16. _____ and E. C. McMonagle, *Virology* **130**, 427 (1983).
17. S. K. Weller *et al.*, *Mol. Cell. Biol.* **5**, 930 (1985).
18. D. A. Vlazny, A. Kwong, N. Frenkel, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 1423 (1982).
19. R. W. Honess and B. Roizman, *J. Virol.* **12**, 1346 (1973).
20. E. K. Wagner, in *The Herpesviruses*, B. Roizman, Ed. (Plenum, New York, 1985), vol. 3, pp. 45–104.
21. R. W. Honess and B. Roizman, *J. Virol.* **14**, 8 (1974).
22. _____, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 1276 (1975).
23. A. J. Conley, D. M. Knipe, P. C. Jones, B. Roizman, *J. Virol.* **37**, 191 (1981).
24. D. M. Knipe, W. T. Ruyechan, B. Roizman, I. W. Halliburton, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 3896 (1978).
25. D. M. Knipe, W. T. Ruyechan, R. W. Honess, B. Roizman, *ibid.* **76**, 4534 (1979).
26. R. A. F. Dixon and P. A. Schaffer, *J. Virol.* **36**, 189 (1980).
27. L. E. Post and B. Roizman, *Cell* **25**, 227 (1981).
28. C. Greene and P. Schaffer, *Abstracts of the Ninth International Herpesvirus Workshop* (1984), p. 11.
29. A. E. Sears, I. W. Halliburton, B. Meignier, S. Silver, B. Roizman, *J. Virol.*, in press.
30. P. J. Godowski and D. M. Knipe, *ibid.* **47**, 478 (1983).
31. M. L. Fenwick and M. J. Walker, *J. Gen. Virol.* **41**, 37 (1978).
32. G. S. Read and N. Frenkel, *J. Virol.* **46**, 498 (1983).
33. S. Mackem and B. Roizman, *ibid.* **43**, 1015 (1982).
34. _____, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4917 (1982).
35. _____, *J. Virol.* **44**, 939 (1982).
36. W. Batterson and B. Roizman, *ibid.* **46**, 371 (1983).
37. M. E. M. Campbell, J. W. Palfreyman, C. M. Preston, *J. Mol. Biol.* **180**, 1 (1984).
38. P. E. Pellett, J. L. C. McKnight, F. J Jenkins, B. Roizman, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
39. R. J. Watson, C. M. Preston, B. Clements, *J. Virol.* **31**, 42 (1979).
40. S. Mackem and B. Roizman, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 7122 (1980).
41. T. M. Kristie and B. Roizman, *ibid.* **81**, 4065 (1984).
42. T. J. Hill, in *The Herpesviruses*, B. Roizman, Ed. (Plenum, New York, 1985), vol. 3, pp. 175–240.
43. C. Morgan, H. M. Rose, B. Mednis, *J. Virol.* **2**, 507 (1968).
44. W. Batterson, D. Furlong, B. Roizman, *ibid.* **45**, 397 (1983).
45. F. Constanzo, G. Campadelli-Fiume, L. Foa-Tomas, E. Cassai, *ibid.* **21**, 996 (1977).
46. H. M. Keir and E. Gold, *Biochim. Biophys. Acta* **72**, 263 (1963).
47. R. J. Jacob, L. S. Morse, B. Roizman, *J. Virol.* **29**, 448 (1979).
48. T. Ben-Porat and S. Tokazewski, *Virology* **79**, 292 (1977).
49. W. Gibson and B. Roizman, *J. Virol.* **10**, 1044 (1972).
50. D. K. Braun, B. Roizman, L. Pereira, *ibid.* **49**, 142 (1984).
51. R. W. Darlington and L. H. Moss III, *Prog. Med. Virol.* **11**, 16 (1969).
52. J. Schwartz and B. Roizman, *J. Virol.* **4**, 879 (1969).
53. S. Kit and D. R. Dubbs, *Virology* **26**, 16 (1965).
54. G. H. Cohen, *J. Virol.* **9**, 408 (1972).
55. D. Huszar and S. Bacchetti, *ibid.* **37**, 580 (1981).
56. F. Wohlrab and B. Francke, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 100 (1980).
57. V. G. Preston and F. B. Fisher, *Virology* **138**, 58 (1984).
58. D. R. Dubbs and S. Kit, *ibid.* **22**, 493 (1964).
59. D. Lando and M. L. Rhyiner, *C. R. Acad. Sci.* **269**, 527 (1969).
60. F. L. Graham, G. Velihaisen, N. M. Wilkie, *Nature (London) New Biol.* **245**, 265 (1973).
61. P. Sheldrick *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **70**, 3621 (1973).
62. W. T. Ruyechan, L. S. Morse, D. M. Knipe, B. Roizman, *J. Virol.* **29**, 677 (1979).
63. J. Hubenthal-Voss and B. Roizman, unpublished observations.
64. R. B. Tenser and M. E. Dunston, *Virology* **99**, 417 (1979).
65. M. Ackermann, M. Sarmiento-Batterson, B. Roizman, in preparation.
66. J. Hubenthal-Voss and B. Roizman, *J. Virol.* **54**, 509 (1985).
67. F. J. Jenkins, M. J. Casadaban, B. Roizman, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4773 (1985).
68. B. A. Castilho, P. Olfson, M. J. Casadaban, *J. Bacteriol.* **158**, 488 (1984).
69. N. D. Stow, E. C. McMonagle, A. J. Davison, *Nucleic Acids Res.* **11**, 8205 (1983).
70. K. L. Poffenberger, E. Tabares, B. Roizman, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2690 (1983).
71. P. Mauromara-Nazos, J. Hubenthal-Voss, B. Roizman, unpublished observations.
72. S. Silver and B. Roizman, *Mol. Cell. Biol.* **5**, 518 (1985).
73. G. T.-Y. Lee, K. L. Pogue-Geile, L. Pereira, P. G. Spear, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 6612 (1982).
74. M. G. Gibson and P. G. Spear, *J. Virol.* **48**, 396 (1983).
75. M.-F. Shih, M. Arsenakis, P. Tiollais, B. Roizman, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5867 (1984).
76. M. Arsenakis, K. L. Poffenberger, B. Roizman, in preparation.
77. M. Hummel *et al.*, in preparation.
78. C. Tackney, G. Cachianes, S. Silverstein, *J. Virol.* **52**, 606 (1984).
79. K. L. Pogue-Geile, G. T.-Y. Lee, P. G. Spear, *ibid.* **53**, 456 (1985).
80. K. L. Poffenberger and B. Roizman, *ibid.* **53**, 587 (1985).
81. F. J. Jenkins and B. Roizman, unpublished observations.
82. M. Mackett, G. L. Smith, B. Moss, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 7415 (1982).
83. G. L. Smith, M. Mackett, B. Moss, *Nature (London)* **302**, 490 (1983).
84. G. L. Smith, B. R. Murphy, B. Moss, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 7155 (1983).
85. M. Mackett, G. L. Smith, B. Moss, *J. Virol.* **49**, 857 (1984).
86. S. Gillard, D. Spehner, P. Drillien, *ibid.* **53**, 316 (1985).
87. Supported in part by grants from the National Cancer Institute (CA08494 and CA19264), United States Public Health Service, American Cancer Society (MV2T), and Institut Merieux. F.J.J. is a postdoctoral trainee (CA09241) of the National Cancer Institute.

