MOLECULAR GENETIC STUDIES IN *Arabidopsis thaliana*:

I. Cloning and Characterization of the Alcohol Dehydrogenase Gene;
II. Complementation of an Alcohol Dehydrogenase Mutant;
III. A Restriction Fragment Length Polymorphism Map of the Genome

Thesis by

Caren Chang

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1988

(Submitted June 3, 1988)

This thesis is dedicated to the memory of my mother

who was always a source of inspiration to me.

ACKNOWLEDGEMENTS

I am highly indebted to my thesis advisor Elliot Meyerowitz for his knowledge, enthusiasm, and encouragement throughout this work. It has been a pleasure and an honor to be part of his lab. I also thank the other members of my thesis committee: Giuseppe Attardi, Norman Davidson, Melvin Simon and Barbara Wold.

I am grateful to Bob Pruitt and Leslie Leutwiler for breaking ground in *Arabidopsis* molecular genetics which provided the foundation for this thesis. I am also grateful to those who shared in the RFLP mapping; without John Bowman and Art DeJohn the map would not have been realized. These guys were terrific to work with, and they almost managed to make a good time out of 10^2 DNA preps and 10^2 hybridizations. Michael Ahn and Sherry Kempin contributed toward getting the last few markers onto the map. I also want to thank Sherry for producing large stocks for most of the 90 markers. Eric Lander and Mark Daly generously and speedily computed the final linkage map. Thank you 'Turbo Man' Stan Tamaki, wherever you are, for your special appearances from UC Riverside. Most importantly, without occasional invasions of Mark Garfinkel's bench, none of the RFLP work would have been possible.

I thank all the members of the Meyerowitz lab for making life in the lab very enjoyable, and for contributing to my enlightenment. David Smyth gave me excellent advice on several aspects of this thesis within the short time I have known him. John, Mark, Chris Martin, Patty Pang, Marty Yanofsky, Hong Ma, Tony Bleeker, and Pete Mathers provided valuable discussions as well as comaraderie. Thanks to Mike Palazzolo for words of encouragement; no ice machine in my future will ever replace the one in 108.

Finally, I thank Art for motivating me to think about graduating and to think about getting on with other important things in life such as him.

(My graduate studies were supported partially by National Institutes of Health training grant 5T32GM07616 and partially by National Science Foundation grant DCB-8703439.)

ABSTRACT

This thesis examines several genetic aspects of the flowering plant *Arabidopsis thaliana* which has previously been shown to possess several attributes for molecular genetic experiments such as a short life cycle, small size, fecundity, and a strong background of classical genetics. In particular, the nuclear genome is remarkably small and consists almost entirely of single copy sequences.

Two approaches are established for the isolation and study of plant genes using *Arabidopsis* as a model system. One approach demonstrates complementation of a mutant phenotype using the *Arabidopsis* alcohol dehydrogenase (ADH) gene. The ADH gene was first isolated from a genomic DNA library by cross-hybridization with a maize ADH1 gene probe. The gene structure was studied by DNA sequence analysis and mapping of the transcript. The gene conferred wild-type ADH activity to an *Arabidopsis* ADH null mutant when introduced by *Agrobacterium*-mediated transformation. Transformed plants were subjected to genetic and molecular analyses. The ADH gene may have utility as a gene tag in transformation experiments.

The second topic of this thesis is the construction of a genetic linkage map of restriction fragment length polymorphisms (RFLPs). The map provides an approach to cloning genes about which nothing more is known than a mutant phenotype and a map location because the RFLP markers can serve as starting points for the isolation of overlapping clones from a genomic DNA library. The map contains 90 RFLP markers distributed randomly throughout the nuclear genome. Since the genome consists of about 70,000 kilobase pairs, the markers are at an average physical spacing of approximately 780 kilobase pairs. The map is based on meiotic segregation of RFLPs in two different crosses detected with the restriction enzymes *Eco*RI, *Bgl*II, and *Xba*I. The RFLP linkage groups have been aligned with the standard genetic map of approximately 80 mutation markers.

TABLE OF CONTENTS

INTRODUCTION

Molecular Biology of Plant Growth and Development: *Arabidopsis thaliana* as an Experimental System

(Updated version of a chapter published in
Developmental Biology: A Comprehensive Synthesis,
vol. 5, ed. L. Browder, Plenum, New York, 1988.)

Introduction

Plant development differs from animal development in several fundamental respects. Since plant cells are immobilized in a rigid cell wall, morphogenesis is dependent upon control of cell growth and plane of cell division, rather than cell migration as occurs in animal development. The meristematic cells of plants display a degree of plasticity not found in animal cells; they remain embryonic throughout the life of the plant, and produce both adult organs and germ cells. Further, individual differentiated cells from vegetative plant parts are able to dedifferentiate and regenerate into new, fertile plants.

Arabidopsis thaliana is a flowering plant that has many characteristics that recommend it for genetic and molecular studies of development, much as certain attributes of *Drosophila* recommend it for such studies. In fact, *A. thaliana* was recognized as a "botanical *Drosophila*" in the 1940s (Whyte, 1946) because of its low chromosome number, short generation time, ease of culturing, and high fecundity. Over the past 40 years an extensive literature has accumulated on the genetics and ecology of this member of the mustard family (reviewed by Rédei, 1970, 1975a, 1975b). Recently, molecular characterization of the *Arabidopsis* genome has shown that the plant has an extraordinarily small and simple genome, properties that facilitate the molecular analysis of development (Leutwiler *et al.*, 1984; Pruitt and Meyerowitz, 1986). In this chapter, we provide an overview of *Arabidopsis thaliana* as an experimental system by describing the classical and molecular genetics of the plant and discussing the attributes which make it a valuable model system for genetic and molecular studies of plant development.

General Description

Arabidopsis thaliana is a typical member of the mustard family (Cruciferae). Though a number of other crucifers, such as cabbage and turnips, are crops, *Arabidopsis* has no

known food or other economic value. The mature plant is small, but is as morphologically complex as other flowering plants (Figure 1, 2). Typically about 30 cm tall, the mature plant consists of a rosette of small leaves surrounding a main stem. Flowers in racemes are at the apex of the stem. Lateral branches, arising from the main stem as the plant ages, also develop apical inflorescences. The flowers are bisexual and are approximately 3 mm long and 1 mm wide. They consist of four sepals surrounding four white petals, which are alternate with the sepals; within the petals are six stamens (two short and four long) and a pistil. The anthers of the longer stamens rise above the stigma surface and release pollen, while the two shorter stamens are not thought to contribute to fertilization (Meinke and Sussex, 1979a). The pistil consists of two fused carpels with parietal placentation (that is, the ovules are borne on the ovary wall), and the ovary is divided into two chambers by a vertical partition of complex origin. After fertilization the ovary matures to a fruit called a silique, which is 12 to 16 mm in length and contains 30 to 60 seeds arranged in two rows. A single plant can produce several hundred siliques and may therefore yield more than 10,000 seeds. The seeds are very small, measuring approximately 0.5 by 0.3 by 0.2 mm, and weighing 16 to 20 μg. Detailed descriptions of the wild type plant are given by Röbbelen (1957a), Müller (1961), and Napp-Zinn (1963).

The length of the *A. thaliana* life cycle depends on growth conditions and genetic background. A typical time from germination to mature silique is five to six weeks for the most rapidly flowering ecotypes in long-day conditions. Seed germination occurs within a few days of the start of imbibition, followed by vegetative growth of rosette leaves. The primary stem bolts in two to three weeks, and the primary inflorescence begins its development shortly thereafter. As the plant ages, lateral branches develop from the leaf axils of the main stem, and occasionally also develop from the base of the plant. While the flowers normally self-fertilize, cross-pollination is easily effected under a dissecting microscope. In nature, up to several percent of the flowers may be cross-pollinated,

evidently by small insects and perhaps wind; under laboratory conditions, in the absence of insect infestation, spontaneous cross-fertilization is virtually absent. The siliques reach maturity about two weeks after pollination. The most apical fruits on each stem are the youngest, so that developmental sequences of seeds may be obtained simply by picking the fruits in the order in which they appear on each stem. If left on the plant, the siliques dry out, their valves open and vibration causes the seeds to drop. For experimental work, individual fruits may be harvested before they open, or the seeds of dried siliques can be harvested by inverting the plants and shaking off the seeds. An *Arabidopsis* plant will continue to grow and produce seed for months. Mature seeds remain germinable for years, and probably decades, if stored in dry conditions. A detailed description of the life cycle is given by Meinke and Sussex (1979a).

The exact taxonomic placement of *A. thaliana* is unclear; different taxonomic treatments of the mustards regard the plant differently. The exact geographical origin of *Arabidopsis* is also unclear. Most collections have been made in Europe and Asia, and the plant is thought to be native to the Old World. However, *Arabidopsis* has been found on all continents (except the Antarctic) in a wide variety of environments, ranging from the temperate Himalayas to semi-desert regions to high elevations of the tropics. A great deal of genetic variation is found in wild populations even within small geographical areas; different populations vary in responses to day length, requirements for vernalization, times to flowering, and leaf morphology and color. Many ecotypes that are adapted to a wide range of ecological niches have been collected and described (Röbbelen, 1965); seeds from these collections are among those available to researchers from the seed bank maintained at the Goethe University in Frankfurt, Federal Republic of Germany, by Prof. Dr. A. R. Kranz (Kranz, 1978; Kirchheim and Kranz, 1981).

Growing *Arabidopsis* in the laboratory is not difficult. Several dozen plants can be grown to maturity in a 5 or 6 cm diameter pot. The plants grow well in soil with occasional

watering, and will also grow in a variety of defined simple media (Langridge, 1957). Rapid growth occurs in a broad range of temperatures (14 to 26°C); above 30°C the number of seeds per silique is reduced, and at colder temperatures growth and maturation are slow. Continuous lighting produces the most rapid growth to maturity; in short days the rosette stage lasts several times longer than in long days, resulting in much larger rosettes. Growth in short-day conditions is thus a method for producing large amounts of plant tissue from individual plants. Plants can be grown aseptically in either solid or liquid media. Whole plants can be regenerated from callus at high frequency (Feldmann and Marks, 1986). Haploid callus and plants have been generated by anther culture (Gresshof and Doy, 1972).

Classical Genetics

Arabidopsis thaliana has a haploid chromosome number of five (Laibach, 1907). The small chromosomes, ranging in length from 1.1 to 3.7 micrometers at meiotic telophase I, can be distinguished by their different lengths and arm ratios (Steinitz-Sears, 1963).

The classical genetic analysis of the plant started about 40 years ago with the isolation of the first induced mutations by Reinholz (1947) who used X rays as the mutagenic agent. It has since been shown that a large number of chemical mutagens are effective in the induction of new mutations (Rédei, 1970; Ehrenberg, 1971). Since *Arabidopsis* is self-fertilizing and matures rapidly, it has been a convenient plant for evaluating the effects of mutagenic agents. This is exemplified by the embryo-lethal assay of Müller (1963) whereby plants grown from mutagenized seeds (M_1 seeds) are screened for siliques that contain approximately 25% aborted and 75% normal M_2 seed. Such siliques result from the induction of recessive embryo-lethal mutations in the cells of the M_1 seed that give rise to flowers. When the flowers self-fertilize, one quarter of the seeds that develop are homozygous for the newly induced mutation, while two thirds of the phenotypically normal

seeds maintain the mutation in a heterozygous condition. Since each seed contains several cells that will develop into reproductive structures, new mutations are normally limited to one sector of the mature plant; only the first five siliques need to be scored to assay all of the sectors. The embryo-lethal screen is simple and reliable: aborted seeds are distinct in color and size in both immature and mature siliques, and the frequency of spontaneous abortion in *Arabidopsis* is low and unaffected by fluctuations in temperature, water, and nutrient supply (Meinke and Sussex, 1979a).

Rédei (1970) has reviewed the genetic effects of about 100 mutagens on *A. thaliana*, measured by various techniques. Ethyl methanesulfonate (EMS) is a very effective mutagen giving a high ratio of mutants to nonsurvivors, in addition to producing a large number of visible mutations relative to induced sterility. Soaking dry or imbibed seeds in solutions of EMS for several hours is sufficient for mutagenesis. Other methanesulfonic esters (hydroxyethyl, methoxy, and methoxyethyl) are comparable to EMS. Certain nitrosoguanidines, nitrosoamines, and nitrosoureas also serve as effective mutagens. While not as efficient as chemical mutagenesis, ionizing radiation applied to seeds has also been used successfully to induce mutations. *A. thaliana* is highly resistant to ionizing radiation, presumably because its extremely small nuclear volume (chromosome volume) presents such a tiny target (Sparrow, 1964). This resistance was initially observed by Reinholz (1947). As a result of these and other studies, a wide spectrum of morphological and physiological mutations have been induced (Reinholz, 1947; Röbbelen, 1957b; McKelvie, 1962, 1963; Bürger, 1971; Koornneef *et al.*, 1983), a few of which will be discussed here to provide an introduction to the range of mutations available for developmental, physiological, and molecular genetic research.

The easiest mutations to identify are those that have a visible or lethal phenotype. Visible mutations affecting every part of the plant have been identified. There are mutations affecting plant and embryo color (*albina, chlorina, sulfurata*), leaf morphology (*serrate,*

angustifolia), trichome morphology (*distorted trichomes*, *glabra*), and plant size (*erecta*, *compacta*, *miniature*), to name only a few of the mapped mutations. Another mapped mutation, *immutans*, interferes with plastid differentiation such that some cells are normal (green) while other cells lack both chlorophylls and carotenoids (white) (Rédei, 1967). This mutation is suppressed by 6-azauracil or low light intensity. *Chloroplast mutator* also causes variegated leaves; mutations in this nuclear gene induce plastid mutations that are then cytoplasmically inherited (Rédei, 1973).

Many mutations are known to affect floral development in ways that are of particular interest to developmental biologists, because these mutations cause homeotic transformations of floral organs. For example, *apetala-2* is a recessive mutation that shows a conversion of petals to stamens that ranges from partial to complete, and an apparent conversion of sepals to leaves. The flowers thus have ten stamens, no petals or sepals, and are surrounded by four leaves. Another striking homeotic transformation is shown by *apetala-3* homozygotes, in which the petals are converted to sepals, and each of the six stamens is changed to a single unpaired carpel having what appear to be ovules attached to its margin. In *pistillata*, petals are replaced by sepals, and anthers are absent. There are other floral mutations that cause fundamental changes in floral structure that are meristic rather than homeotic. For example, *agamous* flowers develop a large and indefinite number of sepals and petals, but no stamens or carpels. Another such mutant is *apetala-1*, in which flowers have rudimentary or absent petals, and frequently have flower buds arising from within the whorl of sepals. The floral mutations described here, and others as well, have been mapped to their chromosomal locations (Koornneef *et al.*, 1983; Bowman and Meyerowitz, unpublished). Several of these mutations exhibit temperature sensitivity (Bowman and Meyerowitz, unpublished).

Another group of developmental mutations has been isolated using Müller's embryo-lethal screen. Isolation and characterization of embryo-lethal mutants is the first step in

identifying essential functions at the various stages of plant embryogenesis. In maize, defective kernel mutants and viviparous mutants have been studied in detail; in carrot, somatic embryogenesis *in vitro* has been examined in mutant cell lines. In *Arabidopsis*, about 40 EMS-induced embryo-lethal mutants have been analyzed, representing a wide range of phenotypes (Meinke and Sussex, 1979b; Meinke, 1985). These aborted seeds and embryos differ with respect to color, size, stage of developmental arrest, and extent of abnormal development. Evidence for the action of essential genes in the gametophyte, prior to fertilization, is provided by the nonrandom distribution of aborted seeds within the siliques of ten of the mutants (Meinke, 1982; Meinke, 1985). The low proportion of aborted embryos in some of the mutants indicates an effect on fertilization. Other embryo-lethal mutants have been used for ultrastructural analysis (Marsden and Meinke, 1985) and for embyro culture. From the culture of arrested embryos, Meinke *et al.* (1985) were able to produce homozygous plants exhibiting mutant phenotypes. By culturing arrested embryos on a nutrient medium it should be possible to identify nutritional lethals and to rescue them, while mutants with a cellular basis would not be rescued in this manner; developmental lethals should not be able to differentiate in culture, but should be able to produce callus. In addition to their use in understanding essential developmental processes, embryo lethals may be useful for studying chlorophyll synthesis during embryogenesis since defects in chlorophyll and carotenoid pigments, i.e. white or pale green arrested seeds and embryos, are commonly seen, but mutants for pigmentation apparently undergo normal embryo development.

In addition to mutations causing morphological abnormalities and arrested embryonic development, a number of mutants have been obtained that affect synthesis of and response to phytohormones. Mutations at five different genetic loci in *A. thaliana* are known to result in plants that are deficient in gibberellins. The phenotype of some of the mutations is failure of germination; those that allow germination cause dwarfing (Koornneef and van der Veen,

1980). When exogenous gibberellins are sprayed on the mutant plants, they are restored to normal or near-normal phenotype. Abscisic acid deficient mutants have been obtained by selecting for second-site revertants of the nongerminating gibberellin mutations, based on the rationale that seed germination is stimulated by gibberellins, but inhibited by abscisic acid (Koornneef *et al.*, 1982). The phenotype of many of the abscisic acid-deficient mutants is precocious seed germination and wilting of maturing plants. Phytohormone-insensitive mutants have also been isolated. Mutations in at least three loci confer resistance to high levels of abscisic acid; these mutants resemble the abscisin-deficient mutants in phenotype, except that endogenous abscisin levels are normal or greater than normal (Koornneef *et al.*, 1984). Mutants resistant to the auxin 2,4-dichlorophenoxyacetic acid, including a dominant mutation, have been mapped to two loci (Maher and Martindale, 1980). Alleles at both loci display increased rates of root growth and altered geotropism (Mirza *et al.*, 1984). A dominant mutation that confers insensitivity to ethylene has been isolated which lacks a number of diverse ethylene responses such as leaf senescence and inhibition of cell elongation (Bleeker *et al.*, 1988).

Mutations affecting the levels of activity of a number of specific enzymes are also known. Mutants with reduced nitrate reductase activity have been selected by isolating plants resistant to chlorate, which is converted to toxic chlorite by nitrate reductase (Braaksma and Feenstra, 1982). Some of these mutations affect chlorate uptake, others reduce nitrate reductase activity by affecting its regulation or the molybdenum cofactor necessary for its function. Mutants lacking alcohol dehydrogenase have been selected by isolating plants resistant to allyl alcohol, which is converted to a toxic aldehyde by alcohol dehydrogenase. These recessive mutants appear to have genetic lesions in the structural gene for the enzyme (Negrutiu *et al.*, 1984). Dominant sulfonylurea-resistant mutations have been isolated which have a defect in acetohydroxy acid synthase which catalyzes the first enzymatic step common to the biosynthesis of the branched-chain amino acids (Haughn

and Somerville, 1986). These mutants are resistant to concentrations of sulfonylurea herbicides up to approximately 300-fold greater than that required to inhibit growth of the wild-type. A number of mutations affect the levels of enzymes in the photorespiration pathway; among the affected enzymes are glutamate synthase, glycine decarboxylase, phosphoglycolate phosphatase, serine-glyoxylate aminotransferase and serine transhydroxymethylase (Somerville and Ogren, 1982). It is not yet known if any of these mutations are in the structural genes for the enzymes. Other biochemical mutations are known in *Arabidopsis*, including mutations in several loci that cause the plant to require exogenous thiamine or thiamine precursors (Langridge, 1955; Li and Rédei, 1969). Some of these conditional lethals are temperature-sensitive. It is likely that temperature-sensitive mutations and other conditional mutations can be isolated for different genes.

Recently, the first amino acid auxotroph on the whole plant level was isolated in *Arabidopsis* (Last and Fink, 1988). The mutation is in the biosynthesis of tryptophan from anthranilate and was isolated by selection of plants on 5-methylanthranilic acid which is converted to toxic 5-methyltryptophan by the wild-type enzyme activity.

Many of the known *Arabidopsis* mutations have been mapped to their positions on the five *Arabidopsis* linkage groups, which correspond to the five chromosomes. Koornneef *et al.* (1983) have published a map showing the genetic locations of 76 loci; this map is a compilation of all previously published linkage information and a great deal of new linkage data. The correpondence between linkage groups and chromosomes has been established by analysis of trisomic lines (Koornneef and van der Veen, 1983); centromeres have been mapped by study of complementation in telotrisomic lines. The entire genetic map currently measures approximately 430 centiMorgans.

The *Arabidopsis* Genome

While *Arabidopsis* thus has a large number of useful and developmentally interesting mutations, it is not distinguished from other experimentally useful angiosperms such as maize and tomato in this respect. Although it is easier and quicker to isolate new mutations in *Arabidopsis* than in these other plants, the properties that truly set *Arabidopsis* apart are those that are important for experiments in molecular genetics, primarily an extremely small genome and paucity of dispersed repetitive DNA in the nuclear genome.

Arabidopsis has the smallest known angiosperm genome. The haploid nuclear genome of *Arabidopsis* is 70,000 kilobase pairs (kb), as measured both by DNA reassociation kinetic analysis (Leutwiler *et al.*, 1984) and by quantitative genome blot hybridization (Pruitt and Meyerowitz, 1986). This is only five times the genome size of the yeast, *Saccharomyces cerevisiae*, and only twenty times the size of the *E. coli* genome. Plants that are frequently used in molecular genetic experiments have much larger genomes; tobacco and wheat, for example, have genomes of 1,600,000 and 5,900,000 kb, respectively (Zimmerman and Goldberg, 1977; Flavell and Smith, 1976).

The importance of the tiny *Arabidopsis* genome is that it simplifies the molecular cloning of any specific gene. First, only 16,000 random lambda clones need be screened with any probe to have a 99% chance of finding the desired DNA in the library. In contrast, 370,000 clones would be required for tobacco and 1,400,000 clones for wheat (Meyerowitz and Pruitt, 1985). Second, in genome blot experiments with heterologous probes, the signal-to-noise ratio is much higher than in other plants because of the low complexity of the *Arabidopsis* genome. This allows very diverged probes to be tested for hybridization to an angiosperm genome. Many examples will be described later.

