Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 10831–10835, November 1995

Review

The genome of *Arabidopsis thaliana*

Howard M. Goodman†‡, Joseph R. Ecker§, and Caroline Dean¶*

**Department of Genetics, Harvard Medical School, and †Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114; §Department of Biology, University of Pennsylvania, Philadelphia, PA 19104; and ¶Department of Molecular Genetics, Biotechnology and Biological Sciences Research Council, John Innes Centre, Colney Lane, Norwich, NR4 7UJ, United Kingdom*

ABSTRACT ***Arabidopsis thaliana* is a small flowering plant that is a member of the family cruciferae. It has many characteristics—diploid genetics, rapid growth cycle, relatively low repetitive DNA content, and small genome size—that recommend it as the model for a plant genome project. The current status of the genetic and physical maps, as well as efforts to sequence the genome, are presented. Examples are given of genes isolated by using map-based cloning. The importance of the *Arabidopsis* project for plant biology in general is discussed.**

Our understanding of basic plant biology and our ability to genetically manipulate plants for agronomic improvement will depend to a large extent on our understanding of developmental, genetic, and metabolic processes which are fundamental to plants. One of the major challenges in plant molecular biology is the isolation of genes in which the biochemical function of the gene product is unknown. In a number of plant species, genes controlling a wide range of fundamental developmental and metabolic processes have been identified by mutational analysis and placed on classical genetic linkage maps (1). In most cases, while the mutant phenotypes and genetic map locations are known, virtually nothing is known about the gene product. The cloning of these loci therefore relies solely on their mutant phenotype and genetic map position. The relatively large genome sizes and high content of repeated sequences of the majority of plant genomes have meant that cloning by chromosome walking is not a feasible prospect in most plant species.

The development of map-based cloning has focused on three plant species, *Arabidopsis*, tomato, and rice. The first genes to be cloned this way were reported in 1992 (2, 3). Since then, there has been a massive increase in the number of laboratories using this approach, such that, to date, probably >100 loci are being targeted. Most of this effort is centered on *Arabidopsis*, making the availability of a physical map of the *Arabidopsis* genome a high priority for the *Arabidopsis* research community. The immediate benefits of having a physical map are twofold. First, the map provides ready access to any region of the genome which can be genetically identified. In other words, the physical map serves as a cloning tool by facilitating the movement from the genetic locus to the cloned gene. Given a mutation of known genetic map location, the physical map can be used to easily isolate an overlapping collection of clones encompassing the locus of interest. Second, the physical map provides a starting point for studying global genomic organization. As an increasing number of genes are cloned and molecular biological information is accumulated, one can begin to investigate the physical linkage of cloned genes, study the organization and distribution of repetitive elements, and address questions such as how physical distance and genetic distance are correlated. In this context, the map provides the framework for cataloging and integrating molecular biological information. Ultimately, genome organization will be investigated at the nucleotide level. Clearly, physical maps are the logical substrates for genome-sequencing projects.

Why *Arabidopsis*?

Over the past several years, *Arabidopsis thaliana*, a typical flowering plant and a member of the mustard family, has gained increasing popularity as a model system for the study of plant biology. Its short life cycle, small size, and large seed output make it well suited for classical genetic analysis (reviewed in refs. 4 and 5). Mutations have been described affecting a wide range of fundamental developmental and metabolic processes (reviewed in ref. 6). The pioneering work of Redei and coworkers (7, 8) and Koornneef and coworkers (9, 10) using classical segregation analysis led to the construction of the initial genetic maps of *Arabidopsis* that contained ≈90 loci (9). Significant progress has been made in the last several years in expanding the number of markers and the current map contains >280 visible markers (10). An increasing number of cloned genes and restriction fragment length polymorphisms (RFLPs) are available for the alignment of the genetic map with the physical map (11–13). *Arabidopsis* is well suited for physical mapping studies since it has a very small genome (5) and a remarkably low content of interspersed repetitive DNA (14). The genome size of *Arabidopsis* has been estimated by using a variety of different methods (see refs. 5 and 15 for a more complete discussion) including reassociation kinetics (≈80 Mb; refs. 15 and 16), flow cytometry (86–145 Mb; refs. 17 and 18), chromosome volume (100 Mb; ref. 19), and physical mapping of cosmids (100 Mb; ref. 20). These estimates range from about 50 to 150 Mb, with 100 Mb being taken as a convenient compromise. A more accurate estimate awaits the completion of the individual chromosome maps (see below) and eventually nucleotide sequence analysis. Therefore, the positive features of *Arabidopsis* that recommend it for genome research are its small genome size, the presence of relatively few repetitive sequences, and the large number of mapped markers. These features have been exploited in physical mapping strategies. The lack of single-copy gene *in situ* hybridization, flow sorting of chromosomes, and well-characterized hybrids with other plant species has precluded other strategies that have been effective in the *Drosophila* and human physical mapping projects.