Arabidopsis thaliana and Plant Molecular Genetics

Elliot M. Meyerowitz and Robert E. Pruitt

It is worth understanding the molecular genetics of plants not only because such understanding has practical value in the improvement of crops but also because studies of plants offer an opportunity to gain insight into various basic life processes unique to plants. Plants do not use the same hormones as animals; nor do they use hormones in the same way that animals use them. Most or all of the cells in a plant may both produce and respond to plant hormones at some time. Plants respond to stress differently than animals. Plants also respond to light in diverse and subtle ways, with photosynthesis being only one of the responses of plants to light that are not found in animals. Even the basic developmental processes of plants have features that distinguish them from those of animals. Among these features are the absence of cell migration in plant development and the fact that each flowering plant has certain parts (the meristems) that remain embryonic and can produce adult organs, including germ cells, throughout the life of the plant. Further, individual differentiated cells taken from the vegetative parts of plants can dedifferentiate and regenerate to form entire, fertile plants.

Although it is sensible and necessary to study crop plants for purposes of crop improvement, the crop plants now used for basic classical and molecular genetic studies have disadvantages for some of the types of experimentation used in this work. Classical genetics depends on researchers being able to raise many successive generations of organisms in large numbers. Typical crop plants have generation times of several months, and they require a great deal of field space for growth in large numbers. The genetics of some of these plants is also made more difficult by polyploidy or allopolyploidy. The ease with which recombinant DNA work can be done with any organism depends in part on the size of the organism's nuclear genome; the smaller the genome, the less work is required to screen recombinant DNA libraries and thus to isolate any particular gene. The genomes of the plants presently used for recombinant DNA

Elliot M. Meyerowitz is an associate professor and Robert E. Pruitt is a graduate student in the Division of Biology, California Institute of Technology, Pasadena 91125.

work are large, being similar in size to those found in mammals (Table 1). In addition, the plants generally used for recombinant DNA work contain large amounts of dispersed repetitive DNA (*1*), which makes it very difficult or impossible to perform such procedures as genomic blot analysis with genomic clones and chromosome walking. *Arabidopsis thaliana* provides a system without these specific disadvantages, and it offers the possibility of doing some types of molecular genetic experiments more rapidly, more easily, and at less expense than other plants currently permit.