The 70,000 kb of *Arabidopsis* haploid nuclear DNA has been shown to be largely single-copy by solution hybridization kinetic measurements (Leutwiler *et al.*, 1984). The

organization of the genomic DNA has been examined by molecular characterization of 50 lambda clones chosen randomly from an *Arabidopsis* genomic library made from whole-plant DNA (Pruitt and Meyerowitz, 1986). These clones represent 0.8% of the nuclear genome; they were analyzed by a variety of restriction endonclease digestions and gel-blotting procedures designed to reveal the nature and interspersion of repetitive sequence elements in the genome. In agreement with the results obtained in reassociation experiments, the majority of the clones, 34 out of 50, contained low copy number sequences; 32 of these were entirely unique, while two contained restriction fragments that may be present twice in the genome. The 16 remaining random clones contained DNA belonging to the middle repetitive sequence class. Eight of these clones contained portions of the ribosomal DNA repeat unit, a 9.9 kb unit repeated 570 times per haploid genome, largely or entirely in tandem arrays, and comprising 7.5% of the nuclear DNA. Four of the clones were derived from the chloroplast genome, one was likely derived from the mitochondrion, and only three were nuclear sequences containing dispersed middle repetitive elements. From the frequency and length of the unique-sequence and middle repetitive sequence-containing clones, the average amount of single-copy DNA separating adjacent dispersed repeats was estimated to be 120 kb. This is very different from the DNA of the other angiosperms that have been analyzed: the average size in the predominant class of single-copy sequences in tobacco, for instance, is only 1.4 kb; in wheat it is 1 kb and in pea only 300 base pairs (summarized in Meyerowitz and Pruitt, 1985). The evolutionary significance of the small genome and long-period interspersion of repetitive DNA in the *Arabidopsis* genome is unclear. The experimental significance of the extraordinarily long sequence interspersion pattern in *Arabidopsis* is that it permits chromosome-walking experiments, that is, experiments in which successive isolations of overlapping cloned segments lead from an initial starting clone to any nearby genetic location. Only

Arabidopsis, among all known flowering plants, has a genome with a low enough frequency of interspersed repeats to allow such a procedure to be practically performed.

Molecular Genetics

A number of genes have already been cloned in *A. thaliana*, making it possible to begin using *Arabidopsis* for examining the developmental and environmental control of plant gene expression. Characterization of these genes provides another look at the *Arabidopsis* genome and, additionally, demonstrates its utility for certain molecular approaches. Most of these genes have been cloned using heterologous probes from a variety of organisms. With few exceptions, these cloning experiments show that genes having many copies in other plants exist in a low number of copies, or in only a single copy, in *Arabidopsis*; and that when multiple gene copies have been found in *Arabidopsis*, the copies are usually clustered.

The light harvesting chlorophyll a/b binding protein (LHCP) is a light-inducible, nuclear-encoded component of the light-harvesting antenna complex of the chloroplast thylakoid membranes. The *Arabidopsis* LHCP genes were cloned from a genomic recombinant library by cross hybridization with a LHCP gene probe from the aquatic monocot *Lemna gibba* (Leutwiler *et al.*, 1986). While other plants are known to contain LHCP gene families of seven to sixteen genes, the *Arabidopsis* family contains only three or four genes, with three clustered within a 6.5 kb region. The DNA sequence of the three open reading frames differ at 4% of the nucleotides. The deduced amino acid sequences encoded by the three clustered genes, excluding the transit peptides, are identical; thus the three copies of the LHCP genes do not provide protein heterogeneity.

The alcohol dehydrogenase (ADH) enzymes in higher plants are under environmental and developmental control; ADH activity has been localized in dry seeds, pollen, germinating roots, and anaerobically-treated seedling roots. In maize and *Arabidopsis*,

ADH is induced by anaerobic treatment and by the synthetic auxin 2,4-dichlorophenoxyacetic acid (Dennis *et al.*, 1985; Freeling, 1973; Dolferus *et al.*, 1985). The *Arabidopsis* ADH gene was cloned from a genomic recombinant library by cross hybridization with a maize ADH1 gene probe as described in Chapter One of this thesis (Chang and Meyerowitz, 1986). There is only a single ADH gene in *Arabidopsis* while most plants have two or more ADH genes. The *Arabidopsis* ADH protein sequence (deduced from the DNA sequence) and the maize ADH protein sequences are the same length and are approximately 80% conserved; and the six intervening sequence positions of the *Arabidopsis* ADH gene coincide with six of the nine intervening sequences in both of the maize genes.

Developing plant embryos store nitrogen in the form of various seed storage proteins, such as the 12S globulin storage protein in *Arabidopsis*, which accumulate to high levels in the latter half of embryogenesis and are rapidly broken down during germination. Genes that code for abundant seed RNAs have been isolated from a genomic recombinant library by cross hybridization with *A. thaliana* seed pod cDNA probes (Pang *et al.*, 1988). These clones fall into four homology classes based on cross hybridization and restriction fragment patterns. One class contains the *Arabidopsis* 12S storage protein gene which was identified by homology with a 12S storage protein cDNA clone of the crucifer *Brassica napus*. While there appears to be only one copy in this class of the 12S gene in the Columbia and Bensheim ecotypes of *Arabidopsis*, there are two tandemly-arranged 12S genes in the Landsberg *erecta* ecotype. The 5' upstream DNA sequences of the two tightly clustered genes are not conserved except in the immediate TATA-box region. Such close clustering of genes has not been observed for the 12S protein gene families of other plant species, nor has a gene copy number as low as one or two. The amino acid sequence (deduced from the DNA sequence of one of the two genes) is homologous with both the *B.napus* and the pea

12S storage protein sequences. The intron positions of the *Arabidopsis* gene are coincident with those of the pea gene.

Genes from very divergent species have been used with success to clone homologous genes in *Arabidopsis*. The *Saccharomyces cerevisiae* acetolactate synthase gene was used to clone the *Arabidopsis* acetolactate synthase gene (Haughn *et al.*, 1988). In addition, the mutant gene of acetolactate synthase, which differs from the wild-type gene by a single base pair substitution, was isolated from the sulfonylurea-resistant mutant (Haughn *et al.*, 1988; Haughn and Somerville, 1986). Three genes that were cloned by homology to a *Drosophila melanogaster* 70 kD heat shock gene clone resemble both the maize and *Drosophila* 70 kD heat shock genes (C. Somerville, personal communication). One of the genes is inducible by exposure to elevated temperatures. A *Chlamydomonas reinhardtii* gene for alpha-tubulin was used to isolate two homologous genes from *Arabidopsis* (Ludwig *et al.*, 1987). The protein sequences have 90% identity.

A number of other *Arabidopsis* genes have been cloned using gene probes from other plants. The nitrate reductase gene (Crawford, 1986) and the phytochrome gene (R. Sharrock, C. Gatz, and P. Quail, personal communication) were isolated using squash gene probes. The gene for glycollate oxidase was isolated using a spinach gene probe (C. Somerville, personal communication). The gene for ADP glucose pyrophosphorylase was isolated using a rice gene probe (J. Preiss and C. Somerville, personal communication). The gene for RNA polymerase II was isolated using a soybean gene probe (H. Klee, personal communication). The gene for 5-enolpyruvylshikimate-3-phosphate synthase was isolated with a petunia gene probe (Klee *et al.*, 1987b); the *Arabidopsis* gene contains six introns in the same positions as the introns of the petunia gene, but each of the *Arabidopsis* introns is smaller.

The number of genes in *Arabidopsis* is not always less than that in other plants. For example, there are four tightly-linked genes for the small subunit of ribulose bisphosphate

carboxylase, which were cloned by homology with the pea cDNA, but in other plants there are two to twelve genes. The beta-tubulin gene family in *Arabidopsis* contains seven to eleven genes (or pseudogenes) while some plant species appear to contain more copies (D. Weeks, personal communication; P. Snustad, personal communication).

Another approach for identifying developmentally important genes is to examine the differential expression of genes in specific tissues. In *Arabidopsis*, such work has been initiated by Mesnard *et al.* (1985) who constructed a cDNA library from leaf mRNA. The abundance and relative expression of random genes in two different organs were examined by using individual leaf cDNA clones to probe blots containing seed pod and leaf mRNA. In this manner, two clones were found that have mRNA levels five to ten times greater in pods than in leaves.

A number of additional approaches to the isolation of environmentally and developmentally regulated *Arabidopsis* genes are currently being taken in different laboratories. While no deep understanding of developmental or environmental regulation of transcription has yet been achieved, the ease with which molecular techniques can be applied to analysis of *Arabidopsis* genes assures that progress will continue to be made at a rapid pace.

The Future

The unique combination of molecular and genetic properties of *Arabidopsis* will soon make possible the cloning of *Arabidopsis* genes about which no more is known than their mutant phenotype and genetic map location. The importance of this for plant developmental biology is that this cloning will be the first step in understanding the wild-type functions of the products of genetic loci such as *apetala-2*, *apetala-3* and *pistillata*, and thus a first step in understanding organogenesis and pattern formation in plants. One way such cloning is

being attempted is to start with a clone shown to map near one of these developmental mutations in a meiotic recombination map, then to walk to the developmentally-active gene by successive isolations of overlapping clones.

There are two requirements for such a walk to succeed. One is that an appropriate starting clone must be available, the other is that there must be a way to ascertain when the goal of the walk has been achieved. As described in Chapter Three of this thesis, a high physical density restriction fragment length polymorphism (RFLP) map of the *Arabidopsis* genome has been constructed in order to provide these starting clones (Chang *et al.*, 1988). The map contains 90 molecular markers which are at an average spacing of about 780 kb; over half the genome is within 2 centiMorgans, or 270 kb of the mapped polymorphisms. The map of 501 centiMorgans is has been aligned with the standard genetic map of mutation markers. In several laboratories, mutations of interest are currently being mapped relative to the RFLP markers using appropriate crosses. In our laboratory, mapping has already been performed for a number of mutations in floral development and one in ethylene response, and the isolation of overlapping clones is in progress (Yanofsky, Bowman, Ma, and Meyerowitz, work in progress; Bleeker and Meyerowitz, work in progress). The near absence of dispersed repetitive DNA makes isolation of overlapping clones possible and rapid.

The RFLP map may soon be aligned with a library of contiguous cosmid clones for each *Arabidopsis* linkage group. A physical map is presently under construction (Hauge *et al.*, 1987) which is based on fine structure restriction mapping of cosmid clones with the aid of a computer. The ordered clones will greatly facilitate cloning of genes for which there is only a mapped phenotype. Physical mapping is also in progress in *Saccharomyces cerevisiae* (Olsen *et al.*, 1986) and in *Caenorhabditis elegans* (Coulson *et al.*, 1986) (which has a genome size similar to that of *Arabidopsis*).

The end of a walk can be determined by transforming the cloned DNA segments into the mutant plant, and assaying for complementation of the mutant phenotype. Both transformation and complementation of *Arabidopsis* have been demonstrated using the natural DNA transfer system of *Agrobacterium tumefaciens*. Lloyd *et al.* (1986) used a procedure that is very effective in tobacco and petunia, in which leaf explants are co-cultivated with *Agrobacterium*. Their *Agrobacterium* strain carried an altered Ti-plasmid that contained a chimeric hygromycin resistance gene. This gene was transferred into the plant genome by the bacteria, and after regeneration of fertile plants from transformed callus cells, the resistance gene was inherited as part of one of the plant's chromosomes. In a new, nontissue culture approach, Feldman and Marks (1987) co-cultivated imbibed seeds with an *Agrobacterium* strain containing an altered Ti plasmid that carried a chimeric kanamycin resistance gene, then simply grew plants from the seeds and screened F_2 seeds for kanamycin resistance. Chang and Meyerowitz (Chapter Two of this thesis) have shown complementation of an ADH null mutant with the cloned genomic ADH gene using the transformation method of Lloyd *et al.* (1986). The wild-type phenotype was fully restored, and the transferred DNA was stably inherited in a Mendelian fashion.

Transformation also presents the possibility of cloning genes in the near future by shotgun complementation. *Arabidopsis* would be especially advantageous for this approach due to the small size of its genome. However, a higher efficiency of transformation than is presently obtained in *Arabidopsis* is required. A model experiment in gene rescue was performed in petunia, a plant which can be transformed with high efficiency. Using a genomic library made from transformed *Arabidopsis* callus containing a single T-DNA insertion conferring kanamycin resistance, the resistance was successfully transferred to petunia plants by *Agrobacterium*-mediated transformation (Klee *et al.*, 1987a). In addition to genome size and transformation efficiency, other factors such as the randomness of the

genomic library, the efficiency of transfer from *E. coli* to *Agrobacterium*, and nonselectivity in transfer of sequences to the plant genome need to be characterized and optimized.

One final prospect for cloning developmentally important genes in *Arabidopsis* is by transposon tagging. By searching restriction fragment length polymorphisms over a one hundred kilobase region in 16 different ecotypes, Voytas and Ausebel (1987) have recently identified a 2.3 kb DNA element that may represent a transposon. DNA sequence analysis showed that the element contains a long open reading frame that has 32% identity with the Copia retrotransposon of *Drosophila melanogaster*. The element is flanked by direct repeats suggestive of target duplications caused by transposons. It is not yet known that the element transposes.

Although the use of heterologous probes has been shown to be quite effective, even over broad evolutionary distances, the available methods of cloning *Arabidopsis* genes now extend beyond this approach. The fact that it is possible to clone genes for which there is only a mutant phenotype is very powerful. The set of ordered clones and possibly shotgun transformation and transposon tagging may offer additional approaches. Genes isolated in *Arabidopsis* may then be used to retrieve the homologous genes in other plants. Transformation and complemention in *Arabidopsis* allow the type of detailed analysis of gene structure and function that is currently in progress with the genes of yeast, *Drosophila*, and mice. Thus, *A. thaliana* may soon become the first plant to enter the group of organisms widely used as model systems for the molecular and genetic analysis of development.

References

Bleeker, A.B., Estelle, M.A., Somerville, C., and Kende, H. (1988). A dominant mutation confers insensitivity to ethylene in *Arabidopsis thaliana*. *Science*. In press.

Braaksma, F.J., and Feenstra, W.J. (1982). Isolation and characterization of nitrate reductase-deficient mutants of *Arabidopsis thaliana*. *Theor. Appl. Genet.* 64, 83-90.

Bürger, D. (1971). Die morphologischen Mutanten des Göttinger *Arabidopsis*-Sortiments, einschließlich der Mutanten mit abweichender Samenfarbe. *Arab. Inf. Serv.* 8, 36-42.

Chang, C., and Meyerowitz, E.M. (1986). Molecular cloning and DNA sequence of the *Arabidopsis thaliana* alcohol dehydrogenase gene. *Proc. Natl. Acad. Sci. USA* 83, 1208-1212.

Chang, C., Bowman, J.L., DeJohn, A.W., Lander, E.S., and Meyerowitz, E.M. (1988). An RFLP linkage map to facilitate gene cloning in *Arabidopsis thaliana*. Submitted.

Coulson, A., Sulston, J., Brenner, S, and Karn, J. (1986). Toward a physical map of the genome of the nematode *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* 83, 7821-7825.

Crawford, N.M., Campbell, W.H., and Davis, R.W. (1986). Nitrate reductase from squash: cDNA cloning and nitrate regulation. *Proc. Natl. Acad. Sci. USA* 83, 8073-8076.

Dennis, E.S., Sachs, M.M., Gerlach, W.L., Finnegan, E.J., and Peacock, W.J. (1985). Molecular analysis of the alcohol dehydrogenase 2 (Adh2) gene of maize. *Nucleic Acids Res.* 13, 727-743.

Dolferus, R., Marbaix, G., and Jacobs, M. (1985). Alcohol dehydrogenase in *Arabidopsis*: Analysis of the induction phenomenon in plantlets and tissue cultures. *Mol. Gen. Genet.* 199, 256-264.

Ehrenberg, L. (1971). Higher plants, in: *Chemical Mutagens*, (Hollaender, A., ed.), Plenum Press, New York, pp. 365-386.

Feldmann, K.A. and Marks. M.D. (1987). *Agrobacterium*-mediated transformation of germinating seed of *Arabidopsis thaliana*: a nontissue culture approach. *Mol. Gen.Genet.* 208, 1-9.

Flavell, R., and Smith, D. (1976). Nucleotide sequence organization in the wheat genome. *Heredity* 37, 231-252.

Freeling, M. (1973). Simultaneous induction by anaerobiosis or 2,4-D of multiple enzymes specified by two unlinked genes: Differential *Adh1-Adh2* expression in maize. *Mol. Gen. Genet.* 127, 215-227.

Gresshoff, P.M., and Doy, C.H. (1972). Haploid *Arabidopsis thaliana* callus and plants from anther culture. *Aust. J. Biol. Sci.* 25, 259-264.

Hauge, B., Fritze, C., Nam, H.-G., Paek, K.H., and Goodman, H.M. (1987). Abstract # 32, *Third International Meeting on Arabidopsis*, Michigan State University.

Haughn, G.W., Smith, J., Mazur, B., and Somerville, C. (1988). Transformation with a mutant *Arabidopsis* acetolactate synthase gene renders tobacco resistant to sulfonylurea herbicides. *Mol. Gen. Genet.* 211, 266-271.

Haughn, G.W., and Somerville, C. (1986). Sulfonylurea-resistant mutants of *Arabidopsis thaliana*. *Mol. Gen. Genet.* 204, 430-434.

Kirchheim, B., and Kranz, A.R. (1981). New population samples of the AIS-Seed Bank. *Arab. Inf. Serv.* 18, 173-176.

Klee, H.J., Hayford, M.B., and Rogers, S.G. (1987a). Gene rescue in plants: A model system for "shotgun" cloning by retransformation. *Mol. Gen. Genet.* 210, 282-287.

Klee, H.J., Muskopf, Y. M., and Gasser, C.S. (1987b). Cloning of an *Arabidopsis thaliana* gene encoding 5-enolpyruvylshikimate-3-phosphate synthase: sequence analysis and manipulation to obtain glyphosate-tolerant plants. *Mol. Gen. Genet.* 210, 437-442.

Koornneef, M., and van der Veen, J H. (1980). Induction and analysis of gibberellin sensitive mutants in *Arabisopsis thaliana* (L.) Heynh. *Theor. Appl. Genet.* 58, 257-263.

Koornneef, M., and van der Veen, J.H. (1983). The trisomics of *Arabidopsis* and the location of linkage groups. *Genetica* 61, 41-46.

Koornneef, M., Jorna, M.L., Brinkhorst-van der Swan, D. L. C., and Karssen, C. M. (1982). The isolation of abscisic acid (ABA) deficient mutants by selection of induced revertants in nongerminating gibberellin sensitive lines of *Arabidopsis thaliana* (L.) Heynh. *Theor. Appl. Genet.* 61, 385-393.

Koornneef, M., van Eden, J., Hanhart, C.J., Stam, P., Braaksma, F.J., Feenstra, W.J. (1983). Linkage map of *Arabidopsis thaliana. J. Hered.* 74, 265-272.

Koornneef, M., Reuling, G., and Karssen, C.M. (1984). The isolation and characterization of abscisic acid-insensitive mutants of *Arabidopsis thaliana. Physiol. Plant* 61, 377-383.

Kranz, A.R. (1978). Demonstration of new and additional population samples and mutant lines of the AIS-Seed Bank. *Arab. Inf. Serv.* 15, 118-139.

Laibach, F. (1907). Zur Frage nach der Individualität der Chromosomen im Pflanzenreich. *Beih. Bot. Cbl. 1 Abt.* 22, 191-210.

Langridge, J. (1955). Biochemical mutations in the crucifer *Arabidopsis thaliana* (L.) Heynh. *Nature* 176, 260-261.

Langridge, J. (1957). The aseptic culture of *Arabidopsis thaliana* (L.) Heynh. *Austral. J. Biol. Sci.* 10, 243-252.

Last, R.L., and Fink, G.R. (1988). Tryptophan-requiring mutants of the plant *Arabidopsis thaliana.. Science* 240, 305-310.

Leutwiler, L.S., Hough-Evans, B.R., and Meyerowitz, E.M. (1984). The DNA of *Arabidopsis thaliana*. *Mol. Gen. Genet.* 194, 15-23.

Leutwiler, L.S., Meyerowitz, E.M., and Tobin, E.M. (1986). Structure and expression of three light-harvesting chlorophyll a/b binding proteins in *Arabidopsis thaliana*. *Nucl. Acids Res.* 14, 4051-4064.

Li, S.L., and Rédei, G.P. (1969). Thiamine mutants of the crucifer, *Arabidopsis*. *Biochem. Genet.* 3, 163-170.

Lloyd, A.M., Barnason, A.R., Rogers, S.G., Byrne, M.C., Fraley, R.T. and Horsch, R.B. (1986). Transformation of *Arabidopsis thaliana* with *Agrobacterium tumefaciens*. *Science* 234, 464-466.

Ludwig, S.R., Oppenheimer, D.G., Silflow, C.D., and Snustad, D.P. (1987). Characterization of the alpha-tubulin gene family of *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* 84, 5833-5837.

Maher, E.P., and Martindale, S.J. (1980). Mutants of *Arabidopsis thaliana* with altered responses to auxins and gravity. *Biochem. Genet.* 18, 1041-1053.

Marsden, M.P.F., and Meinke, D.W. (1985). Abnormal development of the suspensor in an embryo-lethal mutant of *Arabidopsis thaliana*. *Amer. J. Bot.* 72, 1801-1812.

McKelvie, A.D. (1962). A list of mutant genes in *Arabidopsis thaliana* (L.) Heynh. *Radiation Botany* 1, 233-241.

McKelvie, A.D. (1963). Studies in the induction of mutations in *Arabidopsis thaliana* (L.) Heynh. *Radiation Bot.* 3, 105-123.