Molecular Markers

Mapping of mutations in *Arabidopsis* has typically been carried out by using classical morphological markers (10) or RFLPs (11, 13). Classical morphological markers are simple to use and require no use of expensive molecular reagents but may suffer from ambiguities in scoring, interference with the phenotype to be mapped, and, in most cases, only a few markers can be reliably followed in a single cross. Random amplified polymorphic DNAs (RAPDs) are easily generated, simple to score, and amenable to automation, and a high-density RAPD map of the *Arabidopsis* genome has been constructed (21). However, a major drawback is that these markers are generally dominant, and the method is exquisitely sensitive to DNA concentration and PCR conditions. Thus, as a practical matter, the transfer of such mapping information among laboratories has been difficult. For these reasons,

Abbreviations: RFLP, restriction fragment length polymorphism; YAC, yeast artificial chromosome.

‡To whom reprint requests should be sent at the † address.

RFLPs are commonly used for mapping of gene mutations. Initial RFLP maps were constructed of the *Arabidopsis* genome by using various different independent mapping populations (11, 13). As common markers were incorporated into both maps, it was possible to mathematically integrate the two data sets (22). RFLP markers from both maps have been fingerprinted and integrated into the overlapping cosmid map (see below), thereby establishing contact with the genetic linkage map. These early maps have been largely superseded by the development of "recombinant inbred" or "single seed descent" mapping populations (12, 21) that are now the standard for the field and on which >430 DNA-based markers have been mapped (C. Lister and C.D., unpublished data).

More recently, PCR-based markers, such as simple sequence length polymorphisms (SSLPs) (23), cleaved amplified polymorphic sequences (CAPS) (24), and amplified fragment length polymorphisms (AFLPs) (25) have been developed for *Arabidopsis*. These markers offer a considerable advantage over RFLPs in that only small amounts of tissue are required and the polymorphisms are visible by electrophoresis rather than blotting and hybridization. While SSLPs and CAPS produce codominant markers by using straightforward PCR (the primers are known DNA sequences), AFLP markers are largely dominant, and the technology requires much more complex manipulation for each DNA sample (adapters/ligation/PCR/affinity capture/differential display). Additionally, sophisticated image analysis capability is also required to interpret the complex pattern of banding that is produced for each sample. However, the distinct advantage of AFLP technology is that literally thousands of markers can be identified in a short time, and, when combined with bulked segregant analysis (26), markers can readily be identified within 10 kb of a locus of interest (C. Thomas and J. Jones, personal communication).

Physical Map

Due to the success of the cosmid-fingerprinting strategy in *Caenorhabditis elegans* and the similar size and complexity of the *C. elegans* and *Arabidopsis* genomes, this method was also chosen as the initial strategy for *Arabidopsis*.

Random Linking of Cosmid Clones. The construction of the cosmid map involved the characterization of random clones by fingerprint analysis (20, 27–29). Approximately 20,000 clones (≈8-fold sampling redundancy) from a primary cosmid library were fingerprinted. By using computer matching programs, the clones were aligned into some 750 overlapping groups or contigs. The 750 contigs encompass ≈91,000 kb representing between 90% and 95% of the *Arabidopsis* genome. Having fingerprinted eight genomic equivalents, the practical limit of random clone mapping was reached—i.e., the stage of the project where the rate of finding new joins was unacceptably low due to the scarcity of the linking clones.