Summary. Arabidopsis thaliana is a small flowering plant with various properties that make it an excellent organism for experiments in molecular genetics. These properties include having a small nuclear genome, a near absence of dispersed repetitive DNA, and a generation time of 4 to 5 weeks. In addition, mutations that affect hormone synthesis and response, many different enzyme activities, and numerous developmental processes have been identified and characterized.

Classical and Biochemical Genetics

Arabidopsis is a flowering plant that has been used in classical genetic work for over 40 years (*2*) and that has an extensive genetic and ecological literature. The plant is a member of the mustard family, along with some more familiar plants, such as cabbages and radishes. It is a harmless weed of no food or other economic value. Nonetheless, the literature on the plant indicates that for several reasons, *Arabidopsis* may be of considerable value in molecular genetic research. The plant is well suited to classical genetic work: it has a generation time of only 5 weeks; it can produce more than 10,000 seeds per plant; and it is of such small size that dozens can be grown in a small pot, and tens of thousands can be grown in a small room (*3*). It requires only moist soil and fluorescent light for rapid growth. The small flowers contain both anthers and pistils, and the plant typically self-fertilizes. Self-fertilization allows new mutations to be made homozygous with minimal effort. When desired, cross-fertilization can be simply and rapidly effected, thereby making possible the crosses necessary for genetic mapping and for the production of multiple mutant stocks. Mutagenesis can be performed by soaking seeds in chemical mutagens such as ethyl methanesulfonate or by irradiating of seeds soaked in water. The mutagenized seeds are planted, grown to maturity, and allowed to self-fertilize, thereby producing seeds that are homozygous for new mutations.

Various mutations have been identified, including visible mutations useful as markers in genetic mapping, with phenotypes affecting every part of the plant. These include mutations affecting the wax coat of the epidermal cells (*eceriferum* mutants) (*4*) and the trichomes normally found on leaves and stems (*distorted trichomes* and *globra* mutants) (*5*), as well as those mutations causing more easily visible effects, such as changes in flower morphology (*agamous apetala* and *pistillata* mutants) (*6*) and growth habit (*erecta* and *compacta* mutants) (*6, 7*). Other mutations affect the embryonic development of *Arabidopsis* and lead to embryonic lethality; the stage at which developmental arrest occurs depends on the specific gene mutated (*8*). Many leaf-color mutants exist, including two that give variegated plants. Plants homozygous for the immutans mutation have leaves that are variegated white and green; the degree of this variegation depends both on the particular allele of the immutans mutation and on the intensity of the light falling on the plant at the time a given leaf develops. This recessive mutation is inherited from both white and green sectors (*9*). Plants homozygous for the chloroplast mutator mutation also have leaves that bear white and green sectors. In this case the white sectors are attributable to the failure of the chloroplasts to develop normally. Although plants heterozygous for the chloroplast mutator mutation do not produce aberrant chloroplasts, chloroplasts that heterozygous plants maternally inherit from homozygous chloroplast mutator plants continue to be of abnormal structure. In fact, from these outcrossed chloroplast mutator strains it is possible to isolate stable homoplastidic lines of *Arabidopsis* in which all the chloroplasts have the same abnormal appearance. It has been proposed that the nuclear chloroplast mutator mutation acts by causing errors during chloroplast DNA replication (*10*).

Biochemical mutants of various types have also been isolated. Mutations in several genes give rise to plants that for growth require thiamine or the thiamine precursors 2,5-dimethyl-4-aminopyrimidine or 4-methyl-5-hydroxyethylthiazole (*11, 12*). Somerville and Ogren have characterized biochemical mutations that affect photorespiration and photosynthesis. These include nuclear mutations that result in the absence of activity of chloroplast glutamate synthase, phosphoglycolate phosphatase, mitochondrial glycine decarboxylase, mitochondrial serine transhydroxymethylase, and serine-glyoxylate aminotransferase (*13*). Mutations that affect the activity of nitrate reductase and of alcohol dehydrogenase have also been characterized (*14*).