Meinke, D.W. (1982). Embryo-lethal mutants of *Arabidopsis thaliana*: Evidence for gametophytic expression of the mutant genes. *Theor. Appl. Genet.* 63, 381-386.

Meinke, D.W. (1985). Embryo-lethal mutants of *Arabidopsis thaliana*: Analysis of mutants with a wide range of lethal phases. *Theor. Appl. Genet.* 69, 543-552.

Meinke, D.W., and Sussex, I.M. (1979a). Embryo-lethal mutants of *Arabidopsis thaliana*: A model system for genetic analysis of plant embryo development. *Dev. Biol.* 72, 50-61.

Meinke, D.W., and Sussex, I.M. (1979b). Isolation and characterization of six embryo-lethal mutants of *Arabidopsis thaliana*. *Dev. Biol.* 72, 62-72.

Meinke, D.W., Franzmann, L., Baus, A., Patton, D., Weldon, R., Heath, J.D., and Monnot, C. (1985). Embryo-lethal mutants of *Arabidopsis thaliana*, in: *Plant Genetics*, UCLA Symposia of Molecular and Cellular Biology, vol. 35 (Freeling, M., ed.), Alan Liss, New York, pp. 129-146.

Mesnard, J.-M., Lebeurier, G., Lacroute, F., and Hirth, L. (1985). Use of labelled cDNA as probe to study differences in messenger RNAs abundance in the leaves and in the pods of *Arabidopsis thaliana*. *Plant Science* 40, 185-191.

Meyerowitz, E.M., and Pruitt, R.E. (1985). *Arabidopsis thaliana* and plant molecular genetics. *Science* 229, 1214-1218.

Mirza, J. I., Olsen, G. M., Iversen, T.-H., and Maher, E. P. (1984). The growth and gravitropic responses of wild-type and auxin resistant mutants of *Arabidopsis thaliana. Physiol. Plant.* 60, 516-522.

Müller, A. (1961). Zur Charakterisierung der Blüten und Infloeszenzen von *Arabidopsis thaliana* (L.) Heynh. *Kulturpflanze* 9, 364-393.

Müller, A. J. (1963). Embryonenstest zum Nachweis rezessiver Letalfaktoren bei *Arabidopsis thaliana. Biol. Zentralbl.* 83, 133-163.

Napp-Zinn, K. (1963). Zur Genetik der Wachsformen. *Beitr. Biol. Pflanz* 38, 161-177.

Negrutiu, I., Jacobs, M., and Cattoir-Reynaerts, A. (1984). Progress in cellular engineering of plants: biochemical and genetic assessment of selectable markers from cultured cells. *Pl. Mol. Biol.* 3, 289-302.

Olsen, M.V., Dutchik, J.E., Graham, M.Y., Brodeur, G.M., Helms, C., Frank, M., MacCollin, M., Scheinman, R., and Frank, T. (1986). Random-clone strategy for genomic restriction mapping in yeast. *Proc. Natl. Acad. Sci. USA* 83, 7826-7830.

Pruitt, R.E., and Meyerowitz, E.M. (1986). Characterization of the genome of *Arabidopsis thaliana. J. Mol. Bio.* 187, 169-183.

Pang, P.P., Pruitt, R.E., and Meyerowitz, E.M. (1988). Molecular cloning and characterization of genes that are expressed in the seeds of *Arabidopsis thaliana*. In Preparation.

Rédei, G.P. (1967). Biochemical aspects of a genetically determined variegation in *Arabidopsis*. *Genetics* 56, 431-443.

Rédei, G.P. (1970). *Arabidopsis thaliana* (L.) Heynh. A review of the genetics and biology. *Bibliogr. Genet.* 20, 1-151.

Rédei, G.P. (1973). Extrachromosomal mutability determined by a nuclear gene locus in *Arabidopsis*. *Mut. Res.* 18, 149-162.

Rédei, G.P. (1975a). *Arabidopsis* as a genetic tool. *Ann. Rev. Genet.* 9, 111-127.

Rédei, G.P. (1975b). *Arabidopsis thaliana*, in: *Handbook of Genetics*, Vol. 2 (King, R.C., ed.), Plenum Press, New York, pp. 151-180.

Reinholz, E. (1947). Auslösung von Röntgenmutationen bei *Arabidopsis thaliana* (L.) Heynh. und ihre Bedeutung für die Pflanzenzüchtung und Evolutionstheorie. *Field Information Agency Technical Report* 1006, 1-70.

Röbbelen, G. (1957a). Über heterophyllie bei *Arabidopsis thaliana* (L.) Heynh. *Ber. Dtsch. Bot. Ges.* 70, 39-44

Röbbelen, G. (1957b). Untersuchungen an strahleninduzierten Blattforbmutanten von *Arabidopsis thaliana* (L.) Heynh. *Z. Indukt. Abst. Vererb.-Lehre* 88, 189-252.

Röbbelen, G. (1965). The Laibach standard collection of natural races. *Arab. Inf. Serv.* 2, 36-47.

Somerville C.R., and Ogren, W.L. (1982). Isolation of photorespiration mutants in *Arabidopsis thaliana*, in: *Methods in Chloroplast Molecular Biology*, (Edelman, M., Hallick, R.B., and Chua, N.H., eds.), Elsevier Biomedical Press, New York, pp. 129-138.

Sparrow, A.H. (1964). Comparisons of the tolerances of higher plant species to acute and chronic exposures of ionizing radiation. *Japan J. Genet.* 40(Suppl.), 12-37.

Steinitz-Sears, L.M. (1963). Chromosome studies in *Arabidopsis thaliana*. *Genetics* 48, 483-490.

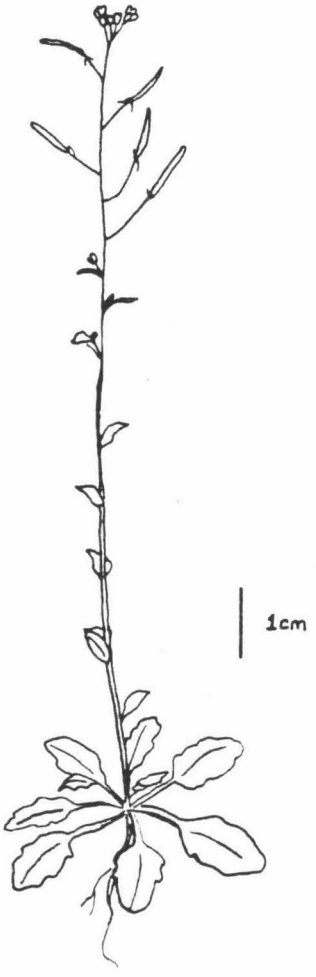
Voytas, D.F., and Ausubel, F.M. (1987). Abstract # 124, *Third International Meeting on Arabidopsis*, Michigan State University.

Whyte, R.O. (1946). *Crop Production and Environment*, Faber and Faber, London.

Zimmerman, J.L., and Goldberg, R.B. (1977). DNA sequence organizations in the genome of *Nicotiana tabacum*. *Chromosoma* 59, 227-252.

Figure 1. The *erecta* mutant of the *Arabidopsis thaliana* Landsberg ecotype. Landsberg *erecta* is commonly used in genetic and molecular genetic experiments because of its rapid life cycle and reduced stature.

Figure 2. Illustration of an *Arabidopsis thaliana* plant (Bensheim ecotype).


1cm

CHAPTER ONE

Molecular Cloning and DNA Sequence of the *Arabidopsis thaliana* Alcohol Dehydrogenase Gene

(Reprinted from *Proc. Natl. Acad. Sci. USA* 83: 1408-1412, 1986)

Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 1408–1412, March 1986
Genetics

Molecular cloning and DNA sequence of the *Arabidopsis thaliana* alcohol dehydrogenase gene

(gene structure/enzyme induction/intron/evolution)

CAREN CHANG AND ELLIOT M. MEYEROWITZ

Division of Biology, California Institute of Technology, Pasadena, CA 91125

Communicated by Ray D. Owen, October 18, 1985

ABSTRACT ***Arabidopsis thaliana* provides an excellent experimental plant system for molecular genetics because of its remarkably small genome size, near absence of dispersed middle repetitive DNA, and short life cycle. We have cloned and determined the nucleotide sequence of a single-copy gene from *A. thaliana* likely to be the gene encoding alcohol dehydrogenase (ADH; alcohol:NAD$^+$ oxidoreductase, EC 1.1.1.1). The gene was isolated from a random recombinant library by cross-hybridization with a maize *Adh1* gene probe. The DNA sequence contains an open reading frame capable of encoding a polypeptide the same length as maize ADH1 and ADH2 (379 amino acids) and having ≈80% homology with both maize enzymes. This open reading frame is interrupted by six introns whose positions are conserved with six of the nine intron positions present in both maize genes. The 5′ and 3′ untranslated regions are, respectively, 58 and 204 base pairs long. Sequences important for eukaryotic gene expression such as the TATA box, polyadenylylation signal, and intron splice-site sequences are found in the expected locations. The gene hybridizes to a specific anaerobically induced RNA in *Arabidopsis* whose appearance correlates with the anaerobic induction of *Arabidopsis* ADH protein.**

Alcohol dehydrogenase (ADH; alcohol:NAD$^+$ oxidoreductase, EC 1.1.1.1) is an easily assayed enzyme whose activity has been observed in numerous higher plants including *Arabidopsis*, maize, pearl millet, sunflower, wheat, and pea (1, 2). Most plants have two or three isozymes of ADH, which exist as both hetero- and homodimers in various organs (1). The enzyme is presumably required by plants for NADH metabolism, via reduction of acetaldehyde to ethanol, during periods of anaerobic stress. High levels of ADH activity are found in dry seeds (3, 4) and in anaerobically treated seeds (5, 6), roots (5), and shoots (7).

The most extensive study of a plant ADH system has been in maize (8) from which both *Adh* genes, *Adh1* and *Adh2*, have been cloned and sequenced (9–11). The coding sequences of these genes are 82% homologous, interrupted by nine identically positioned introns that differ in sequence and length. ADH1 and ADH2 belong to a small group of proteins in maize primary root that are selectively translated in response to anaerobiosis (12); the increased levels of ADH are due to induction of Adh mRNA (9–11, 13).

Arabidopsis ADH is similar to the maize ADHs, although genetic experiments indicate only one *Adh* locus in *Arabidopsis* (14). Examination of *Arabidopsis* ADH in crude extracts has shown that the enzyme behaves as a homodimer of M_r 87,000 (14), close to the maize ADH M_r of ≈80,000 (15). ADH is induced anaerobically in *Arabidopsis* (16) as in maize. ADH is also induced in both maize root and *Arabidopsis* callus by the synthetic auxin 2,4-dichlorophenoxyacetic acid (16, 17), and for *Arabidopsis* this has been shown to be the result of *de novo* synthesis of a poly(A) mRNA (16).

Arabidopsis ADH has potential as a biochemical marker for genetic transformation of *Arabidopsis*: null mutations exist, and ADH is easily induced and assayed. We present here the molecular cloning and characterization of an inducible single-copy gene from *Arabidopsis* likely to be the gene encoding ADH.

MATERIALS AND METHODS

***Arabidopsis* Strains.** The Landsberg *erecta* strain of *A. thaliana* was obtained from F. J. Braaksma (Department of Genetics, Biology Centre, Haren, The Netherlands); the Bensheim strain was obtained from A. R. Kranz (Botanisches Institut, J. W. Goethe-Universität, Frankfurt am Main, Federal Republic of Germany).

General Nucleic Acid Methods. *Arabidopsis* DNA was prepared from whole plants as described in ref. 18. Library construction followed the procedure in ref. 19. The clone nomenclature system is described in ref. 20 with the following additions: f for λ EMBL4 (21); b for pMT21, a 1.9-kilobase-pair (kbp) pBR322 derivative (H. V. Huang, personal communication); j and k, respectively, for pSP65 and pSP64 (Promega Biotec, Madison, WI) containing the SP6 RNA polymerase promoter. Radiolabeled DNA probes were produced by nick-translation (22). Hybridizations with the maize probe were in 50% formamide/5× SSPE (5× SSPE is 5 mM Na_2EDTA/40 mM NaOH/50 mM $NaH_2PO_4 \cdot H_2O$/900 mM NaCl) at 37°C with washes in 0.05× SSPE at 37°C. Genome blot hybridizations were in 50% formamide/5× SSPE at 43°C, with washes in 1× SSPE at room temperature. All DNA manipulations were carried out as described in ref. 23. The DNA sequence was determined by the method of Maxam and Gilbert (24).

Anaerobic Treatment. Seeds were treated for 1–2 days at 4°C in Petri dishes containing distilled H_2O-soaked filter paper (Whatman 3). The seeds were germinated on fresh filter paper on 0.7% agar plates at 25°C with constant illumination (7000 lx). After 3–5 days, seedlings were transferred into 1–2 ml of distilled H_2O (60–100 seedlings per ml). Untreated seedlings were left on plates. At various times, seedlings were placed on filter paper in a Büchner funnel with suction to remove excess water. They were then used for either RNA or protein preparations.

RNA Analysis. Seedlings (60–100) were homogenized for 2 min in a 0.5-ml microtube fitted onto a mini-BeadBeater (Biospec Products). Each tube contained 20 μl of 50% phenol/0.2 M Tris, pH 7.7/10 mM NaCl/75 mg of 0.5-mm zirconium oxide beads (Biospec Products). After homogenization, an additional 20 μl of buffer, 10 μl of 10% $NaDodSO_4$, and 2 μl of yeast tRNA (10 mg/ml) were mixed in by

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: ADH, alcohol dehydrogenase; kbp, kilobase pair(s).

vortexing. The contents were extracted three times with an equal volume of phenol/chloroform/isoamyl alcohol (50:49.5:0.5; vol/vol), twice with chloroform/isoamyl alcohol (99:1; vol/vol), once with ether, and then precipitated with ethanol.

For blot analysis, the RNA was glyoxylated (25), subjected to electrophoresis in agarose gels and transferred to nitrocellulose (23). Radiolabeled RNA probes were synthesized by using SP6 RNA polymerase (26). RNA induction was quantitated by liquid scintillation counting of bands excised from the blots.

Protein Analysis. Seedlings (60–100 per 0.5-ml microtube) were homogenized for 2 min in the mini-BeadBeater. Each tube contained 5.0 μl of 70 mM Tris·HCl, pH 7.2/25% (vol/vol) glycerol/0.8% (vol/vol) 2-mercaptoethanol/0.25% (wt/vol) bromophenol blue/75 mg of 0.5-mm zirconium oxide beads. A hole was punched in the bottom of each tube with a 26-gauge needle and the homogenates were spun into 1.5-ml microtubes. The resulting supernatants were loaded onto nondenaturing polyacrylamide gels consisting of 7.3% (wt/vol) acrylamide/0.2% (wt/vol) *N,N'*-methylenebisacrylamide/70 mM Tris·HCl, pH 7.2/0.8% (wt/vol) ammonium persulfate/0.08% (vol/vol) *N,N,N',N'*-tetramethylethylenediamine. The gels were run at 10 mA at 22°C with recirculated 8.25 mM Tris/30 mM diethylbarbituric acid electrode buffer (pH 7.4). Gels were stained for ADH activity at 37°C in 0.1 M Tris·HCl, pH 7.6/1.5 mM NAD^+/0.25 mM nitroblue tetrazolium/0.26 mM phenazine methosulfate/0.5% (vol/vol) ethanol. Ethanol was omitted in substrate-dependence controls.

5′ and 3′ RNA Mapping. The 5′ end was mapped by primer extension (27) using induced RNA and a synthetic oligonucleotide primer complementary to the RNA sequence at the 3′ end of the believed first exon. The 3′ end was mapped by ribonuclease protection (28) using a gel-isolated *in vitro*-synthesized complementary strand RNA probe, ≈645 bases long, and having one terminus within the last intron.

RESULTS

Isolation of *Arabidopsis* DNA Clones. Four genome equivalents of an *Arabidopsis* DNA library (described in ref. 18) were screened for cross-hybridization with a maize *Adh1* gene fragment probe (kindly provided by M. Freeling). Four positive clones were detected; restriction digests revealed that all four were identical, each containing the same 4.9-kbp *Eco*RI restriction fragment of *Arabidopsis* DNA (the At3001 fragment; Fig. 1*a*) and a 6.6-kbp *Eco*RI stuffer fragment of the λ vector.

To obtain larger clone segments, a new library was constructed consisting of Landsberg *erecta* strain DNA that had been partially digested with *Mbo* I and ligated into *Bam*HI-digested λEMBL4. After amplification of the library, four genome equivalents were screened with nick-translated bAt3001 (a plasmid containing the At3001 fragment). Three positive clones were detected; two contained 15.6-kbp inserts that were identically oriented and indistinguishable by restriction mapping (λfAt3102), and the third contained a 17.2-kbp insert in the opposite orientation (λfAt3101) (Fig. 1*a*).

Genome blots verified that these clones represent the sequence organization present in genomic DNA. Seven different restriction digests of *Arabidopsis* DNA were probed with nick-translated bAt3001 and jAt3011 (a plasmid containing the 2.5-kbp *Sac* I/*Hin*dIII At3011 fragment of λfAt3102, Fig. 1*b*). The hybridized restriction fragments completely agreed with the map produced by the λ clones, suggesting that the region of DNA spanned by the clones exists in a single copy in the *Arabidopsis* genome (data not shown).

The maize *Adh1* gene fragment cross-hybridized with each of the λ clones as expected; the region of hybridization is shown in Fig. 1*a*. The cloned portion of DNA includes at least 7 kbp on either side of this region.

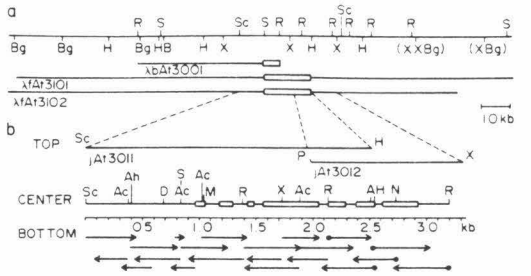


FIG. 1. (*a*) Summary of isolated λ clones. Lines represent the lengths and positions of the three independently cloned inserts relative to the restriction map above. Boxes indicate regions of the clones that hybridize with the maize *Adh1* 2.3-kbp *Hin*dIII gene fragment. (Relative positions have not been verified for sites enclosed in parentheses. Three unmapped *Hin*dIII sites are not shown.) (*b*) DNA sequencing strategy. Top, lines represent the two sequenced subclones, derived from λfAt3102 as indicated by dashed lines. Center, restriction map of sites end-labeled for sequence analysis showing the entire gene structure with boxes representing exons; the gene is shown such that the transcription direction is left to right (determined as in Fig. 2 *a–c*). Bottom, arrows designate the direction and extent of DNA sequence obtained for each fragment. Arrows having a dot denote sequence data from jAt3012; remaining data are from jAt3011. A, *Ava* I; Ac, *Acc* I; Ah, *Aha* III; B, *Bam*HI; Bg, *Bgl* II; D, *Dde* I; H, *Hin*dIII; M, *Msp* I; N, *Nco* I; P, *Pst* I; R, *Eco*RI; S, *Sal* I; Sc, *Sac* I; X, *Xba* I.

RNA Induction Correlates with ADH Protein Induction. The fact that ADH activity increases in anaerobically treated *Arabidopsis* plantlets suggested that anaerobically treated plantlets should contain Adh mRNA to which our *Arabidopsis* clones might hybridize. RNA blots were probed with single-stranded RNA derived from SP6 RNA polymerase transcription of *Hin*dIII-linearized jAt3011 and *Sac*

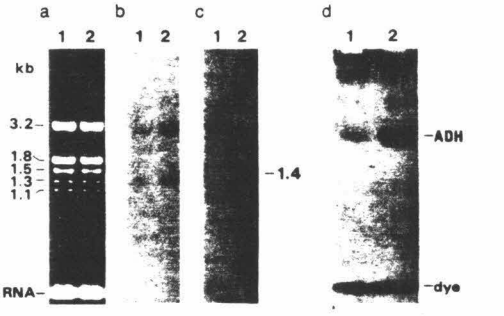


FIG. 2. RNA induction, showing direction of transcription, and ADH protein induction. Lanes 1 and 2, 0 hr and 4 hr of anaerobiosis, respectively. (*a*) Ethidium bromide-stained half of agarose gel containing 10% of total RNA prepared from Bensheim strain plantlets shows similar amounts of RNA for lanes 1 and 2, assessed by the intensity of predominant rRNA bands. (*b*) Autoradiogram of blot of other half of gel, containing 90% of the RNA preparations, probed with a single-stranded RNA probe (described in text) from jAt3011. Little or no specific hybridization is detected. (*c*) Autoradiogram of the same blot hybridized with a single-stranded RNA probe from kAt3011 (after washing off previous probe) shows induction of an RNA in lane 2. An overexposure is shown here to emphasize the low level of hybridization in lane 1. Hybridization conditions were made as close as possible to those of the first hybridization. (*d*) Native protein gel stained for ADH activity shows induction of ADH. Nonspecific protein staining just below the wells indicates that roughly equal amounts of protein were loaded onto both lanes.