Linking Cosmid Contigs by Using Yeast Artificial Chromosome (YAC) to Contig Hybridization. The majority of the gaps in the cosmid contig map were expected to be small and attributable to the fact that the linking cosmids were either nonexistent or underrepresented in the cosmid libraries. Another potential difficulty was the instability of the cosmids—i.e., they are very prone to delete during growth of the cells and tend to give low yields of DNA—hence, they have to be handled with considerable care (20). An attempt was made to join the cosmid contigs by using linking clones from two (EG and EW) of the available *Arabidopsis* YAC libraries listed in Table 1 (30–33). The large insert sizes which can be propagated in YAC vectors (35) means that fewer clones need be examined, but more important, the yeast cloning system offers the potential to give a random or at least different representation of sequences than are obtained with cosmids. The linking strategy was to use YAC clones to probe ordered cosmid grids representative of the contigs and the unattached clones (36). Cosmids within a contig were chosen so that there was minimal overlap between flanking clones, yet the clones were representative of the entire contig. The colonies were then gridded at high density by interleaving 16 Microtiter dish patterns over an area of 8 × 12 cm. The representative collection of cosmids could therefore be arrayed on two filters the size of a Microtiter plate. The YAC clones used as probes were first separated from the host chromosomes by clamped homogeneous electric field (CHEF) gel electrophoresis and cut out of the gel. The gel slices containing the YACs were then labeled by random priming in Microtiter plates (20). Consequently, 96 probes could be prepared simultaneously, thereby facilitating multiple hybridizations. Since the cosmid libraries were constructed in the lorist series of vectors (37), which have no significant homology with the YAC vectors (35), there was no need to separate the insert from the vector sequences. The overlaps were established by probing the ordered cosmid grids with the labeled YAC clones. The YACs to be used as probes can either be picked at random or preselected by hybridization. The latter approach has the advantage that it can be used to rapidly establish linkage in specified regions of the genome. The disadvantage is that two hybridizations are required: preselection of the YAC clones followed by hybridization to the cosmid grid. Both approaches were employed, with the primary emphasis being the use of randomly selected YAC clones. Although this approach of YAC to cosmid contig hybridization was somewhat fruitful (B. M. Hauge and H.M.G., unpublished data), it was considerably less valuable for contig linking than originally anticipated (36) because the two YAC libraries (EG and EW, see Table 1) used more or less exclusively for these experiments were highly chimeric and many also contained chloroplast or ribosomal DNA. However, the gridded cosmids and their filters have been helpful for linking the YAC contigs and as substrates for genomic sequencing (see below).

Physical Mapping by Anchoring of YAC Clones with Molecular Markers. The *Arabidopsis* genome has ideal properties to undertake a strategy for physical mapping by anchoring YAC clones with genetically mapped molecular markers (38–40). The basis of this method is to identify clusters of overlapping YAC clones that share short "anchor" sequences; with a great many such anchors scattered randomly throughout the genome, contiguous arrays (contigs) of YACs sharing common anchors may be identified. With a sufficient number of anchors, these contigs become assembled into a true physical map with long-range continuity. Furthermore, by using genetically mapped markers as an-

Table 1. Characteristics of available *Arabidopsis* YAC libraries

Library name	Average insert size, kb	No. of clones	% of clones in the library (*n*)				DNA used for library	DNA fragmentation method	Vector	Ref.
			With rDNA	With 180 bp	With Chl DNA	With 160 bp				
EG	≈150	2300	15.4 (354)	1.6 (37)	26.4 (606)	?	COL nuclear	*Bam*HI partial	pYAC41	33
EW	≈150	2200	2.3 (51)	8.45 (186)	6.6 (146)	?	COL nuclear	Random shear	pYAC3	30
yUP	≈250	2300	9.5 (218)	0.7 (16)	7.8 (180)	?	COL nuclear	*Eco*RI partial	pYAC4	31
CIC	≈450	1152	8.9 (103)	5.8 (67)	12.9 (148)	0.09 (1)	COL nuclear	*Eco*RI partial	pYAC4	34

Chl, chlorophyll; COL, Columbia ecotype.

chors, the meiotic and physical maps will be tightly aligned and facilitate the prediction of physical map positions on the basis of linkage data. This is a major benefit for the positional cloning of genes.

Mathematical analysis of a general scheme for mapping the *Arabidopsis* genome by anchoring has been described (40). Several additional theoretical studies about the efficiency of mapping strategies have also been published for other genomes (41–44). Although different assumptions need to be applied for each of the different strategies, important similarities are found in these predictions. Independent of the chosen strategy, physical mapping projects will be more successful if both the number of molecular markers and the insert size of the library are large. Insert size dramatically affects the number and the size of contigs for a specific number of markers, and these two parameters are decisive in the evaluation of the possibilities for map closure (39).

Anchoring strategies have several important advantages over the random-content mapping approaches. Of primary importance is the knowledge of where contigs are being built. When a set of YAC clones are detected with a genetically mapped probe, their position in the genome is immediately known and allows for the correlation of the physical and genetic linkage data. Additionally, the potential relationships between genetically positioned contigs can be assessed and it may be possible to predict which contig ends are likely to cross-hybridize. In this sense, anchoring resembles the approach being used for physical mapping of the *Drosophila* genome (45) in which YACs have been "cytologically anchored" by *in situ* hybridization to salivary gland chromosomes (46, 47).

