our reliance on nonrenewable sources of petrochemicals, which are needed to make plastics and related materials. The goal is to modify plants genetically so that instead of producing edible oils, the plants will produce industrial oils and polymers, which are now made from petroleum stocks.

Somerville and his colleagues have shown that it is possible to genetically engineer *Arabidopsis* to produce granules of polyhydroxybutyrate (PHB), a polyester used for biodegradable plastic containers which is usually obtained from a bacterium, Alcaligenes eutrophus. The researchers did this by taking two genes the bacterium uses to make PHB and putting them into *Arabidopsis*. At first, the genetically transformed plants produced only minute granules of this valuable plastic. However, after a few years of tinkering, researchers inserted modified genetic constructs into the plants and could increase production enough to attract considerable commercial interest.

No one is suggesting that tiny *Arabidopsis* plants be developed as a commercial crop for plastics production, but whatever is learned from the *Arabidopsis* model can be used with other, more practical, crops, such as canola or soybean. Ganesh Kishore of Monsanto in St. Louis, which is one of several major companies studying the commercial possibilities of this system, says, "Our goal is to produce biodegradable plastic from a renewable source. We will create a whole new paradigm for the plastics industry."

Somerville adds that the production of plastics in plants will have the added benefit of giving farmers new markets for their harvests. "American agriculture produces too much of too few products. And novel plant varieties, created with the aid of *Arabidopsis* genes, will give farmers new cash crops," he says. What's more, he suggests that plastics are just the beginning, and that croplands of genetically transformed plants may be

devoted to the production of, for example, hydraulic oils, lubricants, nylons, drugs, and valuable enzymes.

Analysis and Recommendations for the Coming Year

Analysis and Recommendations for the Coming Year

By the end of the fourth year of the Multinational Coordinated *Arabidopsis thaliana* Genome Research Project, dramatic progress has been made in six prime areas of effort described in the original long-range plan:

- Genome analysis
- Technology development
- Biological resource centers
- Informatics
- Human resource development
- Workshops and symposia

The specifics of this progress are presented in the preceding sections of this report. This section presents an analysis of the progress and describes plans to pursue this progress into the immediate future.

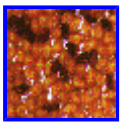
Genome Analysis and Technology Development

In the area of genome analysis -- which includes physical and genetic mapping, DNA sequencing, and functional analysis of mutated and cloned genes -- progress has been excellent. The physical map now covers a high proportion of the genome, with large chromosomal regions, such as almost all of chromosome IV, covered in fine detail. New yeast artificial chromosome (YAC) libraries containing very large genomic DNA inserts have been produced in the past year, adding to the several other YAC libraries that have been made available in previous years. P1 libraries have been made and are now being prepared for general distribution by the stock centers. Although still incomplete, the linkage of the *Arabidopsis* genome using YACs, cosmids, and now, P1 clones, has been sufficient to increase the numbers of genes cloned by map-based methods.

An important goal for 1995 is to complete the physical linkage of YACs so an ordered set of YACs spanning the entire genome can be generated. This set of YACs will allow immediate mapping of any DNA probe and greatly ease chromosome walking. Constructing a library of *Arabidopsis* DNA in a bacterial artificial chromosome vector -- which, like P1 and YAC libraries, contains large genomic DNA fragments but which also easily facilitates the use of Escherichia coli colonies for screening and amplification -- is another goal for the program's next year.

In complementary DNA (cDNA) sequencing, one goal for the coming year is to add another 10,000 expressed sequence tags (ESTs) to the more than 8,000 cDNA *Arabidopsis* ESTs already in the database. Another goal is to map 500 of these ESTs to YACs or to chromosome locations using restriction fragment length polymorphism, simple sequence length polymorphism, or cleaved amplified polymorphic sequence technology. Meeting this goal will provide a set of mapped markers located at high density throughout the genome.

The major new initiative for the coming year is to begin the systematic sequencing of the nuclear genome. In both the European Community and the United States, this will involve preparing stretches of overlapping clones ready for sequencing as well as starting large-scale sequencing projects at several sites. This genomic sequencing is a major initiative and is in accord with the original goal of completing the entire sequence of the *Arabidopsis* genome. The effort has just begun in Europe: Starting a complementary effort in the United States is a high priority.



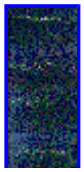
Progress has been remarkable in the areas of cloning genes known through their mutant phenotypes, isolating new and revealing mutations, and understanding the functions and interactions of genes. As described elsewhere in this report, the basic processes of all higher plants -- such as flower and root development, hormone action, and disease resistance -- are becoming better understood as a result of *Arabidopsis* research. This work should continue in 1995, with new emphasis on collections of interacting genes, so that a complete description of the development and physiology of higher plants can be achieved.

A new area of concentration should be subcellular function, so that the cellular basis of plant processes can be understood at the level of gene and gene products. The problem of cloning genes, once mutant forms are available, has been solved, though faster methods are still needed. However, the problem of learning the function of a gene when only its sequence is available remains to be solved. One of the major scientific aims of the program therefore must be to develop a gene knockout system, which allows the identification of gene function from gene sequence. Together, sequencing and functional data will allow modification of plant function in new and useful ways.