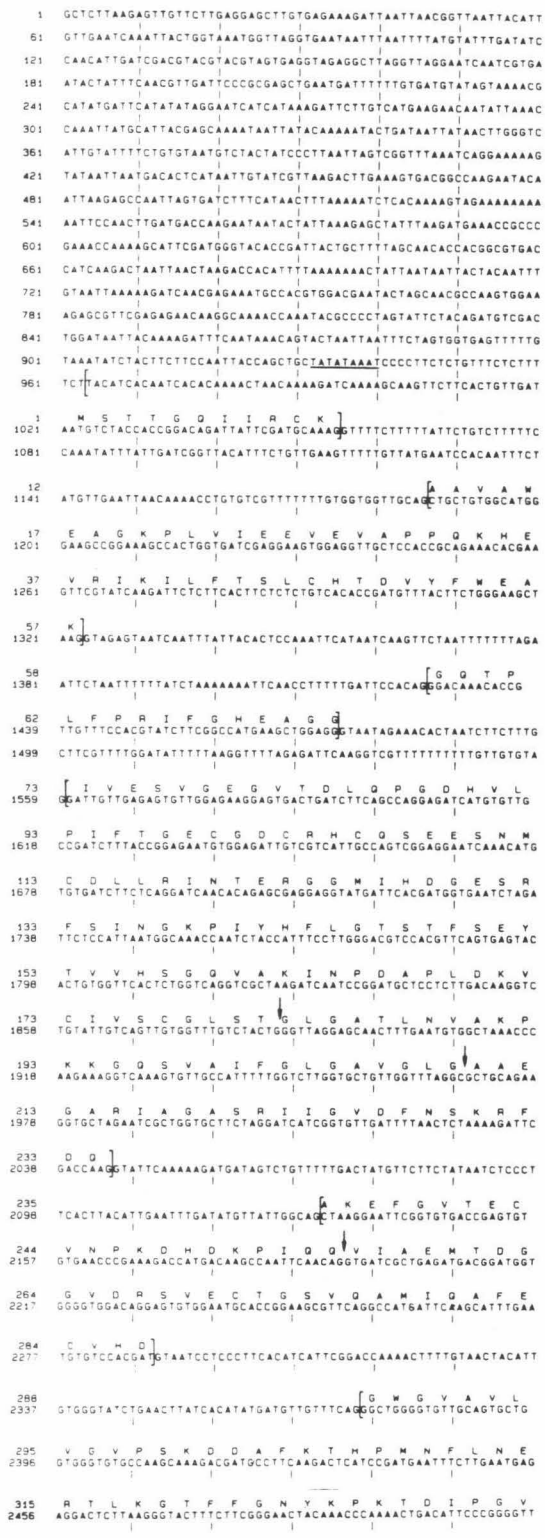


FIG. 3. Complete DNA sequence of the *Adh* gene including introns and flanking regions. The derived amino acid sequence is shown in the one-letter code above the DNA sequence and is numbered separately. Exons are bracketed. Arrows indicate intron positions that are present in the maize *Adh* genes but absent in *Arabidopsis Adh*. The TATA box is noted with a solid underline and the putative polyadenylylation signal is noted with a dotted underline.

I-linearized kAt3011 templates. (kAt3011 is identical to jAt3011 except for the opposite orientation of the At3011 fragment with respect to the SP6 promoter.) These experiments revealed a homologous RNA, ≈1450 nucleotides long, that clearly increased upon anaerobiosis. The induction for Landsberg *erecta* strain plantlets at 2 hr and 4 hr of anaerobiosis was ≈5-fold and 10-fold, respectively. Hybridization to the induced RNA occurred only with the SP6 transcripts from kAt3011 (not from jAt3011), revealing that the *in vivo* transcription direction of the gene, as depicted in Fig. 1, is from left to right (Fig. 2).

Examination of protein from plantlets treated in parallel with those used in the RNA experiments demonstrates that the induction of RNA correlates with induction of ADH activity. Protein extracts were subjected to electrophoresis in nondenaturing polyacrylamide gels and stained for ADH activity. Two bands normally appeared; we identified one of these bands as arising from ADH because of its substrate specificity as well as its shifted migration when *Arabidopsis* ADH electrophoretic variants (14) were used (data not shown). The ADH band showed a marked increase in intensity after anaerobic treatment (Fig. 2*d*).

DNA Sequencing. We determined the DNA sequence of a 3.2-kbp segment (Fig. 1*b*) that contains the region that had hybridized with the induced RNA. The sequence is presented in Fig. 3. Extensive homology was discovered between the *Arabidopsis* gene and the protein coding sequence of both maize *Adh* genes. The homology is interrupted by six probable introns in the *Arabidopsis* gene; although their lengths and sequences differ from the corresponding maize introns, their positions are coincident with six of the nine intron positions present in both of the maize *Adh* genes (Fig. 4). All six begin with the dinucleotide GT and end with the dinucleotide AG, as seen consistently at the intron/exon junctions of eukaryotic genes (29). The average A+T content of these regions is 71%, while the A+T content of the open reading frame is 52%. Translation of the DNA sequence in all possible reading frames produces an open reading frame only when all six of the putative intron sequences are omitted.

The deduced polypeptide encoded by the 1137-nucleotide open reading frame (Fig. 3) shows conservation with both maize ADH sequences: it is of identical length and has 80.5% identity with maize ADH1 and 79% identity with maize ADH2. (The maize ADH proteins are 87% conserved.) The

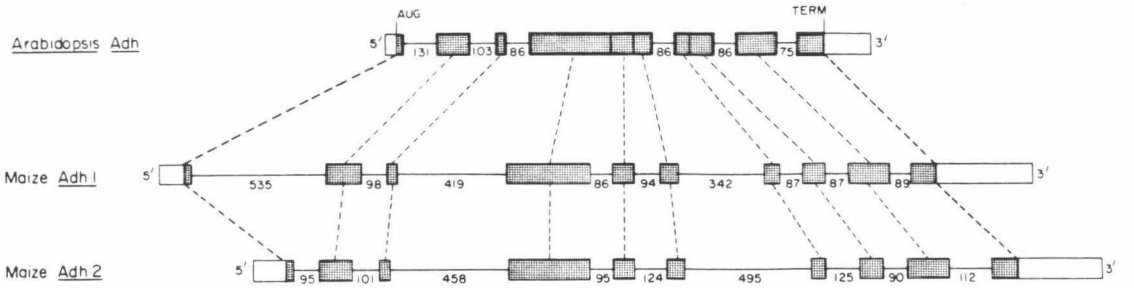


FIG. 4. Comparison of exon/intron structure of *Arabidopsis Adh* and maize *Adh* (11) genes. Boxes represent relative lengths and positions of exons; shaded portions designate protein-coding regions of exons. Solid lines represent relative lengths and positions of introns that are at identical positions in all three genes with respect to the protein sequence; the number below each line gives the length of the intron in nucleotides. Dashed lines connect homologous exons.

nucleic acid sequence coding for this polypeptide has 73% homology with maize *Adh1* and 72% homology with maize *Adh2*. (The maize *Adh* coding regions are 82% conserved.) Of the nucleotide differences with *Adh1*, 53% are silent at the amino acid level and 17% result in conservative amino acid substitutions.

5′ and 3′ RNA Mapping. Primer extension on the RNA resulted in a DNA transcript 58 ± 1 nucleotides beyond the translational initiation codon (data not shown). The position of the initiation codon was determined by homology with the deduced protein sequences of both maize enzymes. Neither the upstream DNA (1021 nucleotides) nor the 5′-untranslated RNA contains an alternative initiation codon that would allow read through. Ribonuclease protection mapping of the 3′ end gave a length of 204 ± 2 bases for the 3′-untranslated RNA using DNA ladder size standards calibrated for RNA (data not shown). The actual length may be as low as 200 nucleotides since the 2–4 terminal adenosine nucleotides of the protected probe are resistant to the ribonucleases used.

DISCUSSION

Sequence analysis indicates that we have cloned the *Arabidopsis Adh* gene. Our clone is single copy in the genome supporting genetic evidence of there being a single *Adh* locus in *Arabidopsis* (14). The calculated molecular weight of the deduced polypeptide is 41,200, which is close to the value 44,000 measured by the migration of *Arabidopsis* ADH monomers in polyacrylamide gels (14). The length of the anaerobically induced RNA is consistent with the predicted exon sizes of the gene, allowing for a poly(A) tail of ≈50 nucleotides.

The gene contains sequences characteristic of expressed eukaryotic genes. A TATA box is located 23 base pairs (bp) upstream of the mapped 5′ end of the RNA, consistent with other genes expressed in plants (30). A possible polyadenylylation signal, AATATAAA, similar to the plant consensus signal (30, 31), is situated 19 bases upstream of the mapped 3′ end. An AGGA box, possibly involved in transcription regulation, is occasionally found 36–59 bp upstream of the TATA box of plant genes (30). Such a sequence for the *Arabidopsis* gene might be TAAACAGTACT, which is similar to the plant consensus sequence $\frac{C}{T}A_{2-5}\frac{G}{T}NGA_{2-4}\frac{C}{T}\frac{C}{T}$ and is located 57 bp upstream of the putative TATA box.

Upstream sequences may be required for ADH expression during anaerobiosis or 2,4-dichlorophenoxyacetic acid induction. Although the DNA sequences upstream of the coding sequences have generally diverged, distinct homologous stretches exist 5′ of the translation start sites of the *Arabidopsis Adh* and maize *Adh1* genes (Fig. 5). These sequences are not found in maize *Adh2*, with the exception of the TATA box. The three 8-bp sequences shared by the two maize genes in their 5′-untranslated regions (11) are not seen in the 1021 nucleotides upstream of the initiation codon of the *Arabidopsis* gene.

The six introns of the *Arabidopsis* gene possess the 5′ and 3′ splice site consensus sequences of eukaryotic genes (29) (Fig. 6). Recently, internal signals, such as the conserved TACTAAC sequence at 20–55 nucleotides from the 3′ border of yeast introns (32), have been discovered to be required for gene splicing. Each *Arabidopsis Adh* intron contains the consensus sequence $\frac{C}{T}T\frac{A}{G}AT$ at 16–39 nucleotides from the 3′ border, which is homologous with the animal consensus sequences found in identical positions in diverse animal gene introns (33) and homologous with the 5-base-long sequences used *in vitro* for formation of the lariat intermediate in intron excision (34). In addition to this consensus sequence, a second highly conserved sequence was found within the *Arabidopsis* introns (Fig. 6); this second consensus sequence may have some role in *Arabidopsis* gene expression.

The deduced *Arabidopsis* ADH polypeptide is 47% conserved with horse liver ADH. Structurally and functionally important residues defined by the tertiary structure of the horse enzyme (35), such as the seven residues that provide ligands for the catalytic and noncatalytic zinc atoms, are conserved, suggesting that the *Arabidopsis* enzyme has a similar structure.

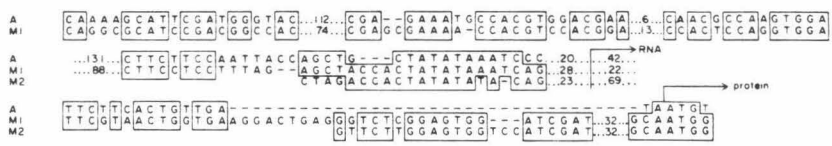


FIG. 5. 5′ sequence homologies for *Arabidopsis Adh* (A), maize *Adh1* (M1), maize *Adh2* (M2). The number of bases between each homologous stretch is given. Contiguous bases, spread apart for alignment purposes, are linked with dashes. The homologies between M1 and M2 are from ref. 11.

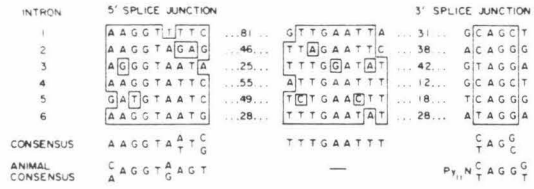


FIG. 6. Intron consensus sequences for the *Arabidopsis Adh* gene. The number of bases between each consensus block is given.

Comparison of the nucleic acid coding sequences and deduced protein sequences of the *Arabidopsis Adh* gene and of the two maize *Adh* genes shows that the two maize genes are more related to each other than either is to the *Arabidopsis* gene. It is thus likely that the maize and *Arabidopsis Adh* genes descended from a single ancestral gene that was subsequently duplicated in the maize but not in the *Arabidopsis* lineage. Further comparison with other angiosperm *Adh* gene sequences, when they become available, should indicate whether the ADH isozymes of other monocots and dicots are likewise descendants of a single gene in their last common ancestor or are the result of more ancient gene duplications, as is generally supposed (1, 11).

The low copy-number sequence content of the maize genome is >20-fold that of the *Arabidopsis* genome (36, 37). This difference is partially reflected in both the size and the copy number of the *Adh* genes of the two species. The *Arabidopsis* gene has three fewer introns than either of the two maize genes; the sum of the intron lengths in the *Arabidopsis* gene is >1000 bases shorter than the comparable sum in the maize *Adh* genes. Other *Arabidopsis* genes have also been seen to be present in fewer copies per haploid genome than the homologous genes in other angiosperms (38).

This work demonstrates the feasibility of isolating a dicot plant gene by homology with a monocot gene. Cross-hybridization is proving to be a useful approach for isolating plant genes. A plant such as *Arabidopsis*, which lends itself well to molecular genetics and cloning, might be exploited for rapid cloning of genes that cross-hybridize with specific genes in other plants (38). Our cloned *Arabidopsis* gene can be tested for the ability to restore ADH function to existing *Arabidopsis* ADH null mutants (14) via Ti plasmid-mediated transformation. Should this approach succeed, we hope to answer questions concerning the regulation of ADH in plants and to develop *Arabidopsis* ADH as a marker in gene fusion experiments.

We are grateful to C. Rice for technical advice and to T. Hunkapiller and C. Martin for DNA sequence analysis programs. We thank M. Garfinkel, P. Mathers, K. Fryxell, and R. Pruitt for comments on the manuscript. This work was supported by Grant PCM-8408504 from the National Science Foundation (E.M.M.). C.C. was supported by National Research Service Award 5 T32GM07616 from the National Institutes of Health.

1. Gottlieb, L. D. (1982) *Science* **216,** 373–380.
2. Leblová, S., Zimáková, I., Barthová, J. & Ehlichová, D. (1971) *Biol. Plant.* **13,** 33–42.
3. Efron, Y. & Schwartz, D. (1973) *Proc. Natl. Acad. Sci. USA* **61,** 586–591.
4. Leblová, S., Zimáková, I., Sofrová, D. & Barthová, J. (1969) *Biol. Plant.* **11,** 417–423.
5. Freeling, M. (1973) *Mol. Gen. Genet.* **127,** 215–227.
6. Banuett-Bourrillon, F. & Hague, D. R. (1979) *Biochem. Genet.* **17,** 537–552.
7. App, A. A. & Meiss, A. N. (1958) *Arch. Biochem. Biophys.* **77,** 181–190.
8. Freeling, M. & Birchler, J. A. (1981) in *Genetic Engineering, Principles and Methods*, eds. Stelow, J. K. & Hollaender, A. (Plenum, New York), Vol. 3, pp. 223–264.
9. Gerlach, W. L., Pryor, A. J., Dennis, E. S., Ferl, R. J., Sachs, M. M. & Peacock, W. J. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 2981–2985.
10. Dennis, E. S., Gerlach, W. L., Pryor, A. J., Bennetzen, J. L., Inglis, A., Llewellyn, D., Sachs, M. M., Ferl, R. J. & Peacock, W. J. (1984) *Nucleic Acids Res.* **12,** 3983–4000.
11. Dennis, E. S., Sachs, M. M., Gerlach, W. L., Finnegan, E. J. & Peacock, W. J. (1985) *Nucleic Acids Res.* **13,** 727–743.
12. Sachs, M. M., Freeling, M. & Okimoto, R. (1980) *Cell* **20,** 761–767.
13. Ferl, R. J., Brennan, M. D. & Schwartz, D. (1980) *Biochem. Genet.* **18,** 681–691.
14. Dolferus, R. & Jacobs, M. (1984) *Biochem. Genet.* **22,** 817–838.
15. Freeling, M. & Schwartz, D. (1973) *Biochem. Genet.* **8,** 27–36.
16. Dolferus, R., Marbaix, G. & Jacobs, M. (1985) *Mol. Gen. Genet.* **199,** 256–264.
17. Freeling, M. (1973) *Mol. Gen. Genet.* **127,** 215–227.
18. Pruitt, R. E. & Meyerowitz, E. M. (1986) *J. Mol. Biol.*, in press.
19. Meyerowitz, E. M., Guild, G. M., Prestidge, L. S. & Hogness, D. S. (1980) *Gene* **11,** 271–282.
20. Meyerowitz, E. M. & Martin, C. H. (1984) *J. Mol. Evol.* **20,** 251–264.
21. Frischauf, A.-M., Lehrach, R., Poustka, A.-M. & Murray, N. (1983) *J. Mol. Biol.* **170,** 827–842.
22. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113,** 237–251.
23. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
24. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
25. McMaster, G. K. & Carmichael, G. G. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 4835–4838.
26. Melton, D. A., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K. & Green, M. R. (1984) *Nucleic Acids Res.* **12,** 7035–7070.
27. Ghosh, P. K., Reddy, V. B., Piatak, M., Lebowitz, P. & Weissman, S. M. (1980) *Methods Enzymol.* **65,** 580–595.
28. Zinn, K., DiMaio, D. & Maniatis, T. (1983) *Cell* **34,** 865–879.
29. Mount, S. M. (1982) *Nucleic Acids Res.* **10,** 459–472.
30. Messing, J., Geraghty, D., Heidecker, G., Hu, N.-T., Kridl, J. & Rubenstein, I. (1983) in *Genetic Engineering of Plants*, eds. Kosuge, T., Meredith, C. P. & Hollaender, A. (Plenum, New York), pp. 211–227.
31. Dhaese, P., De Greve, H., Gielen, J., Seurinck, J., Van Montagu, M. & Schell, J. (1983) *EMBO J.* **2,** 419–426.
32. Langford, C. J. & Gallwitz, D. (1983) *Cell* **33,** 519–527.
33. Keller, E. B. & Noon, W. A. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 7417–7420.
34. Ruskin, B., Krainer, A. R., Maniatis, T. & Green, M. R. (1984) *Cell* **38,** 317–331.
35. Eklund, H., Nordström, B., Zeppezauer, E., Söderlund, G., Ohlsson, I., Boiwe, T., Söderberg, B.-O., Tapia, O. & Brändén, C.-I. (1976) *J. Mol. Biol.* **102,** 27–59.
36. Hake, S. & Walbot, V. (1980) *Chromosoma* **79,** 251–270.
37. Leutwiler, L. S., Hough-Evans, B. R. & Meyerowitz, E. M. (1984) *Mol. Gen. Genet.* **194,** 15–23.
38. Meyerowitz, E. M. & Pruitt, R. E. (1985) *Science* **229,** 1214–1218.

CHAPTER TWO

Transformation of *Arabidopsis thaliana*: Complementation of an Alcohol Dehydrogenase Mutant with the Alcohol Dehydrogenase Gene

INTRODUCTION

This chapter describes the genetic complementation of an *Arabidopsis thaliana* alcohol dehydrogenase (ADH) null mutant with the *Arabidopsis* ADH gene introduced by *Agrobacterium*-mediated transformation. Transformation of plants is becoming a valuable tool for detailed analyses of gene structure and function in plants, as it has been in other experimental organisms such as yeast and *Drosophila*. Of importance for many of these experiments is the ability to complement a mutant phenotype with the wild-type gene, an ability not previously demonstrated in plants. Further, complementation by transformation allows genes to be identified during chromosome walking as described in Chapter Three. The availability of these approaches should enhance the usefulness of *Arabidopsis* for molecular genetic experiments.

The ADH gene was chosen for these experiments for several reasons. First, ADH can serve as a gene tag in transformation experiments because the enzyme activity is readily assayed histochemically. This has been demonstrated recently in transformations using upstream DNA sequences of two *Arabidopsis* seed-specific genes fused to the *Arabidopsis* ADH coding sequence (M. Yanofsky and E. Meyerowitz, unpublished; P. Pang and E. Meyerowitz, unpublished). In one fusion that involved a seed gene of unknown function, the activity was found to be expressed specifically in the developing seed coat. The second reason for using this gene is that tight selection against ADH activity in seeds is provided by exposure to allyl alcohol. The enzyme converts allyl alcohol to acrylaldehyde, a potent protein-alkylating agent which is toxic. This counter-selection has potential for allowing isolation of mutations in trans-acting regulators of gene expression. Third, ADH provides an agreeable system for complementation studies; there is only one ADH gene in *Arabidopsis*, and null mutations in *Arabidopsis* have been isolated using allyl alcohol selection (Negrutiu *et al.*, 1984). The enzyme is nonessential since null mutations are not lethal. The null mutant used in this study, R002, lacks detectable ADH activity in tissues

where it is expressed in the wild-type and has reduced levels of ADH mRNA (P. Pang, personal communication).

The best currently available method for gene transfer in dicotyledonous plants uses the natural DNA transfer system of the soil bacterium *Agrobacterium tumefaciens* (Stachel and Zambryski, 1986). *Agrobacterium* causes crown gall disease in dicotyledonous plants as a result of integrating a portion of its large tumor-inducing (Ti) plasmid into the plant genome. The transferred portion is called the T-DNA, which encodes oncogenic functions involved in phytohormone biosynthesis. The genomic sites of integration are apparently random (Zambryski *et al.*, 1982; Chyi *et al.*, 1986); and, once integrated, the T-DNA is inherited in a Mendelian fashion (Horsch *et al.*, 1984; Budar *et al.*, 1986). Two regions of the Ti plasmid are required for transformation: the virulence (VIR) region which functions in *trans* in the T-DNA transfer process, and the cis-acting right border region of the T-DNA which directionally and physically defines the boundary for the start of integration. The left border of the T-DNA usually defines the other boundary, but its presence in the T-DNA is not essential for transfer (Joos *et al.*, 1983; Jen and Chilton, 1986). Many transformation experiments use nononcogenic *Agrobacterium* strains in which the Ti plasmid carries a functional VIR locus but has been disarmed by replacement of the T-DNA with foreign DNA sequences (Zambryski *et al.*, 1983). The DNA to be transferred can be cloned into a smaller vector, called a binary vector, that carries either both T-DNA border sequences or only the right border sequence (Hoekema *et al.*, 1983). The binary vectors are able to replicate in *E. coli*, as well as *Agrobacterium*, facilitating DNA manipulation. Although *Agrobacterium*-mediated transformation is now a routine procedure for many plants, the mechanism by which DNA is transferred and integrated into the plant genome is not well understood.

The work presented here describes some of the DNA structures that can result from this method and helps to characterize *Agrobacterium*-mediated transformation in *Arabidopsis*.