In addition, mutations have been obtained that appear to disrupt the normal action of several of the plant growth regulators. Koornneef and his collaborators have recovered mutations that cause absence of gibberellins or of abscisic acid, as well as mutations that confer resistance to high levels of exogenous abscisic acid. The mutations that affect gibberellins fall into five complementation groups. Three of these contain alleles that, when homozygous, prevent germination unless the seeds are treated with exogenous gibberellins (*15*). A mutant strain lacking abscisic acid has been used by Karssen *et al.* to examine the relative effects of maternal and embryo-derived abscisic acid on seed dormancy (*16*). Mutations that result in resistance to high levels of a synthetic auxin have also been recovered, and some alleles of these mutations produce plants with agravitropic roots (*17*). One of these mutations is dominant and produces a plant with a dwarf rosette and inflorescence when heterozygous, but is lethal when homozygous.

More than 75 of the known mutations have been assembled into a genetic linkage map by Koornneef *et al.* (*6*). This genetic map consists of five linkage groups, in concordance with the haploid chromosome number of five. The location of the centromeres on the genetic map has been determined by the use of strains that are trisomic for one arm of a chromosome (*6, 18*). *Arabidopsis* stocks containing many of the mutations on the genetic map in combinations suitable for mapping specific chromosome regions, together with many strains collected from the wild, are available from a central international seed collection (*19*).

In addition to growing them in soil, it is possible to grow whole *Arabidopsis* plants on sterile, biochemically defined media; both solid and liquid media have been used (*11, 20*). *Arabidopsis* cells

Table 1. Haploid genome size in various flowering plants and the number of lambda clones that must be screened to have a 99 percent chance of isolating a single-copy sequence from these genomes. The genome sizes are calculated from kinetic complexity measurements (*38*). The library sizes are calculated assuming a random nuclear DNA library with an average clone insert length of 20 kb; 4.6 genome equivalents must be screened for a 99 percent probability of isolating any individual unique sequence.

Plant	Haploid genome size (in kilobase pairs)	Number of lambda clones in complete library
Arabidopsis	70,000	16,000
Mung bean	470,000	110,000
Cotton	780,000	180,000
Tobacco	1,600,000	370,000
Soybean	1,800,000	440,000
Pea	4,500,000	1,000,000
Wheat	5,900,000	1,400,000

have also been grown in tissue culture, and plants have been regenerated from such cells (*21*). Some types of biochemical mutations have been isolated by selection of cultured cells; it may also be possible to use biochemical selections to assay transformation of cells in culture.

Molecular Genetics: Genome Size and Organization

Three lines of evidence indicate that *Arabidopsis* has a small genome. Microspectrophotometry of Feulgen-stained nuclei gives an estimate of 0.2 pg, or approximately 2×10^8 base pairs (bp), for the quantity of DNA in the *Arabidopsis* haploid genome. This is the smallest published size for an angiosperm genome (*22*). A second line of evidence indicating that *Arabidopsis* has a small genome is based on the proportionality of genome size and nuclear volume. This relationship implies that *Arabidopsis* has a haploid genome size of roughly 10^9 bp (*23*). These estimates indicate that the *Arabidopsis* genome is small enough to greatly simplify the task of screening recombinant DNA libraries to isolate genes.

We have performed a more detailed analysis of the *Arabidopsis* genome to verify its small size and to determine the fraction and nature of repetitive sequences present. Our initial study was a reassociation analysis of DNA extracted from whole plants (*24*). This gave an estimated genome size of 7×10^7 bp, a size only five times larger than that of the yeast genome (*25*) and much smaller than that of other flowering plants (Table 1). Our analysis also showed that the whole-plant DNA contains 10 to 14 percent very rapidly reannealing sequences (either highly repeated sequences or inverted repeat sequences) and 23 to 27 percent middle repetitive sequences. By using a labeled tracer containing cloned chloroplast sequences, we demonstrated that almost all the middle repetitive sequences are from the chloroplast genome—and thus that *Arabidopsis* has far less nuclear repetitive DNA than other angiosperms (*1*) (Table 2). Leutwiler *et al.* have also examined the degree of cytosine methylation in *Arabidopsis* and have found that only 4.6 percent of the cytosines derived from whole-plant DNA are present as 5-methylcytosine; this is the lowest value known for a flowering plant (*24*).