Biological Resource Centers

The seed and clone stock centers need expanded capability and space. They now provide a critical service to the *Arabidopsis* community in distributing biological materials. However, they need to grow in order to cope with the ever-increasing demands from both *Arabidopsis* and other plant scientists for the many new ESTs, mutants, specific cloned genes and DNA libraries now available. Any expansion will need a concomitant increase in staffing support. Renewed effort must be placed on finding sources for long-term support for the stock centers and their critical mission. One new initiative, an electronic *Arabidopsis* newsletter called Weeds World, has been started by Mary Anderson of the Nottingham *Arabidopsis* Stock Centre at the University of Nottingham, United Kingdom. Efforts to make the excellent stock centers more accessible, and to make them sources of new information and research materials, should be supported.

Informatics



With ever-increasing amounts of data being produced by the sequencing, mapping, and positional cloning projects, the database network needs to be upgraded. This need is a measure of the success

of the Arabidopsis project itself. The goal for the coming year is to move the existing database system to a new generation of databases, with appropriate curatorial support. The Multinational Science Steering Committee has endorsed this assessment, and a proposal for the next generation *Arabidopsis* database has been submitted to the National Science Foundation for review. One of the objectives of the new database is to collect all available information on YAC positions, mutant locations, and EST mapping. Sharing of these data will benefit all *Arabidopsis* researchers by making new information available before publication, presenting data organized in a way that avoids duplication of effort, and accommodating completion of the physical linkage map.

Human Resources

A key goal in the human resources area is to find support for multinational postdoctoral fellowships, short-term exchanges, and short courses. While the Cold Spring Harbor *Arabidopsis* course is now well-established, and other courses have been given, there is little support for multinational postdoctoral fellowships. Training young researchers and coordinating international efforts are the very future of *Arabidopsis* genome work. Both improved training and coordination result from researchers being able to study and work in different *Arabidopsis* laboratories at early stages of their careers. At present, there is inadequate support for this key area. A goal for 1995 is thus to establish new programs for international postdoctoral fellowships.

Workshops and Symposia

The *Arabidopsis* community has grown to a size and achieved a rate of scientific progress that require an *Arabidopsis* meeting each year. Thus, at the June 1994 meeting of the Multinational Science Steering Committee, plans for future *Arabidopsis* research conferences were announced. Specifically, conferences will be held June 7-11, 1995, at the University of Wisconsin in Madison (the contact is Rick Amasino, Department of Biochemistry, University of Wisconsin); June 24-28, 1996, in Norwich, United Kingdom; and the second week of June 1 997 at the University of Wisconsin. These international meetings, devoted solely to *Arabidopsis* research, should continue to be supported so that they can be attended by all *Arabidopsis* researchers, and thereby serve a coordinating function for the *Arabidopsis* genome project.

Arabidopsis is now such a focus for plant science researchers in such diverse fields as developmental biology, signal transduction, and plant-pathogen interactions, that many genes are being first identified in *Arabidopsis*. The Multinational Coordinated *Arabidopsis thaliana* Genome Research Project has supported many of these advances. Because the project is international and collaborative, it has provided an essential framework for hundreds of researchers. The development of an electronic mail network; various databases; and specialized symposia, meetings, and biological resource centers have fostered an international spirit of cooperation and the free exchange of ideas. As a result, studies of *Arabidopsis* have accelerated, greatly enhancing our understanding of the molecular basis of all plant processes.

Our overall goal for the coming year should be to press ahead with the specific objectives of the project, with the ultimate aim of fully understanding the genetic, molecular, and cellular basis of all plant life.

Appendices

Appendix A. Sequencing Proposal

North American *Arabidopsis* Steering Committee Workshop Proposal For An *Arabidopsis Thaliana* Genome Project (ATGP)

Executive Summary

Barriers that once impeded the identification of genes with important biological functions are vanishing in the 1990's. High through-put genomic sequencing now permits the rapid identification of large numbers of genes that were previously inaccessible to traditional genetic analysis. The rate of gene discovery is now limited only by the ability of scientists to map and sequence an organism's genome. Initial sequencing efforts have focused on so-called model organisms with relatively small genomes. The information obtained from these model genomes is being used to understand gene structure and function in related organisms.

Arabidopsis thaliana, a small flowering plant in the crucifer family, has the smallest genome and the highest gene density so far identified in a flowering plant. During the past ten years, *Arabidopsis* has become established world-wide as the preferred species for molecular-genetic studies in the laboratory. Importantly, because cloned *Arabidopsis* genes can be used to identify corresponding genes in all other plants, continued progress in identifying *Arabidopsis* genes should be considered an important strategic component for maintaining the U.S. preeminence in plant biology. Genes identified in *Arabidopsis* will soon lead to the creation of economically important plants that are more resistant to pathogen attack, that reduce the use of environmentally toxic chemicals, that produce foodstuffs with improved nutritional value, or that yield new kinds of compounds of commercial value. Increasing our knowledge of plant genes has almost limitless potential to improve environmental quality, increase energy production, identify new medicinal compounds, and enhance our ability to respond to the steady increase in human population and changing climatic conditions.