MATERIALS AND METHODS

Strains and plasmids

Agrobacterium tumefaciens strain ASE (Steve Rogers, unpublished work; an analogous system is described in Fraley *et al.*, 1985) and the binary transformation vector pMON410 (Rogers *et al.*, 1987) were obtained from Monsanto Company. ASE is nopaline strain C58C1 (chloramphenicol resistant) harboring the disarmed Ti plasmid pTiT37-SE. The ADH-vector construct was transferred from *E. coli* to ASE by conjugation as described by Rogers *et al.* (1986) in the presence of *E. coli* HB101 containing the mobilization plasmid pRK2013 (Ditta *et al.*, 1980).

Seed of the *Arabidopsis* ADH null mutant, R002 (Negrutiu *et al.*, 1984), was kindly provided by Michel Jacobs, of the Free University of Brussels, Belgium. The mutation was induced in the Bensheim (Be-0) ecotype (Robbelen, 1965) which we used in control experiments. In some controls we also used the Landsberg ecotype (Robbelen, 1965) carrying the *erecta* mutation (Rédei and Hirono, 1964) (L*er*) since the ADH gene was obtained from this line (Chang and Meyerowitz, 1986).

Tobacco feeder cells (*Nicotiana glutinosa*) were kindly provided by the laboratory of T. Murashige of the University of California, Riverside.

Transformant nomenclature

Calli were numbered according to their order of appearance, and each shoot was designated by a letter after the callus number. For example, 2a and 2d are two different plants (T_1 generation) from callus line 2. For each subsequent generation, the plants were given a number preceded by a period to indicate each generation, e.g. 2a.12.1 is the first T_3 plant descended from the twelfth T_2 plant (by self-fertilization) descended from the T_1 plant 2a, again by self-fertilization. Individuals selected with hygromycin were numbered separately from those selected with allyl alcohol such that the same number may be used for

two different plants; the distinction will be made in the text. The control transformants were designated A1-3.

Transformation and regeneration

The leaf inoculation procedure of Lloyd *et al.* (1986) was followed with a few minor exceptions. Leaves were removed from 4 to 5 week old soil-grown plants (before the primary inflorescence appeared) rather than from axenically-grown plants; prior to cutting the leaves in half, the leaves were surface sterilized for 10 minutes with 0.525% wt/vol sodium hypochlorite (10% vol/vol commercial bleach) plus a few drops of Tween 20, followed by several rinses with sterile distilled H_2O. Sterilized leaves were placed on pre-treatment medium for 2 or 7 days. For the initial selections, 25 ug/ml hygromycin B was used; after six weeks, the hygromycin B was lowered to 20 ug/ml. The rooting medium, to which the selected shoots were transferred, contained no hygromycin. Arginine (for stimulating opine production) was omitted from the medium since opine assays were not performed.

Selection of plants on hygromycin

Dry seeds were surface sterilized with 0.525% sodium hypochlorite for 10 minutes. After rinsing thoroughly with distilled H_2O, the seeds were placed on selection media that consisted of MS plant salt mixture (Murashige and Skoog, 1962), 2% sucrose, 0.8% agar, 60 mg/ml hygromycin B (Calbiochem, Behring Diagnostics) and incubated at 25°C with constant illumination (7000 lx). Seedlings were scored for resistance after about 10 days.

Staining of roots for ADH activity

Seeds that had been germinated (at 25°C and constant 7000 lx) on moist filter paper to the point of radicle emergence were submerged in 50% dimethyl sulfoxide for 30-40 minutes at room temperature. After thoroughly rinsing with distilled H_2O, the germinated

seeds were placed in fresh stain solution (0.1M Tris-HCl, pH 7.5, 0.2 mg/ml nitroblue tetrazolium, 1 mg/ml NAD^+, 8% EtOH) at room temperature for 10-20 minutes.

Selection of ADH$^-$ seeds with allyl alcohol

Dry seeds were surface sterilized as above, then submerged in 35 mM allyl alcohol for 2 hours at room temperature. After rinsing thoroughly with distilled H_2O, the seeds were placed on moist filter paper and incubated as above to test for germination. Seeds were treated in parallel with H_2O (no allyl alcohol) to estimate the germination frequency in the absence of selection. Be-0 and L*er* were shown in control experiments to be completely sensitive to 35 mM allyl alcohol, while R002 was resistant to greater than 100 mM.

Staining of chromocenters

Seedlings that had been germinated as above on moist filter paper for several days were fixed in 3:1 EtOH:acetic acid for 2-4 hours at room temperature, or overnight at 4°C. With equivalent results, fixed seedlings were either: (i) stained in a filtered solution of 2% orcein (in 1:1 lactic acid:acetic acid) for 1-2 hours at 65°C, after which root tips were excised and squashed in the stain solution under a cover slip; or (ii) incubated in 45% acetic acid for 10-20 minutes at 70°C, after which root tips were excised and placed in stain solution for 10 minutes at room temperature, then squashed as above. The stained chromocenters (Ambros and Schweizer, 1976) were viewed at 1000x magnification.

DNA gel blots and hybridizations

Plant DNA preparations, restriction digests, and genomic DNA blots were performed as described in Chapter Three of this thesis. To isolate specific DNA sequences for probes, restriction fragments of plasmids were separated by agarose gel electrophoresis then purified by electro-elution (Gobel *et al.*, 1987). The DNA was α^{32}P-labelled by nick-translation (Rigby *et al.*, 1977). Most of the sequences for probes were isolated from plasmid whAt3013 (described in the text); the "wh" vector was used for probe 1. Probe 6

was obtained from plasmid pMON596 (kindly provided by Monsanto Company), and probes 10 and 11 (see text) were obtained from a subclone of probe 9 DNA into pUC19.

RESULTS

The ADH gene-transformation vector

The transformation vector, pMON410, is a wide host-range binary vector which contains the hygromycin phosphotransferase (HPH) coding sequence (Waldron *et al.*, 1985) fused to the constitutive 35S RNA promoter of cauliflower mosaic virus (Odell *et al.*, 1985). Plants transformed by this vector are selectable on growth media containing the aminocyclitol antibiotic hygromycin B. In order to facilitate cloning of the *Arabidopsis* ADH gene into pMON410, the vector's unique *Hpa*I site was replaced with a unique *Bgl*II site by insertion of a *Bgl*II linker. The modified vector pMON410.1 was abbreviated "wh", following the nomenclature system of the laboratory. The *Bgl*II 3013 fragment of λfAt3102 (Chang and Meyerowitz, 1986), containing the *Arabidopsis* ADH gene from the ecotype Landsberg *erecta*, was ligated into the *Bgl*II site to create whAt3013 (Figure 1). The ADH gene in 3013 is flanked by 4.3 kb of upstream DNA and 3.4 kb of downstream DNA. The orientation of the T-DNA right border sequence in the vector is such that transfer of the vector into the plant genome procedes as is shown counter-clockwise in the figure.

Transformation and regeneration of *Arabidopsis*

Using Ti plasmid-mediated transformation, nine independent transformed lines were obtained on the basis of hygromycin resistance. We began by inoculating several hundred leaf explants with a culture of *Agrobacterium* strain ASE that contained the vector whAt3013. By ten weeks after infection, approximately 40% of the leaf explants under hygromycin selection had produced callus, while less than 1% of the explants that had been infected with ASE lacking the whAt3013 vector had produced callus. (Of infected explants grown in the absence of hygromycin, 100% produced callus.) Within the same time span, approximately 40% of the calli gave rise to shoots. Of the excised shoots placed on rooting medium for 1-2 weeks, 70% produced roots. Approximately 60% of the rooted shoots

survived transplantation to soil. At this point, there were 17 regenerated plants (called the T_1 generation) which were derived from 12 independent transformed calli. Fertile flowers were produced by 15 plants of ten calli. Experiments with one of the ten lines were aborted due to low viability. In summary, the frequency of leaf explants that gave rise to fertile plants was approximately 6%, and mature seeds were obtained as early as 12 weeks after infection. (Three fertile plants were regenerated from the control transformation using ASE minus the vector without hygromycin selection.)

Genetic analysis of transformants

Inheritance patterns of ADH activity and hygromycin resistance were analyzed in each of the nine transformed lines. From this analysis we determined that: (i) tissue- and stage-specific ADH activity was restored in the null mutant, although in some cases the level of ADH detected was lower than wild-type; (ii) ADH activity did not necessarily co-segregate with hygromycin resistance; (iii) a majority of the transformants were tetraploid (R002, prior to transformation, was diploid); and (iv) transformation events were inherited in a stable, Mendelian manner.

ADH Staining. ADH activity in the nine transformants was qualitatively measured using two methods: histochemical staining of young roots, and allyl alcohol selection against ADH in mature seeds. The activity was scored in self-fertilized progeny of the regenerated plants. An example of ADH staining is shown in Figure 2a. Eight of the lines exhibited staining identical in both intensity and specificity to that of the wild-type. No positive staining was seen in line 23. In lines 3, 10, and 27, a proportion of progeny stained weakly compared to the wild-type; the stain intensity was subdued, and occasionally it was stronger at the very tip of the root (Figure 2b). A genetic basis for the weak staining in line 27 became clear as described below. However for lines 3 and 10, weak staining was variable and not observed in some experiments at all; this may have been due to several factors which varied between experiments: the length of exposure to the

staining solution, the age of the roots, the growth conditions, and the batch of staining solution. However, no weak staining was ever seen with wild-type control seeds over many different trials. A molecular basis for the ADH staining results will be discussed in the next section of the Results.

Allyl Alcohol Sensitivity. The segregation results for ADH activity (and hygromycin resistance) are shown in Table 1. The segregation of ADH staining was compared with that of allyl alcohol sensitivity. Susceptibility to allyl alcohol was exhibited by all of the transformant lines except line 23 which lacked ADH staining. Co-segregation of allyl alcohol sensitivity and ADH staining was evident in six lines (including line 23), but not in lines 3, 27, and 11 (Table 1).

In both lines 3 and 27, segregation of allyl alcohol sensitivity was consistent with a 3:1 ratio, while positive ADH staining segregated at a much higher proportion. A basis for this was elucidated in the T_3 generation of line 27. First, in the self-cross progeny of two T_2 plants raised from allyl alcohol resistant seeds, weak but positive ADH staining was observed (Table 2). This result was corroborated by genome blots (described in the next section of the Results) that revealed the presence of whAt3013 DNA in both plants. Apparently these two plants escaped allyl alcohol selection due to having a low level of ADH activity as seeds. Second, the weak and normal phenotypes of ADH staining were found to segregate from each other in the T_3 generation of line 27 (Table 2). This was consistent with the fact that variable ADH staining had been observed in the T_2 generation.

Based on the segregation of these two types of staining, we could hypothesize two independent loci, one responsible for the normal staining and one for the weak staining (Table 2). Thus, an individual heterozygous at both loci would produce progeny showing 12:3:1 segregation for normal:weak:negative staining which closely estimates the segregation for 27a.2 (Table 2). With this two locus hypothesis, the T_2 segregation of

ADH staining for both 27a and 3a should be 15:1 which fits very well with the observed ratios (Table 1).

We were interested in testing the weak-staining individuals for resistance to higher doses of allyl alcohol and to see how staining in roots was correlated with allyl alcohol sensitivity. Seeds of plant 27a.3, homozygous for weak-staining (Table 2), were exposed to 0, 35, and 100 mM allyl alcohol for two hours using approximately 50 seeds for each treatment. The seeds were found to be more susceptible to the higher dose, and did not recover as did R002 (Figure 3). The wild-type was completely killed at these concentrations (data not shown). These experiments show that different levels of allyl alcohol can be used for selecting against different levels of ADH expression. In addition, there was a correspondence between low activity in roots, evidenced by the weak staining, and low activity in developing seeds. In another experiment, 200 mM allyl alcohol was applied for 1 hour to seeds of 27a.3, and the result was 57% sensitivity after eight days. A similar experiment was performed on progeny of the two plants that had escaped selection at 35 mM allyl alcohol; the seeds were killed by 100 mM at a frequency of 56% after twelve days.

Although some weak ADH staining had been seen in line 10 (Table 1), there was no obvious effect on allyl alcohol sensitivity in this line, probably because the proportion of weak staining seeds was relatively low.

The aberrant frequency of allyl alcohol sensitivity in line 11 (Table 1) is possibly due to a statistical fluctuation in the correction for germination frequency since only 12 out of 29 seeds germinated in the no allyl alcohol control. If this frequency underrepresented the germination of the treated seeds by 8% (2 or 3 seeds), then the allyl alcohol sensitivity data would have had >5% probability of fitting a 3:1 ratio.

Hygromycin Resistance. In addition to ADH activity, the T_2 progeny were tested for hygromycin resistance (Table 1, Figure 4). Five lines grew well on hygromycin-

containing media. However, resistance was virtually nonexistent in line 27 even though the callus and shoots had been selected with hygromycin. In line 27, the DNA containing the chimeric HPH gene and nearby sequences were found to be methylated (described in the next section of the Results). Three lines (21, 10 and 3) appeared to have feeble resistance since the plants grew poorly and had little or no root development. The growth of these plants resembled *Arabidopsis* plants transformed with the HPH coding sequence fused to the cauliflower mosaic virus promoter for the 19S RNA, rather than the 35S RNA (Patty Pang, personal communication). The 19S promoter is weaker than the 35S promoter (Rothstein *et al.*, 1987). The difficulty in scoring hygromycin resistance is one explanation for the aberrant low ratios in lines 10 and 21.

Hygromycin resistance and ADH activity cosegregated in the first five lines listed in Table 1, assuming a single locus of HPH activity in line 10. The remaining four lines (3, 27, 21 and 23) showed different segregation patterns. In line 3, root tip ADH segregation was consistent with a 15:1 ratio, while hygromycin resistance segregated approximately 3:1. In line 27, root tip ADH segregation was consistent with a 15:1, but none of the progeny seeds tested were clearly hygromycin resistant (6% were weakly resistant). In line 21, ADH activity was present in all progeny seeds tested, but hygromycin resistance segregated approximately 1:1. In line 23, all progeny seeds tested lacked ADH activity, but hygromycin resistance segregated approximately 3:1. A molecular basis for these results will be described in the next section of the Results.

As a further test of cosegregation, progeny of self-fertilized plants grown from allyl alcohol resistant seeds were scored for both ADH staining and hygromycin resistance. This was done with lines 2, 3, 7, 11 and 23. For lines 2, 3, 7 and 11, progeny of allyl alcohol resistant plants completely lacked both ADH activity and hygromycin resistance as predicted (data not shown). For line 23, progeny of two plants segregated for hygromycin resistance. This result, consistent with genome blot hybridization (described in the next

section of the Results), was due to the absence of detectable ADH activity in this line such that allyl alcohol had failed to select against transformed seeds.

As shown in Table 1, lines 2, 21 and 23 were scored for one or two of the phenotypes using a second plant that arose from the same callus line. Callus lines 2, 10, and 21 had each produced two fertile plants and callus line 23 produced three. (The second plant from line 10 produced too few seeds to test segregation.) The phenotypic ratios from the second plants were essentially identical to those of the first plants, indicating that each set of plants arose from a single transformation event.

Ploidy. Segregation of ADH activity and hygromycin resistance was similarly examined in subsequent generations. In several lines, the T_3 and T_4 ratios were inconsistent with the ratios expected from diploid segregation. This is illustrated by the data in Table 3a, b. These ratios were indicative of autotetraploid segregation, which would not have been apparent from the T_2 segregation data for regenerated transformants of the simplex (Aaaa) genotype. For 10a, the observed distribution of T_2 genotypes would have been consistent with diploid genetics if not for the segregation in the progeny of 10a.5 (Table 3a): here, the deviation from a 3:1 ratio expected if the parent was heterozygous diploid is highly significant ($P < 0.1\%$). We also ruled out the presence of two independent loci in a diploid, since only a single locus segregated in the T_2 generation (Table 1). Thus, 10a.5 was most probably duplex tetraploid for the ADH/HPH locus; progeny of a duplex tetraploid would be expected to segregate 35:1 which fits the observed ratio with a high probability (Table 3a). Tetraploidy was also indicated in line 10a.8: the appearance of negative ADH segregants (Table 3b) showed that 10a.8 could not be homozygous as one might have guessed from the T_3 ratio of 1:0. Based on the $\chi 2$ test shown, 10a.8 was unlikely to be heterozygous diploid. The results were consistent with 10a.8 being duplex tetraploid. Similarly, 2a.4.2 and 2a.4.3 were unlikely to be produced from a heterozygous diploid, thus 2a.4 is probably simplex tetraploid. The genotype of

2a.4.11 could have arisen from a simplex parent by double reduction in meiosis (which can occur for loci not linked to the centromere).

To determine the ploidy of each transformed line, chromocenter counts were made from root tip cells of representative T_2 and T_3 plants. Interphase chromocenters can be used as an indication of the mitotic chromosome complement in *Arabidopsis* because each chromosome has heterochromatin in its centromeric region which remains condensed through interphase and is visible as a darkly stained chromocenter (Ambros and Schweizer, 1976). The ten chromocenters in diploid interphase cells and the twenty in tetraploid cells are not always visible due to technical difficulties. The results, shown in Figure 5a, confirmed that five of the transformant lines (2, 4, 10, 21, 23) were tetraploid, while the remaining four lines (3, 7, 11, 27) were diploid, as was R002 (prior to transformation). Tetraploids are often larger and more robust than diploids; plants of line 4 were unusually large. Measurements of seed length showed that the average length of diploid seeds is less than the average length of tetraploid seeds (Figure 5b).

Stable Inheritance. By analyses similar to that in Table 3, the T_3 generation continued to show Mendelian inheritance of the ADH and hygromycin resistant phenotypes, with the exception of those cases for which hygromycin resistance was marginal and difficult to score. Self-cross progeny of T_2 plants were scored using two to sixteen T_2 plants per line (data not shown). In addition, ADH segregation was examined in further generations of diploid line 7a.1 and tetraploid line 2a.12. Nine hygromycin-selected seeds of heterozygous diploid plant 7a.1 were grown and their self-cross progeny (the T_4 generation) were stained for ADH activity. In accordance with the expected proportions, progeny of one third of the plants were all ADH positive while progeny of two thirds segregated 3:1 for ADH activity (data not shown). The same experiment was performed on the T_5 generation derived from duplex tetraploid plants 2a.12.2 and 2a.12.6, and triplex (or

quadruplex) plant 2a.12.7. Again, the expected proportions were obtained (data not shown).

Plants 7a.3 and 7a.4 appeared to be homozygous for the ADH/HPH locus. To test these lines for stable homozygous ADH activity, twelve random self-cross progeny of 7a.3 and ten of 7a.4 were grown and allowed to self-fertilize. Between 31 and 110 germinated seeds per plant were stained for ADH activity; in all, 1203 seeds were stained. All were ADH positive, indicating that 7a.3 and 7a.4 were homozygous for the transformed locus and that transmission of ADH activity in these lines was stable.

Molecular analysis of transformants

Genomic DNA blot hybridizations confirmed the presence of whAt3013 DNA in all of the transformant lines. The number of insertions in each line and their possible restriction maps were deduced from the hybridization data using the combinations of probes and digests shown in Figure 6. We examined two to six different individuals per line without pooling the DNA samples. The results were compared with the genetic observations, and in some cases, DNA methylation or DNA rearrangement could explain the lack of cosegregation between ADH activity and hygromycin resistance. In addition, two other structural aspects of the integrated DNA were revealed. First, the transforming DNA could be joined to genomic DNA at positions other than at the T-DNA right border sequence: in six lines (7, 10, 11, 21, 23, 27), apparent junction sites were located in the 3013 DNA near the 3' end of the ADH gene since new fragments were revealed by some probes and not adjacent ones, in a manner that was inconsistent with DNA insertion or deletion. Second, six transformant lines (2, 7, 11, 21, 23, 27) contained insertions of greater than a unit length. This conclusion is based on their having intact restriction fragments that contained the right border sequence while at the same time having junction fragments at or near a right border. These insertions will be referred to as tandem copies.

For all nine transformant lines, the genetic segregation of ADH/HPH activity was roughly correlated with the hybridization data, although the basis for weak growth on hygromycin in lines 3, 10 and 21 remains unclear. In the lines which had displayed a single locus of ADH/HPH activity (the first five lines listed in Table 1), the number of observed integration junctions agreed with a single transformed locus, which often consisted of tandem configurations. Typically an odd number of junction fragments was observed which is attributed to unobservable junction fragments that were larger than the mean length of the prepared DNA. The four lines (3, 27, 21 and 23) which had failed to show cosegregation are discussed below.

Line 3, which appeared to segregate 15:1 for ADH staining, contained at least two independent integrations of whAt3013 DNA, neither of which involved tandem copies. One copy of a *Bam*HI fragment detected with probe 8 (containing the ADH gene), and another detected with probe 5 (containing the HPH gene) were altered such that new fragments were seen (Figure 7). The extra fragment detected by probe 8 (also with probes 1, 10, and 11) is best explained by a deletion, but may have been related to an integration junction. The fact that it was only present in one of the two siblings (Figure 7) was probably a result of segregation since the parent plant appeared to contain two active ADH loci based segregation analysis. The extra copy detected by probe 5 in both siblings resulted from either a deletion or a junction. The presence of one set of intact copies and one set of altered copies may have been the basis for the 3:1 segregation of both allyl alcohol sensitivity and HPH activity. One explanation for having two loci of ADH staining with only one locus of allyl alcohol sensitivity is that both copies expressed ADH but expression of one copy was capable of escaping allyl alcohol selection as described earlier for line 27. Alternatively the alteration did not affect *cis*-acting sequences involved in ADH expression in the germinating root tip, but did affect sequences required for seed embryo ADH expression.