Construction of an ordered series of overlapping YACs by using molecular markers has been ongoing in several laboratories (32, 34, 38, 39). Screening of the EG, EW, yUP, and CIC YAC libraries (30–34) to isolate clones containing these molecular markers has been carried out by DNA hybridization or by using PCR. By identifying YACs that contain two or more anchors, contigs have been built which consist of overlapping YACs that share common sequences. This approach has provided a facile means of constructing large contigs relatively quickly. Thus far, the experimental progress has closely paralleled the outcome predicted from the mathematical analysis (ref. 38; P. Dunn and J.R.E., unpublished data). Long-range continuity of the physical map is now being achieved by a directed approach toward linking the contigs together with additional YACs and using other large-insert clone libraries (see below).

Directed Linking of YAC Contigs. The directed approach to linking YAC contigs has generally taken the form of utilizing YAC ends as hybridization probes to arrayed YAC libraries (30–34). The end probes were sometimes generated by plasmid rescue or vectorette PCR (48), but, for most of the probes, inverse PCR was the method of choice (49).

The inverse PCR or vectorette PCR products and gel-purified plasmid rescue inserts were radioactively labeled by PCR, hybridized to high-density YAC grids or Southern blots of pulsed-field gel electrophoresis fractionated YAC clones, and all positive signals confirmed by Southern blot analysis to yeast DNA isolated from the YAC-bearing clones. While the linking of YAC clones by using end probes has proven to be straightforward and critical for the linking, it is, nonetheless, labor intensive. YAC-contig expansion with at least two of the early YAC libraries (EG and EW; refs. 30 and 33) was difficult because of their generally small insert sizes (≈150 kb), chimeric nature, and instability of the clones carrying tandemly repeated DNA sequences (Table 1). The problems that arise when trying to construct contigs with libraries with a high proportion of chimeric clones are fairly obvious—e.g., one needs at least two independent YACs confirming each potential join. As a specific example, a contig on chromosome IV utilizing three of the libraries (EG, EW, and yUP; Table 1) contained 84 clones, 22 of which were chimeric—i.e., had both repetitive and unique DNA—and of the 57 end probes generated, only 44 were usable, while 13 remained unlinked or contained repetitive DNA.

The availability of a new library (34) in 1994 (called CIC, see Table 1) consisting of 1152 clones with an average insert size of ≈450 kb and developed by a collaboration among the Center for the Study of Human Polymorphisms, Institut National de la Recherche Agronomique, and Centre National de la Recherche Scientifique in France has facilitated rapid progress in the last year. In many cases, by using the CIC library with larger and less chimeric inserts, the contigs anchored around the RFLP markers coalesce without the use of more markers or chromosome walking. However, integration of the CIC clones with the smaller YAC clones meant that marker order could be established and physical distances reliably estimated.

The current status of the YAC contig map is as follows: chromosome I contains 42 contigs with ≈65% coverage (P. Dunn and J.R.E., unpublished data), chromosome II contains 4 contigs (5 gaps) with ≈80% coverage (Ming-Li Wang, Eve Zachgo, and H.M.G., unpublished data), chromosome III contains 42 contigs with ≈60% coverage (P. Dunn and J.R.E., unpublished data), chromosome IV contains 4 contigs (3 gaps) with >90% coverage (64), and chromosome V contains 35 contigs with ≈85% coverage (R. Schmidt and C.D., unpublished data).

A potential obstacle to closure of the physical map is the absence of certain sequences from the YAC and cosmid libraries. Thus, to construct a "complete" map, it is almost certain that other large-insert clone libraries will be necessary. The recent availability of both bacterial artificial chromosome (**; P. Ronald, personal communication) and P1-based *Arabidopsis* genomic (50) libraries has expanded the possibilities for completing the chromosome physical maps.

AAtDB. A continuously updated collection of relevant information on *Arabidopsis* is available in An *Arabidopsis thaliana* Data Base (AAtDB) (51). The focus of AAtDB is on the genetic and physical maps of the *Arabidopsis thaliana* chromosomes. Genetic maps continue to be refined, and AAtDB contains information on several versions of the five chromosome maps, with visible as well as molecular markers. The physical maps contain data on cosmid and YAC contigs, as well as some sequence information. Additional information related to *Arabidopsis* genetics and genetic research is also included, such as literature citations, *Arabidopsis* sequences from GenBank, germ plasm and DNA resources, and images. AAtDB uses the ACeDB software developed by Richard Durbin (Medical Research Council–Laboratory of Molecular Biology, Cambridge, U.K.) and Jean Thierry-Mieg (Centre National de la Recherche Scientifique, Montpellier, France) for the *C. elegans* genome project. The software runs on UNIX workstations (a Macintosh version is also available) and provides the user with a dense network of graphical and hypertext links to examine data threads of interest. This is accessible through the Internet worldwide computer network. AAtDB also features a wide variety of public information that is presented using graphical, text, and tabular formats. The user interface, which was designed to invite browsing, allows users to explore information by pointing and clicking with the workstation mouse or by using a versatile query facility (Table 2). All the AAtDB information in text form is available via the *Arabidopsis* Research Companion, an Internet Gopher server at Massachusetts General Hospital.