This report contains the recommendations of an ad hoc committee representing the community of *Arabidopsis* researchers and various government agencies that met in Arlington Virginia on June 8 and 9, 1994, to discuss the feasibility of commencing a federally-funded large scale *Arabidopsis* genome project in the United States. The committee discussed the impact that an *Arabidopsis* genome project would have on

the progress of basic plant research as well as on the strategic interests of the United States as they relate to agriculture, energy and the environment. The committee concluded that a large scale *Arabidopsis thaliana* Genome Project (ATGP) should commence as soon as possible. The committee identified essential features that should be considered in any proposals for the initiation of an ATGP. The committee concluded that one or a limited number of linked *Arabidopsis* Genome Centers should be established and that these Centers will serve as important models for other plant genome projects in the future. Finally, the Committee recommended that the United States ATGP be coordinated with a similar effort already underway under the auspices of the European Community. The committee recommended that funds be provided for:

1. Completion of the *Arabidopsis* physical/genetic map and the creation of sequence-ready clone collections by 1997.
2. Pilot sequencing and technology development projects with the goal of completing 10 megabases of *Arabidopsis* genome sequence by 1999.
3. Subsequent scale-up of pilot projects and complete sequencing of the 100 megabase *Arabidopsis* genome by 2004.

Introduction

The general benefits of genome sequencing are increasingly obvious as rapid progress is made toward the goal of sequencing complete chromosomes in other model organisms, such as yeast and Caenorhabditis elegans (a small nematode). While classical mutagenesis, genetic analysis and conventional cloning strategies have uncovered many genes, rough estimates suggest that no more than 20-25% of an organism's genes can be identified by classical genetic techniques, even in organisms with a small fraction of redundant genes. Plants, including *Arabidopsis*, generally exhibit a moderate to considerable redundancy of half or more of their genes. For this and other reasons, mutations that interfere with or eliminate expression of many genes are silent. Thus direct genome sequencing is the only sure way of identifying all of an organisms genes. For plants, it follows that genome sequencing will be required for the identification of most of the economically important genes.

Because genome sequencing projects are still relatively expensive, model organisms have been selected as the initial targets of complete genome sequencing. The evolutionary kinships among organisms justify this approach. Depending on the function of a gene and how well conserved its sequence in evolution, at the very least the gene sequences of the model organism can be used to identify corresponding genes in related species. Thus the selection of model organisms for full genome sequencing is a reasonable policy for conserving limited resources, while maximizing information yield. Model organisms have been chosen by several criteria, including the breadth of existing genetic information, small genome size, and high gene density. *Arabidopsis thaliana* was adopted as a model organism by plant geneticists some years ago because of its small genome size and rapid reproductive cycle. At 100 megabases, the *Arabidopsis* genome is among the smallest known plant genomes. It also has a low repetitive DNA content.

The wisdom of selecting *Arabidopsis* as a model organism for higher plants is becoming increasingly obvious. Initial sequencing efforts suggest that the *Arabidopsis* genome has a very high gene density (~ 1 gene every 5 kb). The relatively close relationship among higher plants due to the fact that they evolved relatively recently in evolutionary time makes it possible to use sequence information obtained from *Arabidopsis* to identify homologous genes in other plants, including agronomically important species with much larger genome sizes, higher gene redundancy and a substantially greater content of repetitive sequences. *Arabidopsis* genes, which are often much easier to clone initially than the corresponding genes of plants with larger genomes, have already been used to identify and manipulate genes in agronomically important species. Scientists at Dupont, for example, have used *Arabidopsis* genes as probes to clone fatty acid desaturase genes from a variety of oilseed species such as soybean and canola. The cloned genes have been modified and reintroduced into the species of origin to alter the composition of the oil for improved health benefits. *Arabidopsis* has also been the initial experimental organism for the introduction of bacterial genes that permit genetically engineered plants to synthesize a biodegradable thermoplastic, polyhydroxybutyrate. The gene system was subsequently transferred to plants that can be used to produce the

plastic on an agricultural scale. It was the ready availability of *Arabidopsis* mutants, as well as the fact that *Arabidopsis* can be genetically manipulated that made this work possible. Additional genes which have been cloned from *Arabidopsis* and which have potential agronomic value include genes that confer resistance to bacterial and fungal pathogens, which are involved in the synthesis of plant hormones, which affect nutritional quality of seeds, and which alter time of flowering.

In addition to its potential agronomic importance, *Arabidopsis* genome mapping and sequencing work has already benefited and will increasingly benefit the community of *Arabidopsis* researchers. It presently takes about three person-years on average to clone an *Arabidopsis* gene identified by a mutation using map-based cloning techniques. The availability of the complete genomic sequence would vastly simplify and reduce the cost of identifying most *Arabidopsis* genes. Although the short-term cost of sequencing the entire *Arabidopsis* genome is substantial (current costs are about $1.00/base, implying a total cost approaching $100 million by project completion), there are long-term savings and benefits for the entire plant research community in accelerating research. Moreover, the high gene density of the *Arabidopsis* genome implies a high ratio of informative to uninformative sequence, maximizing the return on the investment of time and resources. Finally, the information obtained in sequencing the genomes of other model organisms widely used in biological research, such as Escherichia coli, yeast, and C. elegans has contributed greatly to our understanding of the biology of these organisms and clearly demonstrates the important role that genome projects can play in biological research. Equally significant advances in our understanding of plant biology can be expected from an *Arabidopsis* genome project.