Line 27, which had variable levels of ADH activity and lacked hygromycin resistance, contained tandem copies and at least eight *Bgl*II fragments and four *Bam*HI fragments that were derived from junctions between vector DNA and genomic DNA. Probes 4 and 5 hybridized to only two of the *Bgl*II junction fragments. The remaining *Bgl*II junction fragments, detected by probes 2 and 3, were probably located near the right border. One of the junction fragments was located in the 3013 DNA just downstream of the ADH gene since probes 10 and 11 hybridized to both a new 5.9 kb *Bgl*II (Figure 8) fragment and a new 7.7 kb *Bam*HI fragment, while probe 8 did so very weakly. None of the data indicated rearrangements in the DNA to account for the reduced gene activity. However, the relative intensity of two *Sac*II HPH DNA fragments on a genome blot indicated that the fragment containing the 5' half of the chimeric HPH gene was present at a much lower level than the 3' half. Additionally, a large *Sac*II fragment (>20 kb) hybridized to the HPH gene probe.

Since sites containing 5-methylcytosine are resistant to digestion by *Sac*II, methylation of cytosine residues near the chimeric HPH gene was investigated further using blots containing transformant DNA digested separately with *Msp*I and *Hpa*II. *Msp*I and *Hpa*II both cleave the sequence 5'-CCGG-3', however *Msp*I will not cleave if the outer cytosine is methylated, and *Hpa*II will not cleave if the inner cytosine is methylated. In plants, the outer cytosine of the trinucleotide C-C-G is methylated at a lower frequency than that of the dinucleotide C-G (Gruenbaum *et al.*, 1981). Using the chimeric HPH gene as a probe (probe 5), methylation was more extensive in line 27 than in other lines and may also have been present in line 23 (Figure 9). For the other transformant lines (all lines except line 2 were examined), hypermethylation was not detected.

Line 21 displayed a 1:0 ratio for ADH activity and approximately a 1:1 ratio for weak hygromycin resistance. There were at least three independent loci in the genome since at least five *Bgl*II junction fragments were observed with probe 2. At least one of these contained tandem copies (Figure 10). An additional junction appeared to be in the 3013

fragment downstream of the ADH gene since probes 10 and 11 hybridized to both a new *Bam*HI fragment and a new *Bgl*II fragment which was only weakly detected by probe 8 (Figure 10). Another possible *Bgl*II junction or rearrangement in the 3013 DNA was detected with both probe 8 and 11. If at least three unlinked loci expressed the ADH gene, then the ADH segregation could be explained. The reason for weak hygromycin resistance is unknown since the HPH gene was intact and not detectably methylated.

Line 23, which lacked ADH activity, lacked the 9.8 kb *Bam*HI fragment which spans most of the 3013 DNA plus some vector sequence. The 3013 *Bgl*II fragment (which could not be distinguished from the endogenous fragment in blot hybridizations) was most likely absent also. With both *Bgl*II and *Bam*HI blots, three probes (8, 10 and 11) consisting of sequences internal to 3013, hybridized to a fragment 1.7 kb larger than the expected fragments (Figure 11); this new fragment may have been due to an insertion in the 3013 DNA since for both enzymes the fragment length increased by the same amount. In addition, there were two fragments roughly 1.5 kb and 4.5 kb smaller than expected which were detected with probes 10 and 11, and only faintly with probe 8 (Figure 11). For the *Bgl*II digest, the former fragment was also unexpectedly detected with probe 3 indicating that it might have been the result of partial digestion or a deletion. Further, with *Eco*RI genome blot hybridizations, one of the 0.76 kb fragments within the ADH gene was absent and replaced by at least two fragments totalling 2.0 kb (data not shown). An extra *Eco*RI fragment was detected with probe 11, but not probe 8, and its presence is most consistent with its being due to a junction. All the inserted copies could have been tandemly arranged or otherwise closely linked to give only one segregating locus of hygromycin resistance. At least one tandem insertion occurred, and there was evidence of only one locus based on the number of junction fragments.

T_2 and T_3 plants grown from allyl alcohol resistant seeds of six lines (lines 2, 4, 10, 11, 23 and 27) were analyzed by genome blot hybridization to verify that the absence of whAt3013 DNA cosegregated with the absence of ADH activity. As predicted, all the lines

examined except for lines 23 and 27 lacked whAt3013 DNA. In line 23, allyl alcohol treatment failed to select against transformed seeds since detectable ADH activity was absent in this line. In line 27, transformed seeds survived the allyl alcohol selection by virtue of expressing a low level of ADH as described earlier.

In addition to the configurations described above, other lines deviated from having an intact, single integration unit. Line 10, which had shown a single locus of ADH activity, unexpectedly lacked the 9.8 kb vector-3013 *Bam*HI fragment. Instead, there was a new 8 kb *Bam*HI fragment revealed with probe 8 (Figure 12). This new fragment was attributed to a right end junction between vector and genomic DNA. The junction appeared to be located 3' of the ADH gene because probes 3, 10 and 11 from downstream of the ADH gene failed to detect the new fragment. The fragment was not due to a deletion since probe 8, but not probes 2 and 11, detected a larger (10.5 kb) fragment on *Bgl*II blots. On *Eco*RI blots, probe 8 revealed two small fragments not seen with 10 or 11. The unusual position of the integration junction did not effect ADH expression, although it may have had something to do with the weak ADH staining in this line. The DNA from 5' of the ADH gene to 3' of the chimeric HPH gene appeared to be intact with an end junction near the right border. The reason for weak growth on hygromycin was not apparent from the hybridization data.

Lines 7 (in Figure 11) and 11 (data not shown) also seemed to contain new end junctions in the 3013 fragment downstream of the ADH gene based on the presence of new *Bgl*II, *Bam*HI, and/or *Eco*RI fragments in that region.

Finally, different restriction patterns were occasionally seen between siblings in the form of extra bands that could not be attributed to partial digestion (data not shown). These extra bands were seen in vector DNA of plants 7a.3.13, 10a.1, 10a.8, and 4e.13.3, and in 3013 DNA of plants 11b.7 and 27a.1 (27a.1 shown in Figure 8).

DISCUSSION

The work described in this chapter demonstrates complementation of an ADH null mutant with the wild-type ADH gene. Wild-type activity was restored in the mutant based on tissue-specific activity staining and allyl alcohol sensitivity. The coding region of the ADH gene may be useful as a biochemical tag in plant transformation experiments. Unlike other such tags, the potent selection against activity should allow for the selection of mutations in regulatory genes. We have shown that high and low levels of activity can be selected against differentially, as in *E. coli* (Lorowitz and Clark, 1982). Such selectivity would be useful for isolating mutations that reduce but do not eliminate activity.

This work is also the first detailed analysis of *Agrobacterium*-mediated transformation using a one border binary vector. The genetic and molecular data are generally consistent with results that have been described in other transgenic plants obtained by cocultivation with *Agrobacterium* containing two border constructs. Inheritance of the transformed gene appears to be Mendelian, and the introduced DNA seems to be genetically stable. Other aspects of the transformation, some of which are specific to *Agrobacterium*, are discussed below.

Polyploidy is commonly exhibited by plants regenerated from callus culture (Bayliss, 1980). In a study of regeneration of *Arabidopsis* from leaf-derived callus, approximately 30% of the plants were tetraploid (Gaj and Maluszynski, 1987), and in tomato transformations involving regeneration from stem-derived callus, both tetraploid and octoploid plants were produced (Chyi *et al.*, 1986). Polyploidy in cell culture can either occur during the culture or result from selection of pre-exisiting cells from which the culture was established (Bayliss, 1980).

For the tetraploid transformants described in this chapter, the initial transformation event probably occurred in a tetraploid cell, based on the genetic results showing that the T_1 plants were simplex (Aaaa) for the ADH/HPH loci. (An exception was plant 21a which

could have been, say, quadruplex for an ADH locus, although the HPH locus seemed to be simplex.) The other possibility is that the introduced DNA was not stably integrated in the plant genome until DNA replication occurred, after which in some cells a failure in cytokinesis resulted in tetraploidy. For many purposes, tetraploid transformants are undesirable. It may be possible to screen early for tetraploid transformants based on stomatal size or overall plant size. Tetraploidy and other products of tissue culture, such as somaclonal variants, plants with chromosomal rearrangements, and aneuploids might be avoided using the *Agrobacterium*-mediated seed transformation method which was shown in *Arabidopsis* (Feldman and Marks, 1987).

Another caution is that successful transfer of DNA into the plant genome does not guarantee expression of the gene. DNA rearrangements such as those in lines 3 and 23 may occur and thereby affect expression. Rearrangements, such as deletions, in the introduced DNA have been known to occur in transgenic plants (Chyi *et al.*, 1986) including *Arabidopsis* (Ken Feldman, personal communication) and in tumors incited by *Agrobacterium* (Kwok *et al.*, 1985). It is not known if the rearrangements occur before, during, or after integration. The "extra" restriction fragments that were occasionally seen between siblings were not likely to be due to instability since the original fragments were still present (unless the plant was mosaic). There was also no genetic evidence for loss of function which would result from deleterious rearrangement. The extra fragments probably represented aberrant segregating copies that were present in the original regenerated plants.

Even in the absence of DNA rearrangements, the introduced genes may not be active (Chyi *et al.*, 1986; Jones *et al.*, 1987). For example, the number of observed integrations may be greater than the number of genetic loci as in line 27 (see also Feldman and Marks 1987). It is not known whether or not the insertions were genetically linked or if some copies were inactive. Related to this is the fact that expression levels of transformed genes do not correlate with copy number (Jones *et al.*, 1987; Jones *et al.*, 1985; Eckes *et al.*, 1986). For example, tetraploid line 4e contained a single, nontandem integration and

displayed strong hygromycin resistance, while tetraploid line 21 contained multiple integrations with tandem copies and had weak resistance. Further, linked markers on the same T-DNA construct do not always cosegregate (Eckes *et al.*, 1986; Chyi *et. al.*, 1986; Jones *et al.*, 1987), as seen in line 21. The reasons for these results have not been established, but position effects, which are important in transformation of *Drosophila* (Spradling and Rubin, 1983; Giangrande *et al.*, 1987) and mice (Palmiter and Brinster, 1986), are suspected.

One explanation for reduced gene activity in the absence of DNA rearrangement is DNA methylation. Cytosine methylation in animals can be correlated in many cases with reduced transcription activity (Cedar, 1988). T-DNA expression in plants has also been shown to be dependent on methylation (Amasino *et al.*, 1984; Hepburn *et al.*, 1983). Amasino *et al.* showed that suppression of T-DNA gene transcription by methylation was stably inherited. The lack of hygromycin resistance in line 27 was associated with methylation of the chimeric HPH gene. Methylation was detected in both inner and outer cytosines (more in the inner cytosines) of the restriction site 5'-CCGG-3', consistent with observed methylation in plants (Gruenbaum *et al.*, 1981). The chimeric HPH gene and nearby sequences were hypermethylated since in *Arabidopsis* less than 6% of the cytosine residues are methylated (Leutwiler *et al.*, 1984). Methylation is known to be tissue and stage specific, so it is possible that the chimeric HPH gene was not subject to methylation in tissue culture; this explains the otherwise puzzling fact that line 27 survived the initial selection of callus and shoots on hygromycin media. Methylation in other transformants may have been present in a tissue or stage specific manner that was not detectable in DNA preparations of whole plants.

In a variety of transgenic plants and tumor lines incited by *Agrobacterium*, multiple integrations are common. They can be in the form of direct repeats, inverted repeats, or unlinked copies (Zambryski *et al.*, 1982; Simpson *et al.*, 1982; Holsters *et al.*, 1983; Kwok *et al.*, 1985; Spielmann and Simpson, 1986; Jorgensen *et al.*, 1987; Jones *et al.*,

1987; Feldman and Marks, 1987; Budar *et al.*, 1986). Direct repeats which have been analyzed by DNA sequencing have been found to be separated by rearranged T-region sequences of approximately 22 to 389 base pairs (Zambryski *et al.*, 1982; Holsters *et al.*, 1983). Generally when both border sequences are present the unit of integration is from right border to left border, although the precise location of junction sites is variable; a detailed structural analysis of 11 tomato transformants showed that all the right junctions mapped within 1.1 kb of the right border sequence and all the left junctions mapped within 3 kb of the left border sequence for a construct containing approximately 11 kb between the borders (Jorgensen *et al.*, 1987). T-DNA junctions that were cloned from tumor lines also showed heterogeneous junction sites, with less variation at the right border (Zambryski *et al.*, 1982; Yadav *et al.*, 1982). Jen and Chilton (1986) examined the structure of two single-border vectors (one right border and one left border) in tumor lines using one line per construct. In each line they found a junction somewhere in the vector which was short of a unit length of integration.

In our transformants, we found multiple loci (within single plants) and what appeared to be tandem copies. What was striking was that six lines (7, 10, 11, 21, 23, 27) seemed to contain terminal junctions within the 3013 DNA downstream of the ADH gene (roughly 5 kb from the right border sequence). The frequency of junctions around this particular location suggests that there exists a cryptic border site(s) downstream of the *Arabidopsis* ADH gene. Why would this location be chosen over the actual right border sequence? Perhaps the left border sequence signals the stop of transfer so that with constructs lacking a left border, transfer continues until a cryptic border is reached or a particular length is obtained. The cryptic site may also function as a right border. In line 10 the cryptic site may have been where transfer was initiated. Since these lines show intact restriction fragments containing the right border sequence, it is possible that the so-called tandem copies resulted from utilization of the cryptic site. It is not known whether the right border

was bypassed one or more times, or at all. The only other line with tandem copies was line 2 which did not show evidence of a junction in the 3013 DNA.

In tomato, tobacco and petunia (Jorgensen *et al.*, 1987; Jones *et al.*, 1987) and *Arabidopsis* (K. Feldman, personal communication), inverted tandem repeats were exhibited in 50% or more of the transgenic plants obtained using the nopaline *Agrobacterium* strain C58/pGV3850Kan^r, but were much less frequent when octopine strains were used. Yet, in our transformation using a C58 derivative, none of the structures appeared to be inverted repeats. The major difference between our study and the ones cited is that our binary vector contained only the right border sequence. This suggests that the inverted structure is somehow related to having both left and right borders.

The complexities of plant transformation are apparent in this study. It appears that both position effects and the structure of the transferred DNA affect gene activity. The occurrence of nonfunctional inserts is unlikely to be unique to *Agrobacterium*-mediated transformation. The origins of the peculiar structural aspects related to *Agrobacterium* are not understood, but any future models for DNA transfer and integration by this method must account for the observations described here. Nonetheless, these results show that it is possible to obtain single insertions of single-copy genes that are segregated stably over many generations, and expressed appropriately in time and tissue, at normal levels. In addition, mutations in an endogenous gene can be stably and fully complemented by an introduced copy of the same gene derived from a wild-type plant. Thus, with sufficient caution, the methods described here can give stable, well-behaved transgenic *Arabidopsis* plants.

REFERENCES

Amasino, R.M., Powell, A.L.T., and Gordon, M.P. (1984). Changes in T-DNA methylation and expression are associated with phenotypic variation and plant regeneration in a crown gall tumor line. *Mol. Gen. Genet.* 197, 437-446.

Ambros, P., and Schweizer, D. (1976). The giemsa C-banded karyotype of *Arabidopsis thaliana* (L.) Heynh. *Arabidopsis Information Service* 13, 167-171.

Bayliss, M.W. (1980). Chromosomal variation in plant tissues in culture. *Intern. Rev. Cytol.* Suppl. 11A, 113-144.

Budar, F., Thia-Toong, L., Van Montagu, M., and Hernalsteens, J.-P. (1986). *Agrobacterium*-mediated gene transfer results mainly in transgenic plants transmitting T-DNA as a single Mendelian factor. *Genetics* 114, 303-313.

Cedar, H. (1988). DNA methylation and gene activity. *Cell* 53, 3-4.

Chang, C., and Meyerowitz, E.M. (1986). Molecular cloning and DNA sequence of the *Arabidopsis* alcohol dehydrogenase gene. *Proc. Natl. Acad. Sci. USA* 83, 1408-1412.

Chyi, Y.-S., Jorgensen, R.A., Goldstein, D., Tanksley, S.D., and Loaizi-Figueroa, F. (1986). Locations and stability of *Agrobacterium*-mediated T-DNA insertions in the *Lycopersicon* genome. *Mol. Gen. Genet.* 204, 64-69.

Ditta, G., Stanfield, S., Corbin, D., and Helinski, D.R. (1980). Broad host range DNA cloning system for Gram-negative bacteria: Construction of a gene bank of *Rhizobium meliloti*. *Proc. Natl. Acad. Sci. USA* 77, 7347-7351.

Eckes, P., Rosahl, S., Schell, J., and Willmitzer, L. (1986). Isolation and characterization of a light-inducible, organ-specific gene from potato and analysis of its expression after tagging and transfer into tobacco and potato shoots. *Mol. Gen. Genet.* 205, 14-22.

Feldman, K.A., and Marks, M.D. (1987). *Agrobacterium*-mediated transformation of germinating seeds of *Arabidopsis thaliana*: A non-tissue culture approach. *Mol. Gen. Genet.* 208, 1-9.

Fraley, R.T., Rogers, S.G., Horsch, R.B., Eichholtz, D.A., Flick, J.S., Fink, C.L., Hoffman, N.L., and Sanders, P.R. (1985). The SEV system: a new disarmed Ti plasmid vector system for plant transformation. *Biotechnology* 3, 629-635.

Gaj, M.D. and Maluszynski, M. (1987). Genetic variation in callus culture of *Arabidopsis thaliana* (L.) Heynh. *Arabidopsis Information Service* 23, 1-8.

Gobel, U., Maas, R., and Clad, A. (1987). Quantitative electroelution of oligonucleotides and large DNA fragments from gels and purification by electrodialysis. *J. Biochem. Biophys. Methods*. 14, 245-260.

Giangrande, A., Mettling, C., and Richards, G. (1987). *Sgs-3* transcript levels are determined by multiple remote sequence elements. *EMBO J.* 6, 3079-3084.

Gruenbaum, Y., Naveh-Many, T., Cedar, H., and Razin, A. (1981). Sequence specificity of methylation in higher plant DNA. *Nature* 292, 860-862.

Hepburn, A.G., Clarke, L.E., Pearson, L., and White, J. (1983). The role of cytosine methylation in the control of nopaline synthase gene expression in a plant tumor. *J. Mol. Appl. Genet.* 2, 315-329.

Hoekema, A., Hirsch, P.R., Hooykaas, P.J.J., Schilperoort, R.A. (1983). A binary plant vector strategy based on separation of *vir*- and T-region of the *Agrobacterium tumefaciens* Ti plasmid. *Nature* 303, 179-180.

Holsters, M., Villarroel, R., Gielen, J., Seurinck, J., DeGreve, H., Van Montagu, M., and Schell, J. (1983). An analysis of the boundaries of the octopine TL-DNA in tumors induced by *Agrobacterium tumefaciens*. *Mol. Gen. Genet.* 190, 35-41.

Horsch, R.B., Fraley, R.T., Rogers, S.G., Sanders, P.R., Lloyd, A., and Hoffmann, N. (1984). Inheritance of functional foreign genes in plants. *Science* 223, 496-498.

Jen, G.C., and Chilton, M.-D. (1986). Activity of T-DNA borders in plant cell transformation by mini-T plasmids. *J. Bacteriol.* 166, 491-499.

Jones, J.D.G., Dunsmuir, P., and Bedbrook, J. (1985). High level of expression of introduced chimaeric genes in regenerated transformed plants. *EMBO J.* 4, 2411-2418.

Jones, J.D.G., Gilbert, D.E., Grady, K.L., and Jorgensen, R.A. (1987). T-DNA structure and gene expression in petunia plants transformed by *Agrobacterium tumefaciens* C58 derivatives. *Mol. Gen. Genet.* 207, 478-485.

Joos, H., Inze, D., Caplan, A., Sormann, M., Van Montagu, M., and Schell, J. (1983). Genetic analysis of T-DNA transcripts in nopaline crown galls. *Cell* 32, 1057-1067.

Jorgensen, R., Snyder, C., and Jones, J.D.G. (1987). T-DNA is organized predominantly in inverted repeat structures in plants transformed with *Agrobacterium tumefaciens* C58 derivatives. *Mol. Gen. Genet.* 207, 471-477.

Kwok, W.W., Nester, E.W., and Gordon, M.P. (1985). Unusual plasmid DNA organization in an octopine crown gall tumor. *Nucleic Acids Res.* 13, 459-471.

Leutwiler, L.S., Hough-Evans, B.R., and Meyerowitz, E.M. (1984). The DNA of *Arabidopsis thaliana*. *Mol. Gen. Genet.* 194, 15-23.

Lloyd, A., Barnason, A.R., Rogers, S.G., Byrne, M.C., Fraley, R.T., and Horsch, R.B. (1986). Transformation of *Arabidopsis thaliana* with *Agrobacterium tumefaciens*. *Science* 234, 464-466.

Lorowitz, W., and Clark, D. (1982). *Escherichia coli* mutants with a temperature-sensitive alcohol dehydrogenase. *J. Bacteriol.* 152, 935-938.

Murashige, T. and Skoog, F. (1962). A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol. Plant.* 15, 473-497.

Negrutiu, I., Jacobs, M., and Cattoir-Reynaerts, A. (1984). Progress in cellular engineering of plants: biochemical and genetic assessment of selectable markers from cultured cells. *Plant Molecular Biology* 3: 289-302.

Odell, J.T., Nagy, F., and Chua, N.H. (1985). Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. *Nature* 313, 810-812.

Palmiter, R.D., and Brinster, R.L. (1986). Germ-line transformatin of mice. *Ann. Rev. Genet.* 20, 465-499.

Rédei, G.P., and Hirono, Y. (1964). Linkage studies. *Arabidopsis Information Service* 1, 9-10.

Rigby, P.W.J., Dieckmann, M., Rhodes, C., and Berg, P. (1977). Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick-translation with DNA polymerase I. *J. Mol. Biol.*113, 237-251.

Robbelen, G. (1965). The Laibach standard collection of natural races. *Arabidopsis Information Service* 2, 36-47.