More recently, we have used recombinant DNA techniques to examine random segments of genomic DNA (*26*). We have subjected randomly selected recombinant lambda phage to a variety of experimental manipulations to determine if they contain unique or repetitive sequences (or both) and to characterize the nature of the repetitive sequences that are contained in the clones. These experiments confirm and extend our previous analysis. A majority of the 50 clones analyzed contain only unique sequences, and most of the clones that contain repetitive sequences are derived either from the chloroplast genome (four clones) or from the nuclear DNA that codes for the large ribosomal RNA's (eight clones). There are approximately 570 copies of the ribosomal DNA per haploid genome, each ribosomal DNA repeat unit is about 10 kilobase pairs (kb) in length, and the repeats are largely arranged in tandem array. In addition, there are two clones that appear to contain duplicated sequences, one clone that contains a low copy number repeat unit that is apparently conserved in all copies (this may represent a fragment of mitochondrial DNA), and three clones that contain both repeated sequences and unique sequences interspersed with each other. These results indicate that the *Arabidopsis* nuclear genome consists of predominantly unique sequences and that most of the nuclear repetitive DNA is ribosomal DNA. The experiments show that much of the nuclear DNA of *Arabidopsis* is organized as extremely long blocks (averaging 120 kb) of unique sequences.

These two sets of experiments demonstrate that *Arabidopsis* has a remarkably small and simple genome. This fact, taken together with the results from all of the previous work on *Arabidopsis*, indicates that the plant does represent a very good model system for basic research in plant molecular biology.

Molecular Genetics: Specific Genes

We have started to do additional work in studying individual genes from *Arabidopsis* to provide material for analysis of the mechanisms by which gene expression is regulated by both developmental and environmental stimuli. So far, we have examined three different genes or gene families, and this work has led us to two generalizations that emphasize the utility of *Arabidopsis* for experiments in molecular cloning. The first generalization is that it is possible to cross-hybridize genes cloned from a wide variety of flowering plants with the homologous gene or genes from *Arabidopsis*. This perhaps is not surprising because fossil evidence indicates that the first major radiation of the angiosperms took place only 1×10^8 to 1.5×10^8 years ago. Our second generalization is that proteins that are encoded by multiple genes or large gene families in other plants are encoded by single genes or small gene families in *Arabidopsis*.

The first gene we examined was the one that encodes the large seed storage protein (*27*). We cloned this gene from a recombinant library containing genomic DNA by using a complementary DNA clone encoding the 12*S* seed storage protein of *Brassica napus* (*28*). In *Arabidopsis* this protein is encoded by a single gene, which indicates that the heterogeneity observed in other storage protein gene families is not required in this plant. Further, the fact that there is only one gene coding for this protein makes it clear that the tissue- and time-specific regulation of the protein is attributable to the activity of only one gene and also that experiments to understand this regulated gene expression need to deal only with this single sequence and not with a large family of similar genes as in other plants.

A second series of experiments involved the cloning of the genes encoding the light-induced chlorophyll a/b binding

protein of *Arabidopsis* (*29*). Chlorophyll a/b binding protein, or light-harvesting chlorophyll protein, is a nuclear-coded component of the light-harvesting antenna complex of the chloroplast thylakoid membranes. These genes were isolated from an *Arabidopsis* genomic library by using as a probe a genomic clone from the tiny aquatic monocot *Lemna gibba* (*30*). There are three genes encoding this protein in *Arabidopsis*, and they are all located in a small gene cluster of less than 6000 bp. The three genes of the *Arabidopsis* chlorophyll a/b binding protein family contrast with the much larger number in the homologous families in other plants. One example is *Petunia*; the thorough studies of Dunsmuir *et al.* have shown that there may be as many as 16 or more members of this gene family, divided into at least five major subclasses (*31*).

Perhaps even more surprising is the fact that it has been determined by DNA sequencing that all the *Arabidopsis* genes encode an identical product after the transit peptides of the three proteins are removed. As with the seed storage protein, therefore, there does not appear to be any protein heterogeneity for the chlorophyll a/b binding protein encoded in the genome of *Arabidopsis*. Any model for this gene family in *Arabidopsis* must recognize that the purpose of the multiple genes is not to provide a series of variant proteins of slightly different function.

G. An (*32*) has taken advantage of the small number of *Arabidopsis* chlorophyll a/b binding protein genes; because there are only three copies, it is possible in a short series of experiments to test all of the upstream regulatory regions of the genes for function and light inducibility. An has fused two of these regions to bacterial chloramphenicol acetyltransferase genes and has used *Agrobacterium* Ti-plasmid constructs to introduce the fusion genes to tissue culture cells of tobacco. In some instances each of the two *Arabidopsis* DNA fragments has conferred light-inducibility on the bacterial gene.