Workshop Summary

Overview: To assess the feasibility and desirability of a federally funded *Arabidopsis* genome project, the North American *Arabidopsis* Steering Committee organized and convened a workshop in Arlington Virginia on June 8-9, 1994. The workshop participants included the elected members of the North American *Arabidopsis* Steering Committee. Representatives from the National Science Foundation, the U.S. Department of Agriculture, the Department of Energy, the NIH-sponsored human genome project, and the European Community were present as observers. Two scientists involved with the human genome project were also present as technical advisors. A list of participants is given below.

The general goal of the workshop was to assess progress toward meeting the goals of mapping and sequencing the *Arabidopsis* genome and make specific recommendations to the National Science Foundation to direct future US efforts in the Multinational Coordinated *Arabidopsis thaliana* Genome Research Project. A secondary goal was to outline in general terms the main issues which should be addressed in future proposals concerning the development of new or expanded *Arabidopsis* sequencing centers.

The workshop commenced with a summary of the recent *Arabidopsis* genome conference held at the Cold Spring Harbor's Banbury Center and discussion of current funding for Arabidopsis genome research. Mike Bevan and Chris Somerville presented overviews of the EC sequencing program (ESSA or European Scientists Sequencing *Arabidopsis*) and the Michigan State University cDNA sequencing project, respectively. Mary Clutter joined the workshop participants for a brief discussion of *Arabidopsis* genome research funding within NSF. Jen-i Mao and Mark Johnston discussed two different approaches to sequencing taken at Collaborative Research (the multiplex approach) and by the C. elegans sequencing group at Washington University (sequencing machines). The committee discussed the responses from the *Arabidopsis* community to a questionnaire on the *Arabidopsis* genome project. Finally, workshop participants discussed the present status and future of the US *Arabidopsis* genome project, commencing with a detailed consideration of the rationale for genome mapping and sequencing and commentary on the benefits of even the limited effort to date. The following issues were discussed in depth: Should there be an organized Arabidopsis genome project given the current state of *Arabidopsis* research? What is the relative priority of complete genome sequencing compared to completion of a physical map, adding more PCR-based mapping markers to the map, or single-pass cDNA sequencing? Who should pay for an *Arabidopsis* genome project, how should it be organized, how long will it take, and how much will it cost? How will a US-funded ATGP be coordinated with ESSA?

Setting Priorities: Before the workshop, a questionnaire designed to obtain feedback from the *Arabidopsis* community on the desirability of an *Arabidopsis* genome project was posted on the *Arabidopsis* electronic newsgroup, More than 20 responses were obtained which were reviewed and discussed during the course of the workshop. Although most respondents supported the concept of an ATGP, several respondents suggested that a high-density genetic map consisting of PCR-based markers be completed before large scale sequencing be undertaken. Indeed, the relatively small number of DNA markers and the incomplete physical map had already been useful to many investigators and that there had been extremely heavy and immediate demand for the cDNA clones that were being sequenced at MSU and in France. The workshop participants agreed that a high-density genetic/physical map would be of immediate benefit to the community. On the other hand, because it takes considerable time to get a sequencing organization equipped, trained and functioning efficiently, there was general agreement of workshop participants that it is essential to begin setting genome sequencing goals immediately and to initiate pilot sequencing projects in parallel with other aspects of genome analysis.

Progress in Genome Research: The current efforts in several laboratories to establish links between the genetic and physical maps of the *Arabidopsis* genome greatly facilitates the map-based cloning of genes. While many mutations and genes have been mapped by the use of restriction fragment length polymorphism (RFLPs), genetic markers based on the polymerase chain reaction (PCR) are being developed for *Arabidopsis*. Cleaved amplified polymorphic sequences (CAPS) and simple sequence length polymorphism (SSLPs) markers can be used for rapid mapping of plant mutations and as a dense set of sequence tagged sites (STSs) for the construction of a physical map of the *Arabidopsis* genome using an anchoring strategy. In a collaborative effort, investigators at the John Innes Institute, the University of Pennsylvania and Massachusetts General Hospital, are developing an overlapping set of YACs covering the entire genome. Using newly available YAC libraries, total genome coverage in YACs is now estimated to be approximately 60-70%; with even greater coverage on chromosome 4 (about 80%).Furthermore, in preparation for phase one of the European Scientists Sequencing *Arabidopsis* (ESSA), restriction mapping of 500 kb of cosmids from the top of chromosome 4 has been completed and distributed to the participating laboratories. In addition to facilitating the cloning of genes identified solely by phenotype, physical mapping of the genome generates the starting materials for rapid and efficient sequencing and is a key component of a genome project.