Rogers, S.G., Horsch, R.B., and Fraley, R.T. (1986). Gene transfer in plants: Production of transformed plants using Ti plasmid vectors. *Methods in Enzymology* 118, 627-640.

Rogers, S.G., Klee, H.J., Horsch, R.B., and Fraley, R.T. (1987). Improved vectors for plant transformation: Expression cassette vectors and new selectable markers. *Methods in Enzymology* 153, 253-277.

Rothstein, S.J., Lahners, K.N., Lotstein, R.J., Carozzi, N.B., Jayne, S.M., and Rice, D.A. (1987). Ti Promoter cassettes, antibiotic-resistance genes, and vectors for plant transformation. *Gene* 53, 153-161.

Simpson, R.B., O'Hara, P.J., Kwok, W., Montoya, A.L., Lichtenstein, C., Gordon, M.P., and Nester, E.W. (1982). DNA from the A6S/2 crown gall tumor contains scrambled Ti-plasmid sequences near its junctions with plant DNA. *Cell* 29, 1005-1014.

Spielmann, A., and Simpson, R.B. (1986). T-DNA structure in transgenic tobacco plants with multiple independent integration sites. *Mol. Gen. Genet.* 205, 34-41.

Spradling, A.C., and Rubin, G.M. (1983). The effect of chromosome position on the expression of the *Drosophila* xanthine dehydrogenase gene. *Cell* 34, 47-57.

Stachel, S.E., and Zambryski, P.C. (1986). *Agrobacterium tumefaciens* and the susceptible plant cell: a novel adaptation of extracellular recognition and DNA conjugation. *Cell* 47, 155-157.

Waldron, C., Murphy, E.B., Roberts, J.L., Gustafson, G.D., Armour, S.L., and Malcolm, S.K. (1985). Resistance to hygromycin B. *Plant Molecular Biology* 5, 103-108.

Yadav, N.S., Vanderleyden, J., Bennett, D.R., Barnes, W.M., and Chilton, M.-D. (1982). Short direct repeats flank the T-DNA on a nopaline Ti plasmid. *Proc. Natl. Acad. Sci. USA* 79, 6322-6326.

Zambryski, P., Depicker, A., Kruger, K., and Goodman, H.M. (1982). Tumor induction by *Agrobacterium tumefaciens*: analysis of the boundaries of T-DNA. *J. Mol. Appl. Genet.* 1, 361-370.

Zambryski, P., Joos, H., Genetello, C., Leemans, J., Van Montagu, M., and Schell, J. (1983). Ti-plasmid vector for the introduction of DNA into plant cells without alteration of their normal regeneration capacity. *EMBO J.* 2, 2143-2150.

Table 1. Segregation of phenotypes in the T_2 generation. The proportions of T_2 individuals that were positive for ADH staining (ADH stain), sensitive to allyl alcohol (AA^S), and resistant to hygromycin (Hyg^r) are given for self-progeny of thirteen regenerated plants (including A1 which was transformed with ASE minus the vector) plus control lines. The data are a summary of up to five separate experiments. The total number of individuals scored is shown in ( ). Each individual was scored for only one of the three phenotypes. All values conform with a 3:1 segregation at a 10% level of significance unless noted by an asterisk.

T_1 or Control	T_2 Segregation		
	ADH stain	AA^S	Hyg^r
2a	0.70 (44)	0.80 (40)	0.71 (31)
d	0.68 (19)	___	0.74 (27)
4e	0.68 (34)	0.79 (90)	0.73 (49)
7a	0.76 (21)	0.79 (19)	0.64 (39)
10a	0.83 [1] (53)	0.72 (43)	0.54 (39) *
11b	0.70 (27)	0.53 (32) *	0.59 (17)
3a	0.94 [2] (84) **	0.84 (31)	0.68 (25)
27a	0.95 [3] (154) **	0.81 (36)	0.0 (52) **
21a	1.0 (28) **	1.0 (28) **	0.53 (34) *
b	1.0 (21) *	___	0.57 (44) *
23d	0.0 (44) **	0.0 (37) **	0.78 (65)
b	0.0 (49) **	___	___
A1	0.0 (50)	___	0.0 (125)
R002	0.0 (200)	0.0 (200)	___
Be-0	1.0 (200)	1.0 (200)	___
Ler	1.0 (50)	1.0 (54)	___

[1] Includes 2 weak positives observed in 1 of 4 experiments.
[2] " 7 " " " " 2 " " ".
[3] " 8 " " " " 3 " " ".

* <1% and ** <0.1% level of significance for 3:1 segregation based on either the χ^2 value or the binomial expansion.

Table 2. Segregation of ADH activity in line 27. The proportions of T_3 individuals are shown for normal and weak ADH staining, and for sensitivity to allyl alcohol (100 mM for two hours). The total number of individuals scored is shown in ( ). AA^r indicates that plants were selected with 35 mM allyl alcohol at the seed stage. Individuals that were scored for staining were different from those treated with allyl alcohol. The most likely genotypes for ADH activity were deduced from both the T_2 and T_3 segregation data, assuming A is the locus for normal staining, and B the locus for weak staining.

T_2	Adh staining in T_3			T_2 genotype (most likely)
	Normal	Weak		
27a.1	0.0	0.0	(16)	aa; bb
2	0.79	0.21	(14)	Aa; Bb
3	0.0	1.0	(26)	aa; BB
4	0.75	0.0	(8)	Aa; bb
5	0.69	0.0	(16)	Aa; BB
AA^r 27a.1	0.0	0.58 *	(33)	aa; Bb
27a.2	0.0	1.0	(11)	aa; BB

* P (3:1) < 0.05

Table 3a, b. Genetic evidence for tetraploidy in transformant lines 10a and 2a.

(a) For hygromycin selected T_2 plants of 10a and 2a, Table A gives the proportion of T_3 generation individuals that were positive for Adh staining (Adh), resistant to hygromycin (Hyg^r), and sensitive to allyl alcohol (AA^s). The number shown in ( ) is the number of individuals scored. Each individual was assayed for only one of the three phenotypes. The most likely T_2 genotypes were deduced as described in the text, based on the phenotypic segregation data. To calculate the χ^2 value for the segregation in 10a.5, the data for Adh, Hyg^r, and AA^s were pooled. Exact probabilities for the segregation of 10a.8 were calculated using the binomial expansion.

(b) Table B gives T_4 segregation for random progeny of 10a.8 and for hygromycin-selected progeny of 2a.4.

A

T_2	T_3 Segregation: Adh stain	AA^s	Hyg^r	T_2 Genotype (Most likely)
10a.1	0.74 (115)			Aaaa
2	0.78 (40)		0.80 (40)	Aaaa
3	0.79 (29)			Aaaa
4	0.73 (33)			Aaaa
5	0.97 (69) *	0.97 (110) *	0.98 (40) *	AAaa
6	1.00 (44)			AAaa
7	0.68 (25)			Aaaa
8	1.00 (49) †			AAaa
2a.1	0.78 (27)			Aaaa
2	0.77 (39)			Aaaa
3	0.95 (20)			AAaa
4	0.84 (73)		0.71 (76)	Aaaa
5	0.63 (35)			Aaaa
6	1.00 (28)			AAaa
7	0.97 (106)			AAaa
8	0.80 (41)			Aaaa

* $\chi 2$ (3:1) = 56.7; $P < 0.01$
$\chi 2$ (35:1)= 0.03; $P > 0.8$

† P (3:1) = 7.6E-7
P (35:1) = 0.25

B

T_3	T_4 Segregation: Adh stain	AA^s	Hyg^r	T_3 Genotype (Most likely)
10a.8.1	0.78 (58)			Aaaa
2	0.62 (34)			Aaaa
3	1.00 (37)			AAaa
4	0.79 (75)			Aaaa
2a.4.1	1.00 (87)			AAaa
2	0.93 (30)	0.96 (56)	0.98 (47)	AAaa
3	0.89 (36)	0.98 (130)	0.95 (40)	AAaa
4	0.65 (40)			Aaaa
5	0.76 (62)			Aaaa
6	0.67 (36)			Aaaa
7	0.80 (51)			Aaaa
8	0.76 (38)			Aaaa
9	0.76 (45)			Aaaa
10	0.63 (48)			Aaaa
11	1.0 (156)		0.98 (53)	AAAa

Figure 1. A schematic diagram, not to scale, of the binary transformation vector and ADH gene insert. The vector contains the right border sequence and nopaline synthase gene (NOS) of the Ti plasmid. There are two markers for selection in plants: hygromycin phosphotransferase (HPH) and neomycin phosphotransferase (NPT II). The HPH coding sequence is fused at its 5' end to the constitutive 35S RNA promoter of cauliflower mosaic virus and at its 3' end to the 3' untranslated DNA of the NOS gene. The NPT II sequence is fused at its 5' end to the NOS promoter and at its 3' end to the 3' untranslated DNA of the NOS gene. Spectinomycin and streptomycin (Spc/Str) resistance is conferred by the plasmid to bacteria.

whAt3013

(pMON410.1 + 3013)

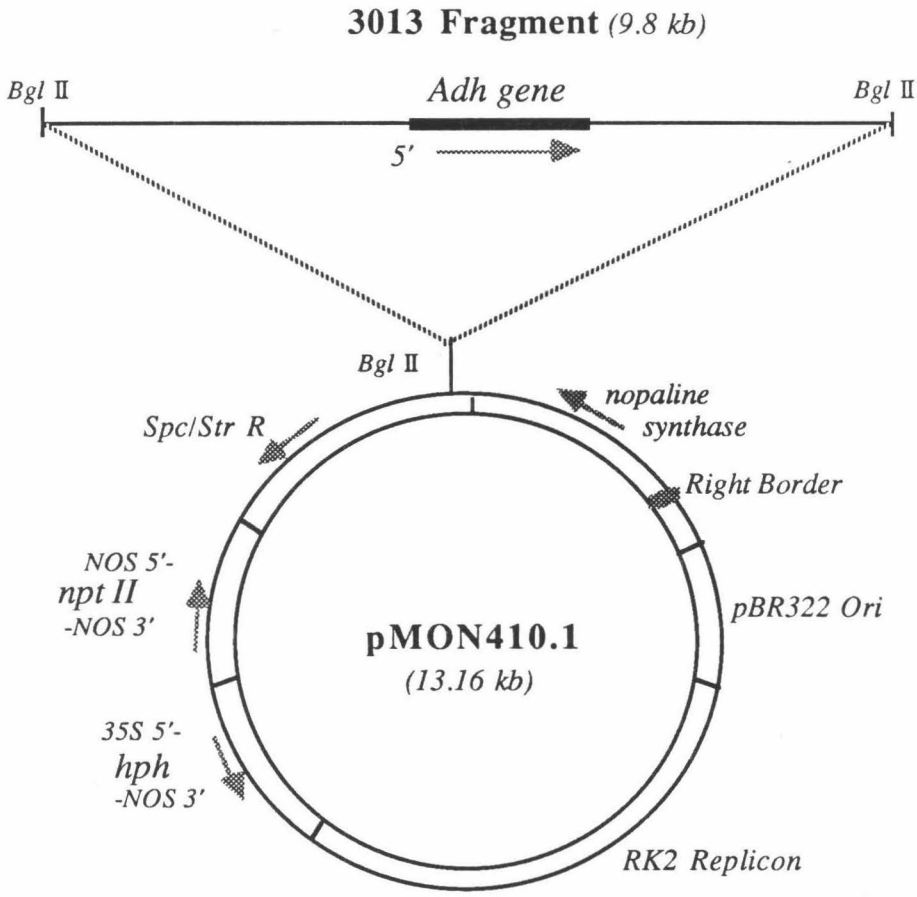


Figure 2a, b. ADH histochemical staining in young roots of transformed ADH null line R002.

(a) Wild-type levels of ADH staining in young roots of 3a progeny.

(b) The range of ADH staining among progeny of 27a. Also shown are negative R002 and positive Be-0 controls.

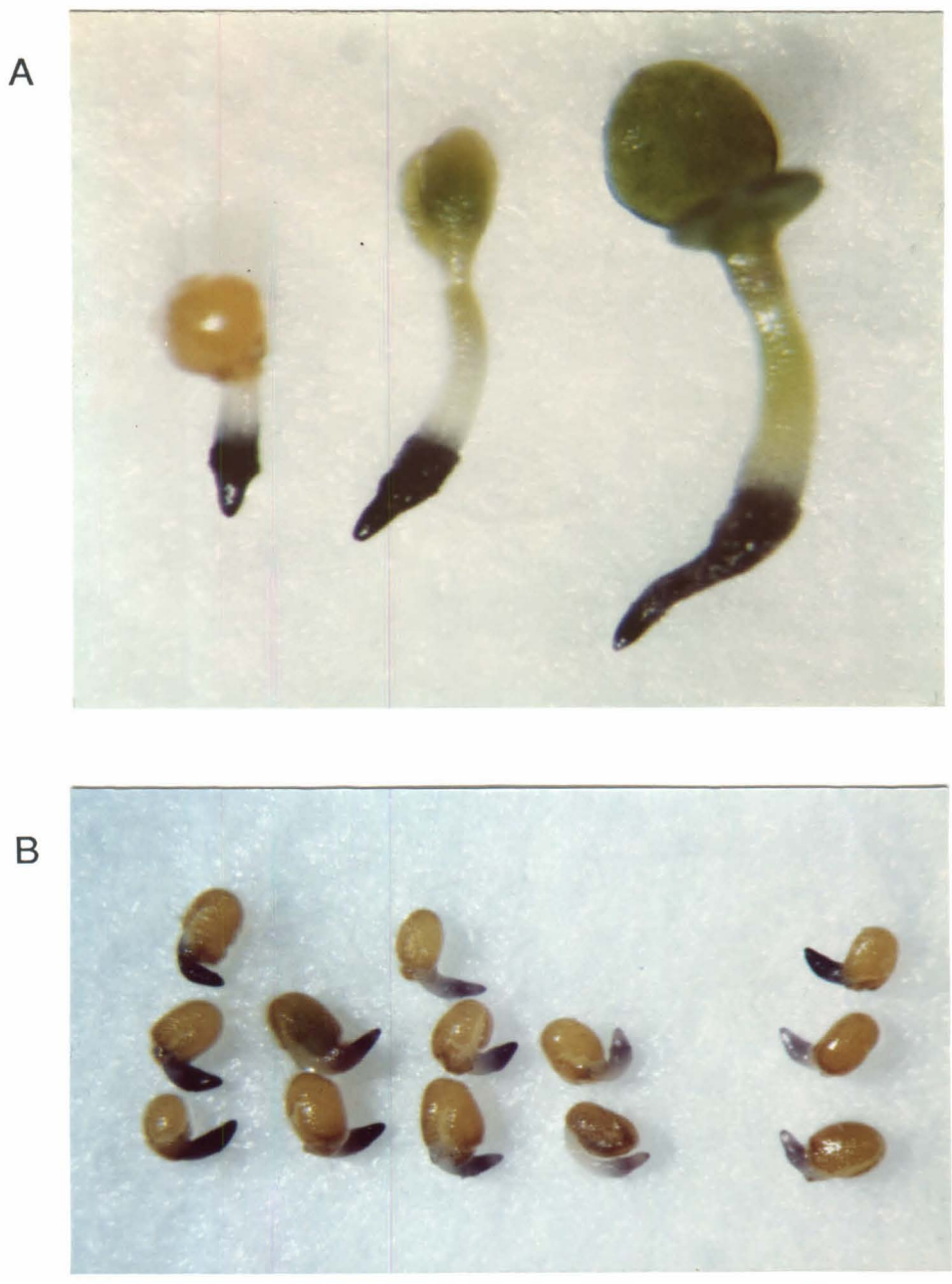

A
B
Strong positive
Weak positive
Weak positive
Negative
2 R002, 1 Be-0

Figure 3. Sensitivity of line 27a.3 to two doses (35 mM and 100 mM) of allyl alcohol compared with that of R002. Sensitivity was measured by the ability to germinate over a period of 12 days. No allyl alcohol was used for a germination control.

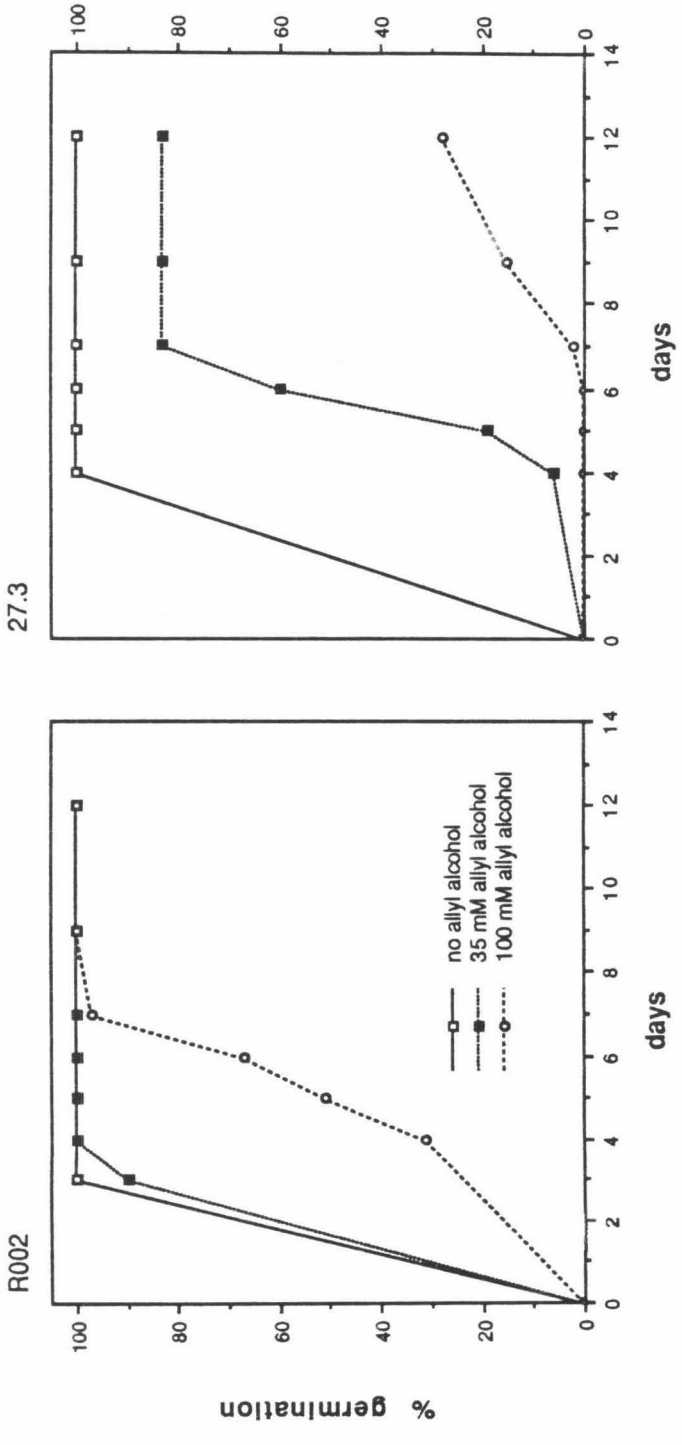

R002
27.3
% germination
days
days
no allyl alcohol
35 mM allyl alcohol
100 mM allyl alcohol
0
20
40
60
80
100
0
2
4
6
8
10
12
14

Figure 4. Hygromycin resistance segregating 3:1 in a transformed line.

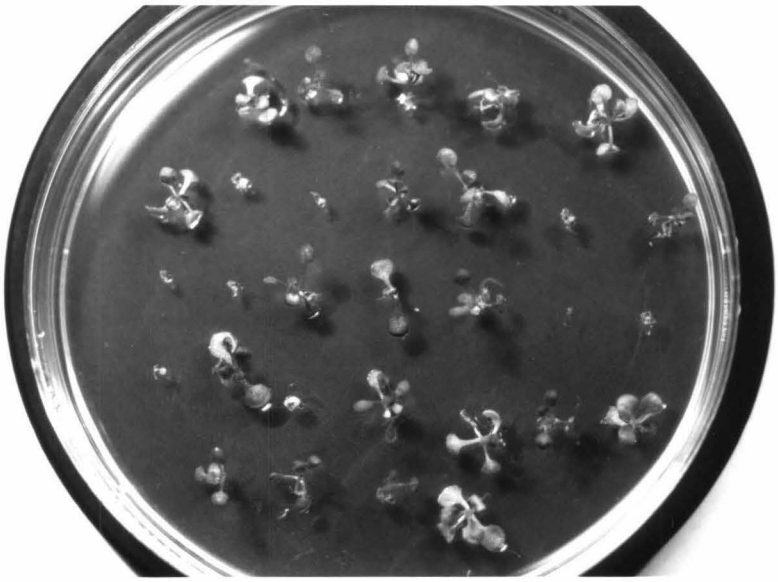


Figure 5a, b. Distributions of chromocenter counts and seed lengths in relation to ploidy. For R002 and each transformant line.

(a) The graph shows the mean number of chromocenters and the standard error of the mean. Counts were based on 14 to 68 nuclei per line.

(b) The graph shows the mean seed length and the standard error of the mean. Measurements were based on 10 to 20 seeds per line.