Sequence Analysis

Most of the genomic sequence data currently available in the data bases (1016 entries) have been obtained by individual sequencing of cloned genes of interest. Large-scale sequencing of the *Arabidopsis*

**Choi, S., Creelman, R. A., Mullet, J. E. & Wing, R. A. (1955) Weeds World (http://weeds.mgh.harvard.edu., Mass. General Hospital, Boston), Vol. 2.

Table 2. Numbers of objects in selected classes of AAtDB (version 3-5)

Object	No.
Map	31
Allele	1295
Author	5425
Clone	19163
Contact	1419
DNA_Resource	2869
Gene_Class	244
Gene_Product	791
Germplasm_Resource	3520
Image	819
Journal	395
Locus	1926
Map_Population	26
Motif	343
Paper	4192
Probe	528
Sequence	34686
Sequence_EST	14915
Sequence_Homol	17257
Sequence_Genomic	2495
Source	60
2_point_data	4877

genome is now underway with a major expansion of activity planned for the next 10 years. The focus will be on the sequence analysis of the ecotype Columbia, used extensively in production of the YAC, P1, and cosmid libraries and a parent in one set of recombinant inbred lines. The first contribution to the large-scale genomic sequencing has been the completion of two 40-kb cosmid clones (H.M.G., P. Gallant, and G. Church, unpublished data). A program to sequence just under 2 Mb of *Arabidopsis* genomic DNA was funded by the European Economic Community program in 1993 and is being coordinated by M. Bevan (John Innes Centre, Norwich, U.K.). A total of 1.5 Mb of contiguous sequence is being generated on the long arm of chromosome IV, centered around a locus controlling flowering time (*FCA*). In addition, 150 cDNA clones transcribed from this 1.5-Mb region will be sequenced. Information from the comparison of the cDNA and genomic sequences will be used to "teach" the GENEFINDER or equivalent software to recognize *Arabidopsis* splice consensus sites. The rest of the genomic sequence is made up of 400 kb throughout the genome, although a large part of this is centered on a region around a locus on chromosome IV controlling floral morphology (*AP2*). This project, termed the ESSA project (for European Scientists Sequencing *Arabidopsis*) is being considered as a pilot program (to be completed by 1996) for a future large-scale project (covering on the order of 10 Mb), funded by the European Economic Community.

The templates for the sequencing of the 1.5 Mb of chromosome IV have so far been exclusively cosmid clones necessitating the production of overlapping cosmid clones from the detailed YAC physical map of that region. Cosmid clones have been identified by hybridization of whole YAC clones to filters carrying representative cosmids from the fingerprinting experiments (see earlier section). In addition, a second cosmid library made from the Columbia ecotype was screened (I. Bancroft, K. Love, and C.D., unpublished data). Approximately 80% cosmid coverage was achieved with this YAC to cosmid hybridization strategy. Currently purified YAC clones are being subcloned into cosmid vectors. The high-redundancy libraries that can be achieved this way (average 30-fold) have so far given complete coverage.

In parallel to these genomic sequencing programs, expressed sequence tags (ESTs) are being generated (52, 53). These are partial sequences of random cDNA clones isolated either from one cDNA library made from mRNA isolated from a mixture of tissues and conditions (52) or from different cDNA libraries made from a variety of tissues and cultured cells (53). So far, 14,915 EST sequences are in the dbEST data base. A first analysis has revealed that 40% of the clones show significant similarity to known genes.

Interim Lessons Learned

What has been learned from the study of the genome of *Arabidopsis thaliana* and how has this contributed to our understanding of plant biology?