Another important component of the ATGP is three cDNA sequencing projects that are underway in Europe, Canada and the US. The European goal is to sequence (from both ends) 3000 unique cDNA fragments (expressed sequence tags or ESTs). ESSA scientists are also mapping their ESTs to YAC clones, regardless of whether the YAC clone has been anchored. Canadian scientists are planning to map 600 ESTs. The US project has already entered 2500 ESTs in publicly available databases and is on the verge of entering an additional 4000 (these have been sequenced only in one direction and relatively little effort has been devoted to eliminating redundancy). The exact number of different gene transcripts represented among this collection of ESTs is unknown; hence the fraction of the estimated 15Ð16,000 *Arabidopsis* genes represented in this collection cannot be determined at present. The workshop participants concluded that mapping cDNAs had merit because it facilitates connecting a mapped mutation to its cognate gene even in the absence of genomic sequence.

Goals for *Arabidopsis* Genome Research: Workshop participants agreed that a pilot genome sequencing project should begin immediately. More specifically, the NAASC recommends that a specific federal program be developed to support Arabidopsis genome sequencing and associated technology development with the goal of completion of the entire genomic sequence by the year 2004. The following steps should be undertaken to achieve this goal:

1. A call for proposals to conduct pilot *Arabidopsis* sequencing projects. This should be in the form of RFPs to make it possible to attract proposals from outside the *Arabidopsis* community.
2. Establishment of several sequencing centers with the short-term goal to obtain 10 megabases of genomic sequence within 3 years from the start of funding (a similar goal to ESSA). The participation of existing DNA sequencing centers, as well as companies with relevant expertise, is encouraged. The

purpose of these pilot projects will be to establish the feasibility of and to develop a detailed strategy to complete the sequencing of the entire *Arabidopsis* genome. To achieve cost-effectiveness, it is not envisioned that this program will fund a large number of small-scale sequencing projects. Pilot sequencing projects should include substantial mapping components, including the goal of finding and mapping at least 1000 PCR-based markers, to generate the appropriate templates for sequencing.

3. Significant expansion of the pilot sequencing centers to achieve the goal of completion of the entire sequence by 2004. It is noted that this phase will require a substantial commitment of equipment, supplies, and personnel.
4. Although specific goals were not set, workshop participants emphasized that a key feature of genome research was the development of methods for the identification of gene function. Some of the more promising methodologies for *Arabidopsis* include antisense mRNA constructs, co-suppression and transposon tagging.

Conference Participants

Dr. Fred Ausubel
Department of Molecular Biology
Massachusetts General
Hospital
Boston, MA 02114

Dr. Mike Bevan
John Innes Institute
Colney Lane, Norwich
NR4 7UJ, United Kingdom

Dr. Joanne Chory
Plant Biology Laboratory
Salk Institute
PO Box 85800
San Diego, CA 92186-5800

Dr. Joseph Ecker
Department of Biology
University of Pennsylvania
Philadelphia, PA 19104-6018

Dr. Mark Estelle
Department of Biology
Indiana University
Bloomington, IN 47405

Dr. Nina Fedoroff
Department of Embryology
Carnegie Institution of Washington
115 West University Parkway
Baltimore, MD 21210

Dr. Howard Goodman
Department of Molecular Biology
Massachusetts General Hospital
Boston, MA 02114

Dr. Mark Johnston

Department of Genetics
Washington University School of Medicine
4566 Scott Avenue
St. Louis, MO 63110-1031

Dr. Jen-i Mao
Collaborative Research Inc.
1365 Main Street
Waltham, MA 02154

Dr. David Meinke
Department of Botany
Oklahoma State University
Stillwater, OK 74078

Dr. Chris Somerville
Plant Biology Department
Carnegie Institution of Washington
Stanford, CA 94305-4170

Observers:

Dr. Machi Dilworth
National Science Foundation
4201 Wilson Blvd, Rm #685
Arlington, VA 22230

Dr. Ed Kalaikau
Program Director
Plant Genome Program
National Research Initiative Competitive Grants
CRS/USDA
AG Box 2241
Washington, DC 20250-2241

Dr. Jerome P. Miksche
Director
Office of Plant Genome Research
Agricultural Research Service
U.S. Department of Agriculture
Bldg. 005, BARC-West,
Beltsville, MD 20705

Dr. Robert Rabson
Director
Division of Energy Biosciences
Office of Basic Energy Sciences
U. S. Department of Energy
ER-17, GTN
Washington, DC 20545

Dr. Robert Strausberg
National Center for Human Genome Research
National Institutes of Health