One final example of a specific *Arabidopsis* gene that has been cloned and characterized is the gene that codes for alcohol dehydrogenase (Adh) (*33*). A classical genetic study of alcohol dehydrogenase allozymes had already demonstrated that there is only a single alcohol dehydrogenase gene in *Arabidopsis*, in contrast to the two or three such genes in many other plants (*34*). The *Arabidopsis* gene was isolated from a recombinant library by cross-hybridization with a labeled maize Adh 1 gene fragment. The gene appears to be single-copy, in confirmation of the earlier genetic results; DNA sequencing shows that the *Arabidopsis* gene shares more than 70 percent nucleic acid homology with the maize Adh 1 gene. The proteins coded by the *Arabidopsis* gene and the maize Adh 1 gene have slightly more than 80 percent identity of their amino acids. Further, the structure of the *Arabidopsis* gene and that of the maize gene are strikingly similar. Both the maize Adh 1 and Adh 2 genes contain nine intervening sequences at identical positions (*35*). The *Arabidopsis* gene has six intervening sequences; the positions of all six are coincident with the corresponding positions of six of the maize introns.

Table 2. Average size of single-copy DNA sequences interspersed with repetitive sequences and total amount of repetitive DNA in various plant genomes. The data for *Arabidopsis* are from (*24*) for amount of repetitive DNA and from (*26*) for size of unique DNA stretches; the unique sequence size is from measurements of random cloned fragments. The data for the other plants are from reassociation analysis (*38*).

Plant	Average size in predominant class of single-copy sequences (in kilobase pairs)	Amount of repetitive DNA in haploid genome (in kilobase pairs)
Arabidopsis	120	18,000
Mung bean	>6.7	160,000
Cotton	1.8	310,000
Tobacco	1.4	1,200,000
Soybean	<3	1,100,000
Pea	0.3	3,800,000
Wheat	1	4,400,000

The existing evidence, therefore, leads to two conclusions. The first is that *Arabidopsis* genes will cross-hybridize with homologous genes from other angiosperms, both monocots and dicots; the second is that genes found in large gene families in other plants can exist either in small families or in single copy in *Arabidopsis*. The practical importance of the first conclusion is that genes of interest can be simply cloned from the extraordinarily small *Arabidopsis* genome and then used as probes for the isolation of the homologous genes from plants of economic value. The importance of the second conclusion is that genes found in gene families can be more easily and more thoroughly studied in *Arabidopsis* than in other angiosperms.

The Future of *Arabidopsis* Research

For *Arabidopsis* to be truly useful as a tool for molecular genetic research, two additional techniques must be developed. First, it must become possible to clone genes about which no more is known than their mutant phenotype. Second, it must become possible to take cloned genes that have been modified in vitro and introduce them back into the plant to assay their in vivo function. Development of the second of these techniques is currently being pursued by many laboratories using various methods. *Arabidopsis* is known to be susceptible to infection by *Agrobacterium tumefaciens*, and it is known that Ti-plasmid strains of *Agrobacterium* cause typical tumors on *Arabidopsis* (*36*). It will very likely prove possible to introduce cloned sequences into the *Arabidopsis* genome by using the same methods currently used in the transformation of other plants. If a method can be found that works efficiently enough, the small genome of *Arabidopsis* may be used to advantage in isolation of genes by so-called shotgun transformation.

Gene isolation would involve transformation of mutant plants or plant cells with a complete random recombinant library derived from wild-type DNA cloned in a Ti-plasmid-based vector, followed by assay of individual transformed cells or plants for complementation of the mutant phenotype by the expression of the introduced DNA. Using a cosmid vector with a capacity of 20 to 25 kb, it would require approximately 3000 transformation events to introduce one genome equivalent of wild-type DNA into a set of mutant plants or cells and 14,000 transformation events to introduce the 4.6 genomes necessary to have a 99 percent chance of recovering any specific gene. Even with an efficient technique, these are large numbers of transformation events unless the gene of interest has a phenotype that can be selected directly.

An alternative approach would be to use a procedure of successive isolations of overlapping cloned segments starting from a known cloned genetic location and proceeding to any nearby genetic locus. The *Arabidopsis* genome is unique among those of flowering plants in its suitability for this sort of process, owing

both to its small size and to the near absence of dispersed repeated sequences. No other flowering plant is known to have a genome with few enough of these sequences, spaced far enough from each other, to allow such a procedure to be practical (Table 2).