Genome Structure. A considerable amount of information on the genome of *Arabidopsis* has been accumulated as outlined in the previous sections. In addition, the first plant telomeres were isolated from *Arabidopsis*, and their partial map positions were determined (54). A recent report also indicates that the genes coding for ribosomal RNA that exist as large rDNA tandem repeats (together these account for ≈6% of the genome) occur in two nucleolar-organizer regions which map near the telomeres of chromosomes II and IV (55). The positions of the centomeres were initially defined genetically for chromosomes I, III, and V by using trisomic lines (9), and later clones were isolated that mapped close to *CEN1* (56). Recently, centromeres have been located more precisely on the YAC contig maps of chromosomes II (H.M.G., unpublished data) and IV (64) by hybridization to a probe, pAL1 (57), containing a 180-bp *Hin*dIII repeat. This had been shown by using *in situ* hybridization to colocalize the centromeres with the paracentromeric heterochromatin of all five chromosomes (58). For chromosome IV, the centromere was located close to the marker mi87, indicating that the short arm of this chromosome contains ≈3.5 Mb of the coding DNA and ≈3 Mb of rDNA at the terminus (55, 64).

The data on genome structure and accumulated sequence data have confirmed previous information that *Arabidopsis* genes are on the average rather small (≈4–5 kb) and in general contain smaller (usually <200 bp) and fewer introns than other plant species. In addition, it appears that there is little space between genes. It can therefore be predicted that *Arabidopsis* contains ≈25,000 genes. On a global scale, the correlation between the genetic and physical maps indicates that there is ≈200 kb per centimorgan (cM), while data from the physical mapping of chromosome IV indicates ≈175 kb per cM for that chromosome. This number can differ significantly in any local region.

Gene Isolation. A large number of *Arabidopsis* genes have been isolated by taking advantage of the integrated genetic, physical, and RFLP maps. As examples, among the first two *Arabidopsis* genes that were isolated by using map-based or positional cloning were the genes corresponding to the *fad3* locus that encodes an $\omega - 3$ desaturase (2) and *ABI3* (3), the wild-type gene corresponding to one of the abscisic acid insensitive mutants (*abi3*) of *Arabidopsis*. These *abi* mutants are altered in various aspects of seed development and germination due to a decreased responsiveness to the hormone abscisic acid. On the basis of its properties, this gene is thought to be involved in the abscisic acid signal transduction pathway.

Of particular note was the recent isolation of a plant disease resistance gene, *RPS2*, from *Arabidopsis* (59, 60). *Arabidopsis* mutant plants were isolated that were resistant to infection by the bacterial pathogen *Pseudomonas syringae* and used for positional cloning of the gene. This gene encodes a protein with 14 imperfect leucine-rich repeats and is probably a receptor for a molecular signal of the pathogen (59, 60). This becomes even more interesting (61) since *RPS2* is similar to another *Arabidopsis* gene, *RPP5*, that confers resistance to a fungal pathogen *Peronospora parasitica* (ref. 62; J. Parker and J. Jones, personal communication) and to the *N* gene of tobacco that confers resistance to tobacco mosaic virus (63).

Conclusions

Considerable progress has been made in a relatively short period of time in isolating and mapping genes and developing a physical map of the model plant *Arabidopsis thaliana*. Work is in progress to complete the physical maps of each of the five chromosomes and to initiate large-scale genomic sequencing. These studies and others have contributed to our understanding of the biology of *Arabidopsis* and of plants in general.