9000 Rockville Pike
Bethesda, MD 20892

Appendix B. RFLP Map

Chromosome I				Chromosome II		Chromosome III			
(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
0.0	RS10	66.1	mi342	0.0	NOR2	0.0	Athb3	112.2	m424
2.4	nga59	68.5	mi133	7.2	ve012	2.5	mi199	113.0	agp29
2.4	PVV4	68.5	GAPB	9.3	mi320	3.5	mi74b	114.8	nga6
2.4	ACC2	70.9	mi72	10.8	m246	3.5	nga32	117.1	nga112
2.4	ve001	70.9	GRF2	13.4	m497A	4.5	nga172		
6.2	ATEAT1	72.9	GRF4	14.3	g4553	6.0	GAPC		
7.3	agp16	80.2	GRF1	17.9	mi310	6.0	mi172		
8.5	O846A	85.3	mi291a	17.9	ve013	9.6	mi355		
9.7	SEP4E	86.4	mi441	18.5	mi444	9.6	mi403		
9.7	ve002	86.9	mi208	18.5	mi421	10.7	m583		
10.3	mi372	87.4	mi106	19.1	g4532	11.4	g4523		
10.3	g4715a	89.0	m213	21.8	SEP2A	12.9	nga126		
10.3	m488	91.7	spl5	22.9	g4133	13.6	JGB3		
10.3	ve003	94.4	mi209	25.9	PR1	14.9	mi467		
10.3	agp102	94.6	nga280	28.5	mi398	15.5	mi357		
10.3	ve004	94.8	nga128	32.6	m216	16.5	apx2		
12.1	PAI1	95.0	mi303	32.6	B33	19.0	mi207		
13.9	apx1	96.6	B34	34.3	mi139	19.3	MS2		
15.5	mi100	98.2	g4026	35.4	mi148	19.6	ATHCHIB		
16.0	m241A	98.7	mi259	38.1	mi238	19.6	spl6		
16.2	mi443	98.7	mi230	38.4	pGC1	20.8	nga162		
16.6	nga63	98.7	mi408	38.7	m251	24.9	g4708		
17.2	ve005	98.7	mi304	38.7	O8O2F	24.9	m228		
20.2	ve006	100.0	ATHGENE	45.1	g6842	25.4	mi289		
20.9	NCC1	102.7	mi324	45.1	GPA1	25.9	mi339		
20.9	pC1	102.7	mi353	48.0	er	30.8	m560B2		
21.6	pC2	102.7	MW1	48.0	B68	33.2	m105		
23.0	O818	104.5	mi424	54.1	pGC2	33.9	mi142		
23.5	EG17G9	108.6	m315	55.1	mi54	34.5	O4O23		
24.5	g3786	114.1	mi185	56.1	ve014	36.4	mi268		
25.5	ve007	114.1	mi193	56.6	m220	41.8	mi386		
28.1	mi348	116.2	g4552	58.9	SEP5B	42.9	mi225		
29.1	mi113	116.7	pCITf117	58.9	m283C	43.9	g4711		
31.5	pFNR	122.1	Athb13	63.3	PR21	50.1	mi178		
33.9	mi203	123.8	Tag1	65.2	ve015	50.1	mi287		
33.9	pC3	124.8	mi462	65.2	spl3	51.1	ve020		
35.3	g3829	125.8	mi103	65.7	mi277	51.1	um579C		
39.2	m235	126.3	PKNAT22	66.0	ve016	52.6	GAPA		
42.0	mi265	126.3	m453A	66.3	g17288	57.3	GL1		
42.5	mi163	129.4	nga111	66.3	m323	60.4	mi413		
45.1	ve008	131.7	mi425	68.0	um579B	63.6	mi79b		
45.1	mi111	131.7	PR5	68.0	ve017	64.1	mi358		
45.1	mi62	133.2	I6	68.3	SAM3	69.1	ve021		
45.1	mi15	134.0	ATHATPAS	68.6	ve018	72.1	g4117		
45.1	mi192	134.6	m532	70.4	LTP	72.1	m249		
46.1	mi116	135.0	ADH	71.4	m429	77.6	O97B1		
46.1	nga248	135.4	petE	71.4	g4514	84.7	I18		
56.2	ve009	135.8	ve011	72.9	O5841	86.1	g4564b		
56.8	m254A	136.9	agp64	72.9	nga168	89.2	g4014		
57.9	m253	140.1	g17311	78.4	m336	93.5	m457		
58.5	P39B2T7	141.1	m132A	81.0	ve019	94.6	mi456		
59.1	ve010	141.1	mi157	81.6	mi473	99.7	t94		

60.8	mi423a	150.4	pAtT32CX	83.7	Athb7	101.8	ve022
62.5	RPS18B			86.4	mi455	104.5	g2778
64.1	mi63			86.4	mi79a	104.5	BGL1
65.1	mi19			88.8	pAtT51	108.1	spl1