For a chromosome walking procedure to be used, it is necessary to have starting points that are known to be located relatively close to the gene of interest. To provide these starting points, we are currently involved in the production of a genetic map using restriction fragment length polymorphisms (RFLP's) (*37*). It is easy to find RFLP's in different wild-type strains of *Arabidopsis*. We have tested a number of different strains and have selected the Niederzenz strain (first collected in Niederzenz, West Germany) and the Landsberg strain (first collected in Landsberg, East Germany) as the parental strains.

Landsberg was chosen because most of the mutations used in construction of the published genetic maps of *Arabidopsis* are in the Landsberg genetic background. Niederzenz was selected as the second parental strain because it grows as quickly as Landsberg (many other wild strains do not) and because it shows many RFLP's when compared to Landsberg. We have crossed these two parental strains and allowed the F_1 progeny to self-fertilize to form F_2 seed. F_2 plants were grown and allowed to self, and F_3 seeds were collected from each plant. These pools of F_3 seeds have been grown and DNA has been prepared from the resulting plants. This DNA contains the alleles present in the F_2 plant from which the F_3 plants descended. By genome blot analysis of the DNA, it will be possible to determine the genotype of each of the F_2 plants at each polymorphic locus. From these data, linkage distances can be calculated just as for any other genetic markers. By including visible markers in the original cross and performing subsequent crosses with other visible markers, it will be possible to align the RFLP map with the published genetic map of visible mutations. The clones used as probes on the RFLP genome blots can then serve as starting points for successive clone isolations.

To tell when the desired gene has been cloned, it will be necessary to transform cloned DNA segments into the plant and to assay them for the ability to complement mutations in the gene. Because only a small number of clones close to the gene of interest need to be tested by introduction into the plant genome, this procedure will not require the high frequency of transformation demanded by the first general gene isolation method described.

The ability to transform *Arabidopsis* not only has the potential to permit cloning of any gene having a known mutant phenotype but will also make possible the type of detailed analysis of these genes that is now being performed with genes of yeast, *Drosophila*, and mice. Among the genes that may be cloned and analyzed are those affecting particular enzyme activities, those with specific effects on the ability of *Arabidopsis* to synthesize and respond to plant hormones, and those in which mutations give specific developmental abnormalities. Our hope is that *Arabidopsis* will soon join the other organisms for which a combined genetic and molecular approach has led to both fundamental and practical scientific advances.