1. O'Brien, S. J., ed. (1990) *Genetic Maps: Locus Maps of Complex Genomes*; Book 6, Plants (Cold Spring Harbor Lab. Press, Plainview, NY).
2. Arondel, V., Lemieux, B., Hwang, I., Gibson, S., Goodman, H. M. & Somerville, C. R. (1992) *Science* **258,** 1353–1355.
3. Giraudat, J., Hauge, B. M., Valon, C., Smalle, J., Parcy, F. & Goodman, H. M. (1992) *Plant Cell* **4,** 1251–1261.
4. Meyerowitz, E. M. (1987) *Annu. Rev. Genet.* **21,** 93–111.
5. Meyerowitz, E. M. (1994) in *Arabidopsis*, eds. Meyerowitz, E. M. & Somerville, C. R. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 21–36.
6. Estelle, M. A. & Somerville, C. R. (1986) *Trends Genet.* **2,** 89–93.
7. Redei, G. P. & Hirono, Y. (1964) *Arabidopsis Inf. Serv.* **1,** 9–10.
8. Koncz, C. & Redei, G. P. (1994) in *Arabidopsis*, eds. Meyerowitz, E. M. & Somerville, C. R. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 223–252.
9. Koornneef, M., Van Eden, J., Hanhart, C. J., Stam, P., Braaksma, F. J. & Feenstra, W. J. (1983) *J. Hered.* **74,** 265–272.
10. Koornneef, M. (1994) in *Arabidopsis*, eds. Meyerowitz, E. M. & Somerville, C. R. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 89–120.
11. Chang, C., Bowman, J. L., De John, A. W., Lander, E. S. & Meyerowitz, E. M. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 6856–6860.
12. Lister, C. & Dean, C. (1993) *Plant J.* **4,** 745–750.
13. Nam, H. G., Giraudat, J., Den Boer, B., Moonan, F., Loos, W. D. B., Hauge, B. M. & Goodman, H. M. (1989) *Plant Cell* **1,** 699–705.
14. Pruitt, R. E. & Meyerowitz, E. M. (1986) *J. Mol. Biol.* **187,** 169–184.
15. Meyerowitz, E. M. (1992) in *Methods in Arabidopsis Research*, eds. Koncz, C., Chua, N.-H. & Schell, J. (World Scientific, Singapore), pp. 100–118.
16. Leutwiler, L. S., Hough-Evans, B. R. & Meyerowitz, E. M. (1984) *Mol. Gen. Genet.* **194,** 15–23.
17. Arumuganathan, K. & Earle, E. D. (1991) *Plant Mol. Biol. Rep.* **9,** 208–218.
18. Galbraith, D. W., Harkins, K. R. & Knapp, S. (1991) *Plant Physiol.* **96,** 985–989.
19. Heslop-Harrison, J. S. & Schwarzacher, T. (1990) in *Abstracts from the Fourth International Conference on Arabidopsis Research* (Univ. of Vienna, Vienna), p. 3.
20. Hauge, B. M. & Goodman, H. M. (1992) in *Methods in Arabidopsis Research*, eds. Koncz, C., Chua, N.-H. & Schell, J. (World Scientific, Singapore), pp. 191–223.
21. Reiter, R. S., Williams, J. G. K., Feldmann, K. A., Rafalski, J. A., Tingey, S. V. & Scolnik, P. A. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 1477–1481.
22. Hauge, B. M., Hanley, S. M., Cartinhour, S., Cherry, J. M., Goodman, H. M., Koornneef, M., Stam, P., Chang, C., Kempin, S. & Medrano, L. (1993) *Plant J.* **3,** 745–754.
23. Bell, C. J. & Ecker, J. R. (1994) *Genomics* **19,** 137–144.
24. Konieczny, A. & Ausubel, F. M. (1993) *Plant J.* **4,** 403–410.
25. Zabeau, M. & Vos, P. (1992) Eur. Patent Appl. 92402629.7 (Publ. 0 534 858A1).
26. Michelmore, R. W., Paran, I. & Kessel, R. V. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 9828–9832.
27. Coulson, A., Sulston, J., Brenner, S. & Karn, J. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 7821–7825.
28. Hauge, B. M., Giraudat, J., Hanley, S., Hwang, I., Kochi, T. & Goodman, H. M. (1991) in *Plant Molecular Biology*, eds. Herrman, R. G. & Larkins, B. (Plenum, New York), Vol. 2, pp. 239–248.
29. Hauge, B. M. & Goodman, H. M. (1992) in *Plant Genomes: Methods for Genetic and Physical Mapping*, eds. Beckmann, J. S. & Osborn, T. C. (Kluwer, Dordrecht, The Netherlands), pp. 101–139.
30. Ward, E. R. & Jen, G. C. (1990) *Plant Mol. Biol.* **14,** 561–568.
31. Ecker, J. R. (1990) *Methods (Duluth)* **1,** 186–194.
32. Hwang, I., Kohchi, T., Hauge, B. M., Goodman, H. M., Schmidt, R., Cnops, G., Dean, C., Gibson, S., Iba, K., Lemieux, B., Arondel, V., Danhoff, L. & Somerville, C. (1991) *Plant J.* **1,** 367–374.