Chromosome IV					Chromosome V		
(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
0.0	BIO217	67.1	m226	0.0	O5629	77.3	mi323
3.7	mi51	67.1	g13683	3.2	um515D	78.4	m247
6.6	mi204	67.1	C18a	3.2	g3715	82.7	g4028
7.2	mi122	67.6	g4539	6.1	ctr1	85.9	DFR
10.4	g3843	68.1	mi112	8.0	mi121	87.0	spl2
13.7	I41G	68.1	mi330	8.0	nga225	88.2	mi194
16.6	mi301	68.6	g3845	9.0	ASA1	88.7	agp6
18.7	ve023	70.8	mi32	10.0	O6569	89.2	mi83
18.7	mi390	73.0	AG	10.5	m217	92.5	mi423b
20.5	GA1	73.0	g19838	11.5	g3837	93.1	mi61
22.4	Gslohp	73.9	pCITd71	13.0	mi97	96.3	ve027
25.3	petC	73.9	g3883	14.6	KG31	99.8	mi271
26.7	m448A	75.5	JGB9	16.5	nga158	100.3	mi226
.7	mi233	79.1	mi422	19.4	nga249	101.6	nga129
27.7	g2616	81.9	mi475	22.2	ca72	106.2	m435
28.2	mi306	83.0	agp66	22.7	nga151	106.3	LFY3
28.7	m506	83.5	SEP2B	22.7	mi174	111.8	m558A
29.5	BIO200	85.5	RPS2	23.3	CHS	112.9	mi70
30.3	mi167	86.5	m600	26.9	mi322	112.9	mi418
30.8	mi87	86.9	PG11	26.9	mi438	112.9	mi74a
30.8	m456A	87.3	mi123	26.9	nga106	112.9	mi184
33.4	nga12	88.4	mi232	28.9	ve026	114.0	mi69
36.1	nga8	88.4	RLK5	29.7	g4560	114.0	SEP5A
39.7	RPS18C	90.6	H1	30.5	TSL	115.3	agp50
40.2	pCITf3	92.8	g8300	32.2	mi138	118.9	m211A
41.7	H2761	93.9	mi431	35.6	mi90	125.4	g2368
41.7	m518A	94.6	O6455	36.1	mi433	126.5	ve028
47.1	BIO206	94.6	g3088	36.1	GslbutAr	132.6	m555
49.4	pCITd23	95.4	pCITd104	37.8	m291	134.9	mi335
52.6	g41 08	100.4	pCITd76	44.1	g4715b	138.2	BIO205
55.0	mi465	100.4	pCITd99	47.9	g455657		
.8	mi128	104.4	m214	48.8	nga139		
58.9	g6837	108.8	AP2	50.1	mi219		
58.9	g2620	114.7	g2486	50.1	AF3		
59.9	um713B1	115.8	um596A	52.6	Tn139		
59.9	mi279	119.4	ve025	57.9	mi125		
60.4	um713B2	124.4	mi369	62.3	um579D		
60.4	mi30	124.4	DHS1	64.0	nga76		
60.9	g10086	124.4	g3713	64.6	mi291b		
60.9	g4564a			67.5	mi137		
61.9	m326			68.5	GRF3		
61.9	ve024						
62.4	mi198						
64.5	mi260						

NOTES: RFLP map prepared by Clare Lister and Caroline Dean, John Innes Centre, Norwich, UK. Map was generated using recombinant inbred lines of *Arabidopsis thaliana* from a cross between Landsberg and Columbia ecotypes. MapMaker software provided the linkage analysis of RFLP scores; recombination frequencies were converted to map distances using the Kosambi mapping function. The maps of chromosomes I, II, III, and V were generated entirely using MapMaker. The map of chromosome IV was based on the physical map data from Renate Schmidt and coworkers. The columns under each chromosome name represent (1) the cumulative centimorgan distance along the chromosome and (2) the corresponding locus.

Appendix C. Brief History of the Multinational Coordinated *Arabidopsis thaliana* Genome Research Project

May 1989 First U.S. planning workshop (National Science Foundation, Washington, DC)

July 1989 Second U.S. planning workshop (Cold Spring Harbor, NY)

October 1989 International planning workshop (Bloomington, IN) UK Agricultural and Food Research Council (AFRC) establishes a coordinated program on *Arabidopsis* biology under the Plant Molecular Biology (PMB) research initiative

December 1989 Seed stock center at University of Nottingham, UK, is established

February 1990 First issue of the AFRC PMB *Arabidopsis* newsletter is published

April 1990 The Multinational Science Steering Committee meets in Denver, CO, and drafts a long-range plan for the Multinational Coordinated *Arabidopsis thaliana* Genome Research Project

May 1990 The European Community (EC) launches a transnational *Arabidopsis* genome research project ¾ a 3-year T-project under Biotechnology Research for Innovation, Development, and Growth in Europe (BRIDGE) aimed at technology development

June 1990 Participants of the Fourth International Conference on *Arabidopsis* Research in Vienna, Austria, endorse the long-range plan drafted by the steering committee in April In Washington, DC, the Department of Energy, the National Institutes of Health, the National Science Foundation (NSF), and the Department of Agriculture sign an interagency agreement to collaborate on *Arabidopsis* genome research

July 1990 A worldwide *Arabidopsis* electronic bulletin board is established

August 1990 Long-range plan for the Multinational Coordinated *Arabidopsis thaliana* Genome Research Project is published (NSF 90-80)

October 1990 NSF receives funding to begin *Arabidopsis* genome research initiative

February 1991 A DNA clone center is established in Koln, Germany, as part of the EC BRIDGE project

April 1991 First annual progress report for the Multinational Coordinated *Arabidopsis thaliana* Genome Research Project is published (NSF 91-60)

September 1991 French cDNA and mapping project begins An *Arabidopsis thaliana* Database (AAtDB) is established at Massachusetts General Hospital in Boston, MA *Arabidopsis* Biological Resource Center at Ohio State University and its associated database, *Arabidopsis* Information Management System (AIMS), are established

January 1992 EC-U.S. workshop on managing *Arabidopsis* genome data (Boston, MA)

March 1992 U.S. cDNA sequencing project at Michigan State University in East Lansing, MI, is initiated