References and Notes

1. R. Flavell, *Annu. Rev. Plant Physiol.* **31**, 569 (1980).
2. F. Laibach, *Bot. Arch.* **44**, 439 (1943); *Naturwissenschaften* **31**, 246 (1943); E. Reinholz, *ibid.* **34**, 26 (1947).
3. G. P. Redei, *Annu. Rev. Genet.* **9**, 111 (1975).
4. L. M. W. Dellaert, J. Y. P. Van Es, M. Koornneef, Arabidopsis *Inf. Serv.* **16**, 10 (1979).
5. W. J. Feenstra, *ibid.* **15**, 35 (1978); S. Lee-Chen and L. M. Steinitz-Sears, *Can. J. Genet. Cytol.* **9**, 381 (1967); M. Koornneef, L. W. M. Dellaert, J. H. van der Veen, *Mutat. Res.* **93**, 109 (1982).
6. M. Koornneef *et al.*, *J. Hered.* **74**, 265 (1983).
7. G. P. Redei, *Z. Vererbungsl.* **93**, 164 (1962).
8. D. W. Meinke and I. M. Sussex, *Dev. Biol.* **72**, 50 (1979); D. W. Meinke, *Theor. Appl. Genet.* **69**, 543 (1985).
9. G. P. Redei, *Genetics* **56**, 431 (1967); G. Robbelen, *Planta (Berlin)* **80**, 237 (1968).
10. G. P. Redei, *Mutat. Res.* **18**, 149 (1973).
11. J. Langridge, *Nature (London)* **176**, 260 (1955).
12. S. L. Li and G. P. Redei, *Biochem. Genet.* **3**, 163 (1969).
13. C. R. Somerville and W. L. Ogren, *Nature (London)* **286**, 257 (1980); *ibid.* **280**, 833 (1979); *Biochem. J.* **202**, 373 (1982); *Plant Physiol.* **67**, 666 (1981); *Proc. Natl. Acad. Sci. U.S.A.* **77**, 2684 (1980).
14. F. J. Braaksma and W. J. Feenstra, *Theor. Appl. Genet.* **64**, 83 (1982); M. Jacobs and D. Schwartz, Arabidopsis *Inf. Serv.* **17**, 88 (1980).
15. M. Koornneef and J. H. van der Veen, *Theor. Appl. Genet.* **58**, 257 (1980); M. Koornneef, M. L. Jorna, D. L. C. Brinkhorst–van der Swan, C. M. Karssen, *ibid.* **61**, 385 (1982); M. Koornneef, G. Reuling, C. M. Karssen, *Physiol. Plant.* **61**, 377 (1984).
16. C. M. Karssen, D. L. C. Brinkhorst–van der Swan, A. E. Breekland, M. Koornneef, *Planta* **157**, 158 (1983).
17. J. I. Mirza, G. M. Olsen, T.-H. Iversen, E. P. Maher, *Physiol. Plant.* **60**, 516 (1984); G. M. Olsen, J. I. Mirza, E. P. Maher, T.-H. Iversen, *ibid.*, p. 523.
18. M. Koornneef and J. H. van der Veen, *Genetica* **61**, 41 (1983).
19. The seed collection is maintained by A. R. Kranz, Botanisches Institut, J. W. Goethe-Universität, Frankfurt am Main 11, West Germany. Kranz also edits and distributes Arabidopsis *Information Service*, an annual newsletter that includes original contributions, reviews, reports of new mutations, and lists of available stocks.
20. G. P. Redei and C. M. Perry, Arabidopsis *Inf. Serv.* **8**, 34 (1971); N. Goto, *ibid.* **19**, 55 (1982).
21. I. Negrutiu, M. Jacobs, W. de Greef, *Z. Pflanzenphysiol.* **90**, 363 (1978).
22. M. D. Bennett and J. B. Smith, *Proc. R. Soc. London Ser. B* **274**, 227 (1976).
23. A. H. Sparrow, H. J. Price, A. G. Underbrink, *Brookhaven Symp. Biol.* **223**, 451 (1972).
24. L. S. Leutwiler, B. R. Hough-Evans, E. M. Meyerowitz, *Mol. Gen. Genet.* **194**, 15 (1984).
25. G. D. Lauer, T. M. Roberts, L. C. Klotz, *J. Mol. Biol.* **114**, 507 (1977).
26. R. E. Pruitt and E. M. Meyerowitz, unpublished manuscript.
27. R. E. Pruitt, D. Ruff, E. M. Meyerowitz, in preparation.
28. A. E. Simon, K. M. Tenbarge, S. R. Scofield, R. R. Finkelstein, M. L. Crouch, unpublished manuscript.
29. L. S. Leutwiler, E. M. Meyerowitz, E. M. Tobin, in preparation.
30. W. J. Stiekema, C. F. Wimpee, J. Silverthorne, E. M. Tobin, *Plant Physiol.* **72**, 717 (1983).
31. P. Dunsmuir, S. M. Smith, J. Bedbrook, *J. Mol. Appl. Genet.* **2**, 285 (1983).
32. G. An, personal communication.
33. C. Chang and E. M. Meyerowitz, in preparation.
34. R. Dolferus and M. Jacobs, *Biochem. Genet.* **22**, 817 (1984).
35. E. S. Dennis *et al.*, *Nucleic Acids Res.* **12**, 3983 (1984); E. S. Dennis, M. M. Sachs, W. L. Gerlach, E. J. Finnegan, W. J. Peacock, *ibid.* **13**, 727 (1985).
36. M. Aerts, M. Jacobs, J.-P. Hernalsteens, M. van Montagu, J. Schell, *Plant Sci. Lett.* **17**, 43 (1979).
37. D. Botstein, R. L. White, M. Skolnick, R. W. Davis, *Am. J. Hum. Genet.* **32**, 314 (1980).
38. The sources of the genome measurements are *Arabidopsis*: (*24*); mung bean: M. G. Murray, J. D. Palmer, W. F. Thompson, *Biochemistry* **18**, 5259 (1979); cotton: V. Walbot, L. S. Dure III, *J. Mol. Biol.* **101**, 503 (1976); tobacco: J. L. Zimmerman and R. B. Goldberg, *Chromosoma (Berlin)* **59**, 227 (1977); soybean: R. B. Goldberg, *Biochem. Genet.* **16**, 45 (1978); pea: M. G. Murray, R. E. Cuellar, W. F. Thompson, *Biochemistry* **17**, 5781 (1978); wheat: D. B. Smith and R. B. Flavell, *Chromosoma (Berlin)* **50**, 223 (1975) and R. B. Flavell and D. B. Smith, *Heredity* **37**, 231 (1976).

39. We thank the other members of the laboratory of E.M.M. for comments on the manuscript. Our *Arabidopsis* work is supported by NSF grant PCM-8408504 to E.M.M.