33. Grill, E. & Somerville, C. (1991) *Mol. Gen. Genet.* **226,** 484–490.
34. Creusot, F., Fouilloux, E., Dron, M., Lafleuriel, J., Picard, G., Billault, A., Le Paslier, D., Cohen, D., Chabouté, M.-E., Durr, A., Fleck, J., Gigot, C., Camilleri, C., Bellini, C., Caboche, M. & Bouchez, D. (1995) *Plant J.*, in press.
35. Burke, D. T., Carle, G. F. & Olson, M. V. (1987) *Science* **236,** 806–812.
36. Coulson, A., Waterston, R., Kiff, J., Sulston, J. & Kohara, Y. (1988) *Nature (London)* **335,** 184–186.
37. Gibson, T. J. & Sulston, J. E. (1987) *Gene Anal. Technol.* **4,** 41–44.
38. Schmidt, R. & Dean, C. (1993) *BioEssays* **15,** 63–68.
39. Matallana, E., Bell, C. J., Dunn, P., Lu, M. & Ecker, J. R. (1992) in *Methods in Arabidopsis Research*, eds. Koncz, C., Chua, N.-H. & Schell, J. (World Scientific, Singapore), pp. 144–169.
40. Ewens, W. J., Bell, C. J., Donnelly, P. J., Dunn, P., Matallana, E. & Ecker, J. R. (1991) *Genomics* **11,** 1–7.
41. Arrartia, R., Lander, E. S., Tavaré, S. & Waterman, M. S. (1991) *Genomics* **11,** 806–827.
42. Barillot, E., Dausset, J. & Cohen, D. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 3917–3921.
43. Palazzolo, M. J., Sawyer, S. A., Martin, C. H., Smoller, D. A. & Hartl, D. L. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 8034–8038.
44. Torney, D. C. (1991) *J. Mol. Biol.* **217,** 259–264.
45. Aijoka, J. W., Smoller, D. A., Jones, R. W., Carulli, J. P., Vellek, A. E. C., Garza, D., Link, A. J., Duncan, I. W. & Hartl, D. L. (1991) *Chromosoma* **100,** 495–509.
46. Garza, D., Ajioka, J. W., Burke, D. T. & Hartl, D. L. (1989) *Science* **246,** 641–646.
47. Lozovskaya, E. R., Petrov, D. A. & Hartl, D. L. (1993) *Chromosoma* **102,** 253–266.
48. Riley, J., Butler, R., Ogilvie, D., Finniear, R., Jenner, D., Powell, S., Anand, R., Smith, J. C. & Markham, A. F. (1990) *Nucleic Acids Res.* **18,** 2887–2890.
49. Ochman, H., Gerber, A. S. & Hartl, D. L. (1988) *Genetics* **120,** 621–623.
50. Liu, Y. G., Mitsukawa, N., Vazquestello, A. & Whittier, R. F. (1995) *Plant J.* **7,** 351–358.
51. Cherry, J. M., Cartinhour, S. W. & Goodman, H. M. (1992) *Plant Mol. Biol. Rep.* **10,** 308–309, 409–411.
52. Newman, T., de Bruijn, F. J., Green, P., Keegstra, K., Kende, H., McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomashow, M., Retzel, E. & Somerville, C. (1994) *Plant Physiol.* **106,** 1241–1255.
53. Hofte, H., Desprez, T., Amselem, J., Chiapello, H., Caboche, M. & Moisan, A. (1993) *Plant J.* **4,** 1051–1061.
54. Richards, E. J. & Ausubel, F. M. (1988) *Cell* **53,** 127–136.
55. Copenhaver, G. P., Doelling, J. H., Gens, J. S. & Pikaard, C. S. (1995) *Plant J.* **7,** 273–286.
56. Richards, E. J., Goodman, H. M. & Ausubel, F. M. (1991) *Nucleic Acids Res.* **19,** 3351–3358.
57. Martinez-Zapater, J. M., Estelle, M. A. & Somerville, C. R. (1986) *Mol. Gen. Genet.* **204,** 417–423.
58. Maluszynska, J. & Heslop-Harrison, J. S. (1991) *Plant J.* **1,** 159–166.
59. Mindrinos, M., Katagiri, F., Yu, G.-L. & Ausubel, F. M. (1994) *Cell* **78,** 1089–1099.
60. Bent, A. F., Kunkel, B. N., Dahlbeck, D., Brown, K. L., Schmidt, R., Giraudat, J., Leung, J. & Staskawicz, B. (1994) *Science* **265,** 1856–1860.
61. Staskawicz, B. J., Ausubel, F. M., Baker, B. J., Ellis, J. G. & Jones, J. D. G. (1995) *Science* **268,** 661–667.
62. Parker, J., Szabo, V., Staskawicz, B., Lister, C., Dean, C., Daniels, M. & Jones, J. (1993) *Plant J.* **4,** 821–831.
63. Whitham, S., Dinesh-Kumar, S. P., Choi, D., Hehl, R., Corr, C. & Baker, B. (1994) *Cell* **78,** 1101–1115.
64. Schmidt, R., West, J., Love, K., Lenehan, Z., Lister, C., Thompson, H., Bouchez, D. & Dean, C. (1995) *Science*, in press.