August 1992 Second annual progress report for the Multinational Coordinated *Arabidopsis thaliana* Genome Research Project is published (NSF 92-112)

October 1992 Second phase of the AFRC PMB *Arabidopsis* program begins in UK

November 1992 First set of *Arabidopsis* expressed sequence tag data is entered into the Database of Expressed Sequence Tags (dbEST), a public database at the U.S. National Center for Biotechnology Information, Bethesda, MD

June 1993 Workshop on database needs for *Arabidopsis* genome research (Dallas, TX)

August 1993 Fifth International Conference on *Arabidopsis* research (Columbus, OH) Germany establishes a special research program, titled "*Arabidopsis* as a Model for the Genetic Analysis of Plant Development"

November 1993 Spain establishes an *Arabidopsis* research network

December 1993 European Scientists Sequencing *Arabidopsis* Project begins in EC Third annual progress report for the Multinational Coordinated *Arabidopsis thaliana* Genome Research Project is published (NSF 93-173) UK *Arabidopsis* electronic bulletin board is established

January 1994 ARANED, an *Arabidopsis* research group, is established in The Netherlands

March 1994 Banbary conference on issues related to a large-scale sequencing of *Arabidopsis* genome (Cold Spring Harbor, NY)

June 1994 Planning workshop for a coordinated *Arabidopsis* genome sequencing project (Arlington, VA)

July 1994 "End of BRIDGE - Beginning of ESSA" EC workshop (Cambridge, UK)

November 1994 First issue of Weeds World, an electronic worldwide newsletter for the Multinational Coordinated *Arabidopsis thaliana* Genome Research Project, is published

January 1995 EC announces plans for the Fourth Framework Program (1996-99), which aims to sequence 10 megabase of the genome and set up a systematic function search program for the *Arabidopsis* genome

Appendix D. Acknowledgments

The following people contributed information, reviews, or illustrative material to this report. Members of the Multinational Science Steering Committee are indebted to their many colleagues who offered their time and talent.

Mary Anderson, Nottingham University,
Nottingham, UK

Frederick Ausubel, Massachusetts General
Hospital, Boston, MA, USA

Barbara Baker, U.S. Department of Agriculture,
Plant Gene Expression Center, Albany, CA, USA

Philip Benfey, New York University,
New York, NY, USA

Michael Bevan, John Innes Centre, Norwich, UK

Barbara Buchanan, National Agricultural Library,
Beltsville, MD, USA

Sam Cartinhour, U.S. Department of Agriculture,
National Agricultural Library, Beltsville, MD, USA

J. Michael Cherry, Stanford University,
Stanford, CA, USA

Joanne Chory, Salk Institute, La Jolla, CA, USA

George Coupland, John Innes Centre, Norwich, UK

Jeff Dangl, Max Planck Institute, Cologne, Germany

Keith Davis, Ohio State University,
Columbus, OH, USA

Caroline Dean, John Innes Centre, Norwich, UK

Michel Delseny, University of Perpignan,
Perpignan, France

Joseph Ecker, University of Pennsylvania,
Philadelphia, PA, USA

Mark Estelle, Indiana University,
Bloomington, IN, USA

Ellen V. Kearns, Massachusetts Institute of
Technology, Cambridge, MA, USA

Maarten Koornneef, Agricultural University,
Wageningen, The Netherlands

David Kristofferson, Intelligenetics,
Palo Alto, CA, USA

Peggy Lemaux, University of California,
Berkeley, Berkeley, CA, USA

Bertrand Lemieux, York University,
North York, Ontario, Canada

Hong Ma, Cold Spring Harbor Laboratory,
Cold Spring Harbor, NY, USA

Sheila McCormick, U.S. Department of Agriculture,

Plant Gene Expression Center, Albany, CA, USA

Susan McCarthy, U.S. Department of Agriculture, National Agricultural Library, Beltsville, MD, USA

Jon Monroe, James Madison University, Harrisonburg, VA, USA

John Morris, Massachusetts General Hospital, Boston, MA, USA

Hong Gil Nam, POSTECH, Kyungbuk, Republic of Korea

Thomas Newman, Michigan State University, East Lansing, MI, USA

Kiyotaka Okada, National Institute of Basic Biology, Okazaki, Japan

Georges Pelletier, Ministere de l'Agriculture-INRA, Versailles, France

Sakti Pramanik, Michigan State University, East Lansing, MI, USA

Daphne Preuss, University of Chicago, Chicago, IL, USA

Ernie Retzel, University of Minnesota, Minneapolis, MN, USA

Chris Somerville, Carnegie Institution of Washington, Stanford, CA, USA

Randy Scholl, Ohio State University, Columbus, OH, USA

Brian Staskawicz, University of California, Berkeley, CA, USA

The following individuals served as consultants during the preparation of this report.

Machi Dilworth, National Science Foundation, Arlington, VA, USA

Etienne Magnien, Commission of the European Communities, Brussels, Belgium

Jerome Miksche, U.S. Department of Agriculture, Beltsville, MD, USA

Jim Tavares, U.S. Department of Energy,

Washington, DC, USA

Anton Vassarotti, Commission of the European Communities, Brussels, Belgium