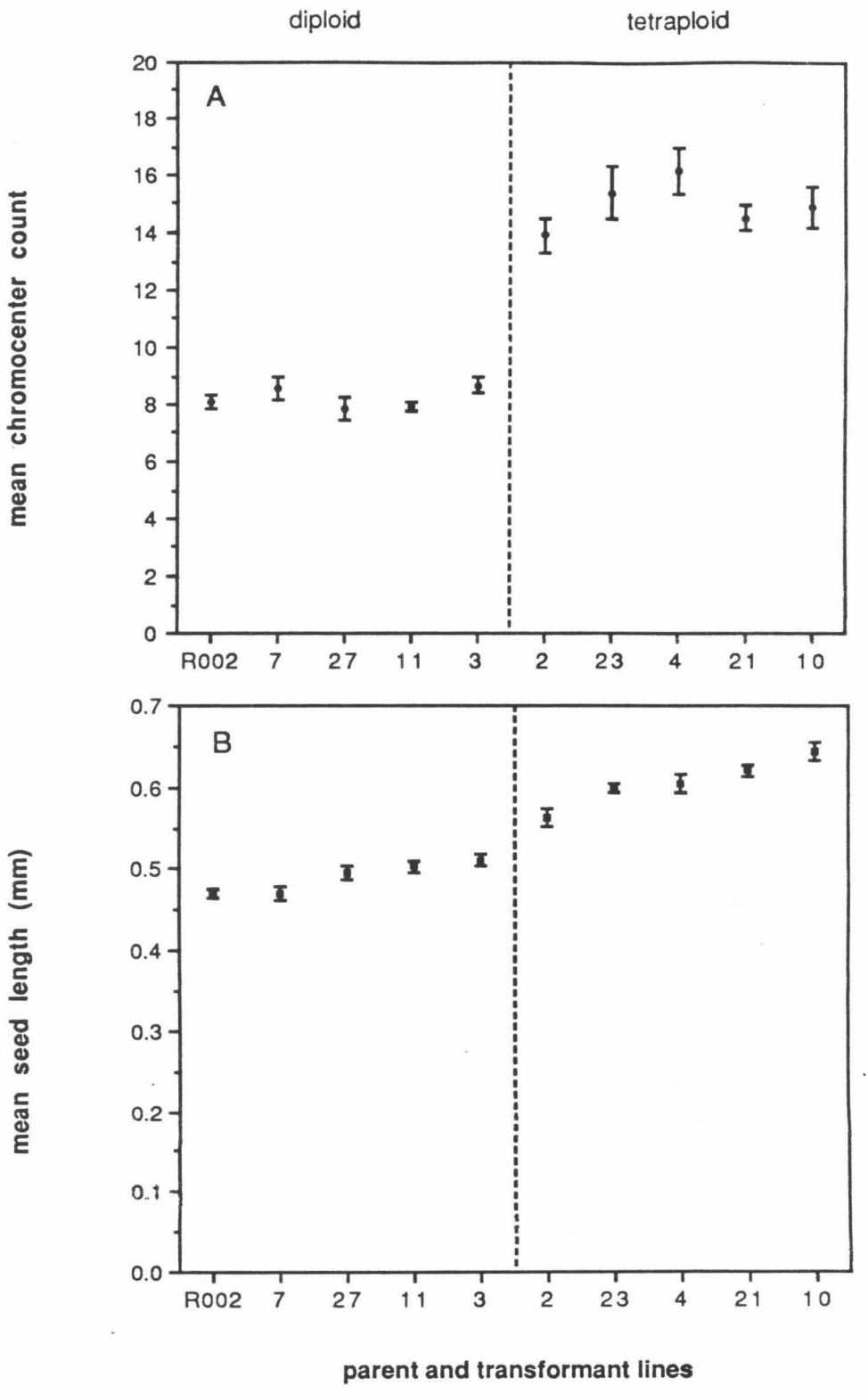
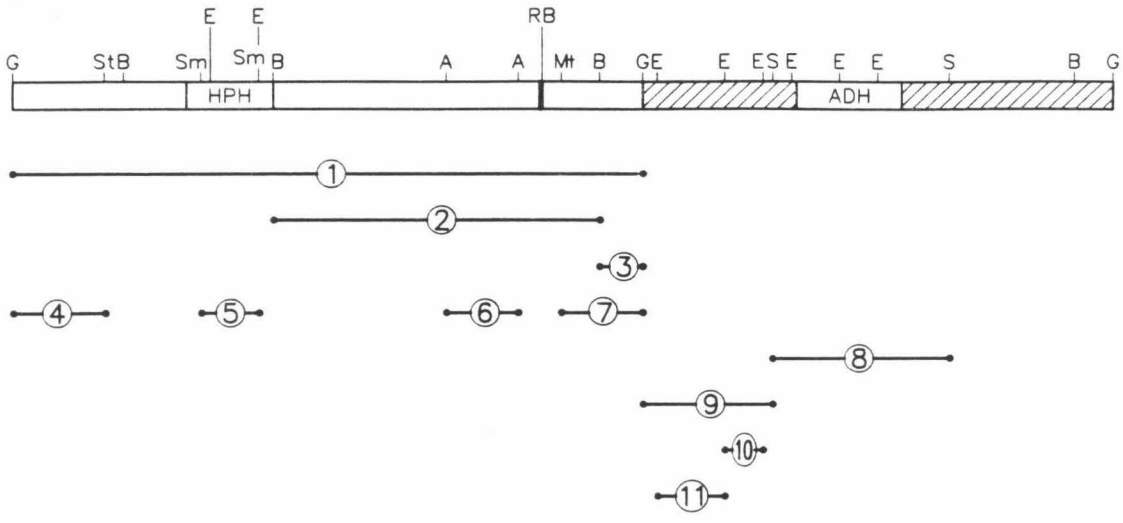

diploid
tetraploid
A
B
mean chromocenter count
mean seed length (mm)
R002 7 27 11 3 2 23 4 21 10
parent and transformant lines

Figure 6a, b. Genome blot hybridization strategy.

(a) The 11 probes used are shown in relation to the restriction map of whAt3013. RB designates the T DNA right border. The restriction enzyme code for (a) and (b) is: A-*Ava*I, B-*Bam*HI, E-*Eco*RI, G-*Bgl*II, H-*Hpa*II, M-*Msp*I, Mt-*Mst*II, S-*Sac*I, Sm-*Sma*I, St-*Stu*I.

(b) The genome blots that were hybridized with specific probes are shown for each transformant line and the R002 parent. The blots are designated by a single letter code that corresponds to the restriction enzyme used to digest the genomic DNA. One or more individuals per line were represented in a blot. The individuals used were: 2a.2, 2a.4, 2a.12, 4e.5, 4e.7, 4e.13.1, 4e.13.2, 4e.13.3, 10a.1, 10a.2, 10a.3, 10a.8, 10a.9, 10a.6.1, 10a.6.2, 7a.3.4, 7a.3.13, 7a.4.1, 7a.4.14, 11b.3, 11b.7, 11b.4.1, 11b.4.2, 21a.1, 21a.2, 21a.33, 21.6.1, 21a.6.2, 23d.2, 23d.17, 23d.19 and 23d.20 (all selected on hygromycin); 3a.2.2.1, 3a.2.2.2 and 27a.6 (all unselected); 2a.12.2, 4e.1, 4e.2, 4e.6.1, 4e.7.1, 4e.7.2, 10a.1, 10a.2, 10a.3, 11b.1, 11b.2, 23d.1, 23d.2, 27a.1 and 27a.2 (all selected with allyl alcohol).

A



B

Probes

Parent and transformant lines	1	2	3	4	5	6	7	8	9	10	11
2a	B	B,E,G	B,E	B,E	B	B,E	B	B,E,G	B	B,E	E,G
3a	B	S		B	B, S M,H	B		B		B	B
4e	B	B,E,G S	B,E	B,E	B,S M,H	B,E		B,E,G	B	B,E	B,E,G
7a	B	G		B	B, M,H	B		B, G	B	B	B, G
10a	B	B,E,G S	B,E	B,E	B,S M,H	B,E		B,E,G	B	B,E	B,E,G
11b	B,G	B,E,G S	B,E,G	B,E,G	B,S,G M,H	B,E,G		B,E,G	B,G	B,E,G	B,E,G
21a	B	G S		B	B, S M,H	B		B, G	B	B	B, G
23d	B,G	E,G	E,G	B,E,G	B, G M,H	B,E,G		B,E,G	B,G	B,E,G	B,E,G
27a	B,G	G S	G	B, G	B,S,G M,H	B, G		B, G	G	B, G	B, G
R002	B	B	B	B	B			B	B	B	

Figure 7. Aberrant fragments in line 3a are shown in a *Bam*HI genome blot hybridized separately with 3013 sequences (probe 11) and HPH sequences (probe 5). For comparison, hybridizations to whAt3013 and line 4e are shown. Probe 11 was expected to hybridize to both the endogenous 22 kb fragment and the introduced 9.8 kb fragment which contains the transforming ADH gene (derived from 3013 and vector *Bam*HI sites). An extra 6.7 kb fragment was revealed in 3a.2.2.1, but not 3a.2.2.2. Probe 5 hybridized to an extra 2.95 kb fragment in both 3a plants.

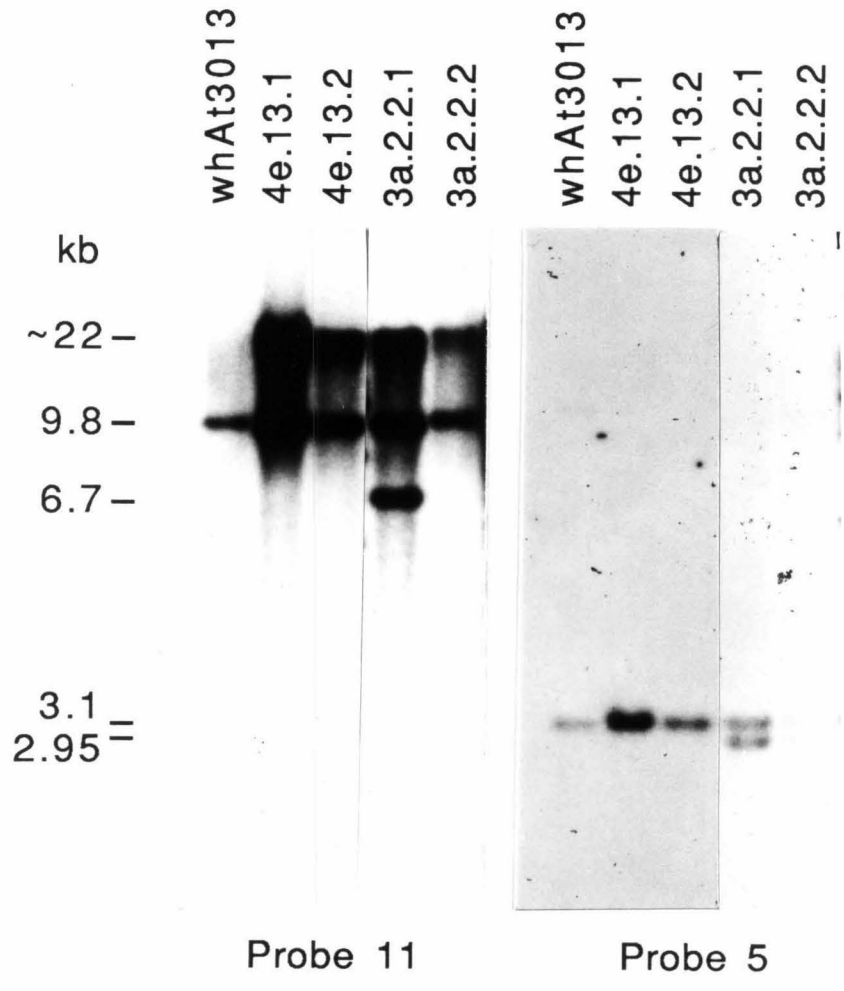

whAt3013
4e.13.1
4e.13.2
3a.2.2.1
3a.2.2.2
whAt3013
4e.13.1
4e.13.2
3a.2.2.1
3a.2.2.2
kb
~22
9.8
6.7
3.1
2.95
Probe 11
Probe 5

Figure 8. Partial evidence for a junction fragment in 3013 DNA in plants 27a.1 and 27a.2 (both selected with allyl alcohol). Sizes are given for the major fragments of a *Bgl*II genome blot hybridized separately with probes 11 and 8. Plant 11b.1 is shown for comparison. Both probes detect the endogenous 9.8 kb fragment containing the ADH gene. Probe 11 also detects a small endogenous fragment (unlabeled). The 5.9 kb fragment is not detected with probe 8. Probe 8 detects the 13.2 kb vector fragment due to contamination of the probe with vector DNA. Note that there is a small "extra" fragment in plant 27a.1 detected with probe 11.

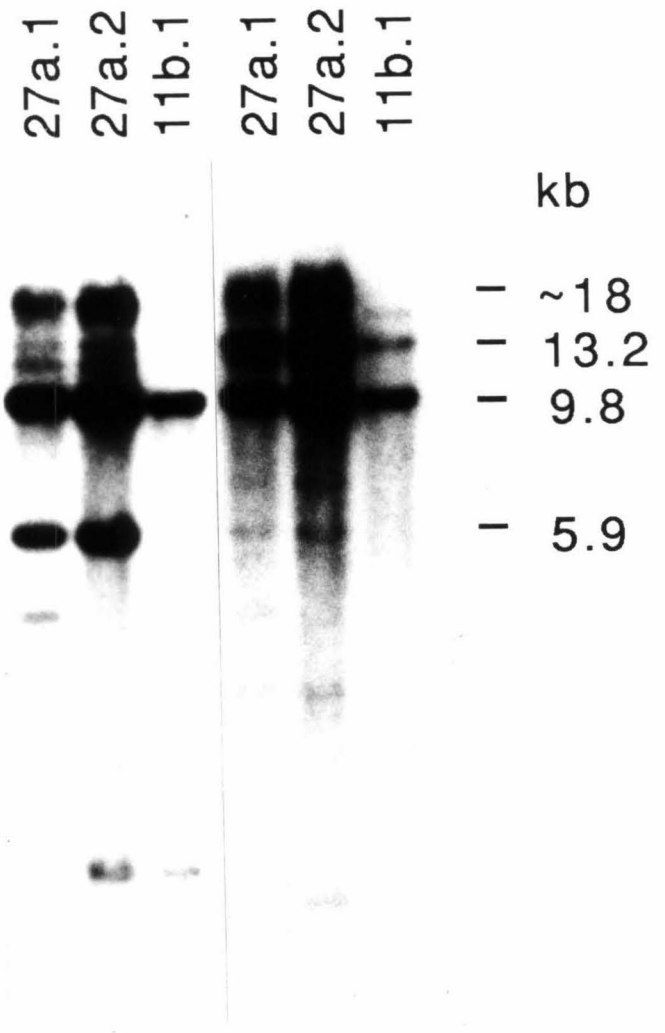

27a.1
27a.2
11b.1
27a.1
27a.2
11b.1
kb
~18
13.2
9.8
5.9
Probe 11
Probe 8

Figure 9. Analysis of cytosine methylation in the chimeric HPH gene using *Msp*I (M) and *Hpa*II (H) restriction digests. Different amounts of DNA per plant were placed in the gel. However for each pair of digests, relatively equal amounts of DNA were loaded. Part of the chimeric HPH gene (probe 5) was used as a probe. Digestion was inhibited in plants 27a.1 and 27a.2 (both selected with allyl alcohol) and plant 23d.1. The differential pattern between *Msp*I and *Hpa*II suggested the presence of methylation.

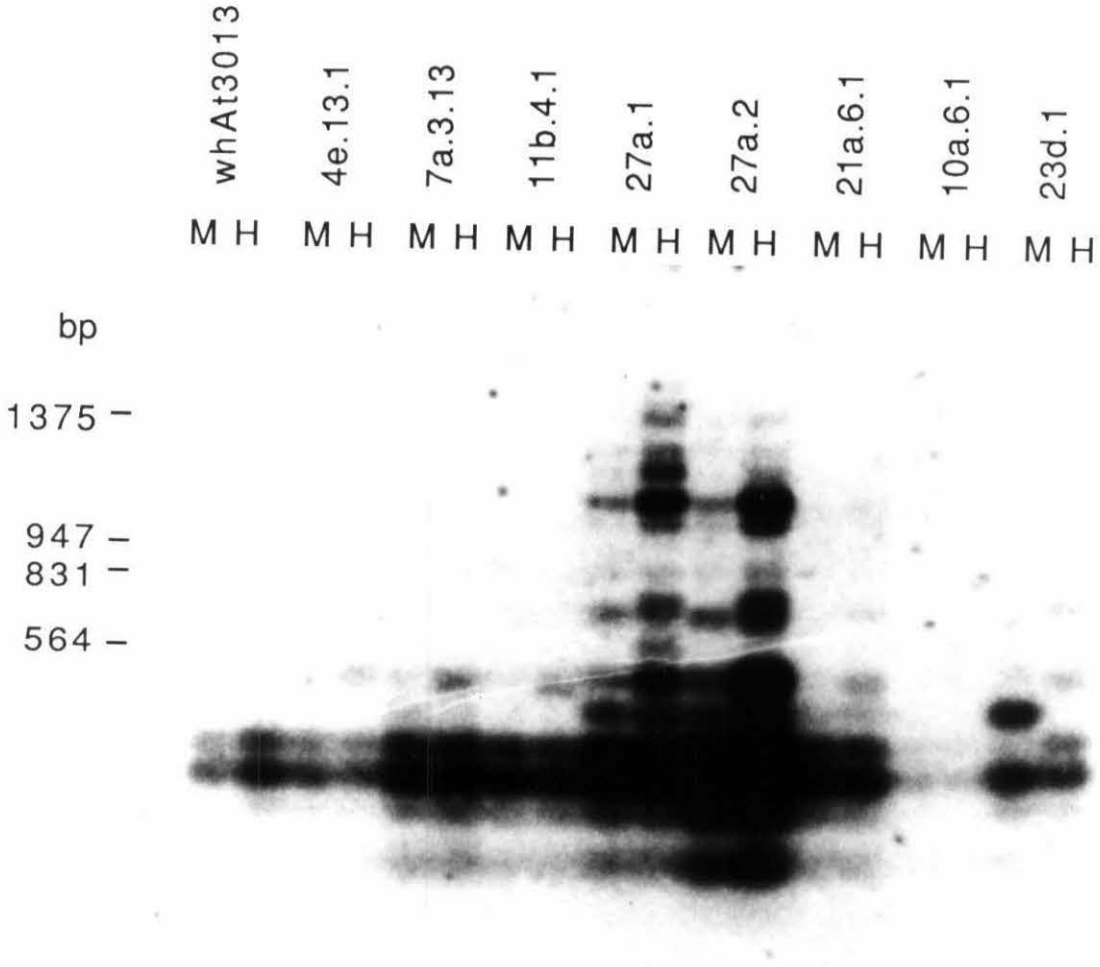

whAt3013
4e.13.1
7a.3.13
11b.4.1
27a.1
27a.2
21a.6.1
10a.6.1
23d.1
M H M H M H M H M H M H M H M H M H
bp
1375
947
831
564

Figure 10. Evidence for a tandem copy and a junction in the 3013 fragment in plants 21a.6.1 (lane a) and 21a.6.2 (lane b). A *Bam*HI blot is shown hybridized separately with probes 11, 8, and 1 (see Figure 6a). The ~22 kb fragment contains the endogenous ADH gene and the 9.8 kb fragment contains the introduced gene. A tandem copy is evident from the presence of the intact 6.7 kb vector fragment detected with probe 1 (also with probe 8 contaminated with probe 1). Of the three probes, only probe 11 detects the 7.8 kb fragment which is presumed to be a junction fragment.

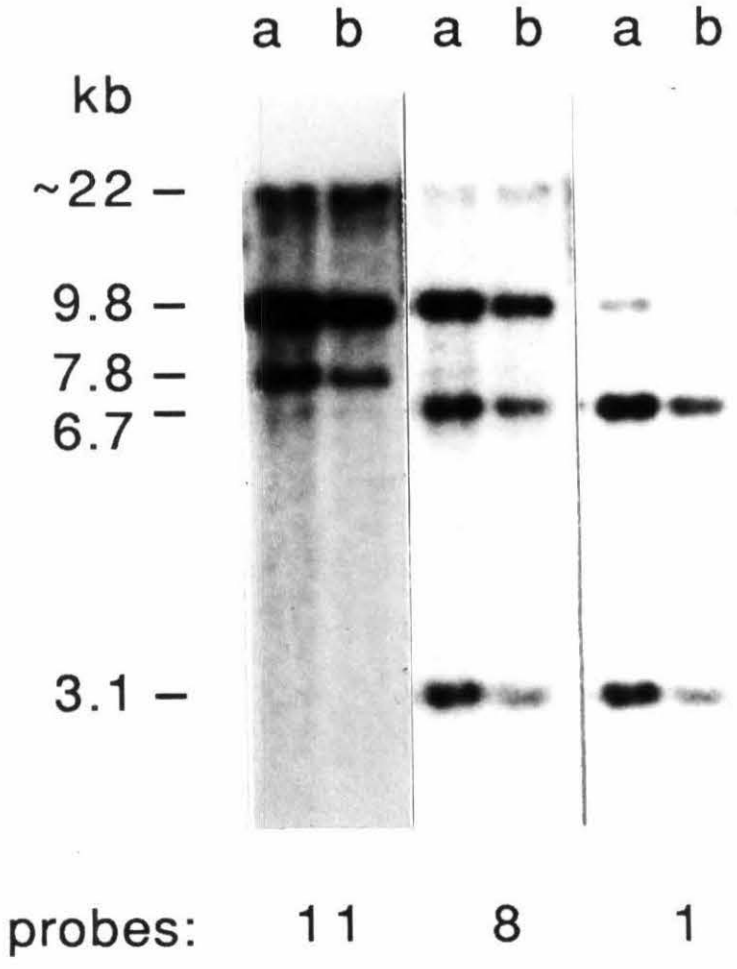

a b a b a b
kb
~22 –
9.8 –
7.8 –
6.7 –
3.1 –
probes: 11 8 1

Figure 11. Evidence for a DNA rearrangement and a junction in the 3013 fragment in plants 23d.1 and 23d.2 (both selected with allyl alcohol). For comparison, DNA of lines 7a and 3a are shown for the *Bam*HI genome blot, and DNA of plant 11b.1 (selected with allyl alcohol) is shown for the *Bam*II genome blot. In the *Bam*HI genome blot, the intact 9.8 kb fragment containing the transforming ADH gene is absent in line 23d, but is present in the other two lines. In the *Bgl*II genome blot, several fragments (including a small, unlabeled endogenous fragment) are detected with probe 11, but only weakly detected with probe 8.

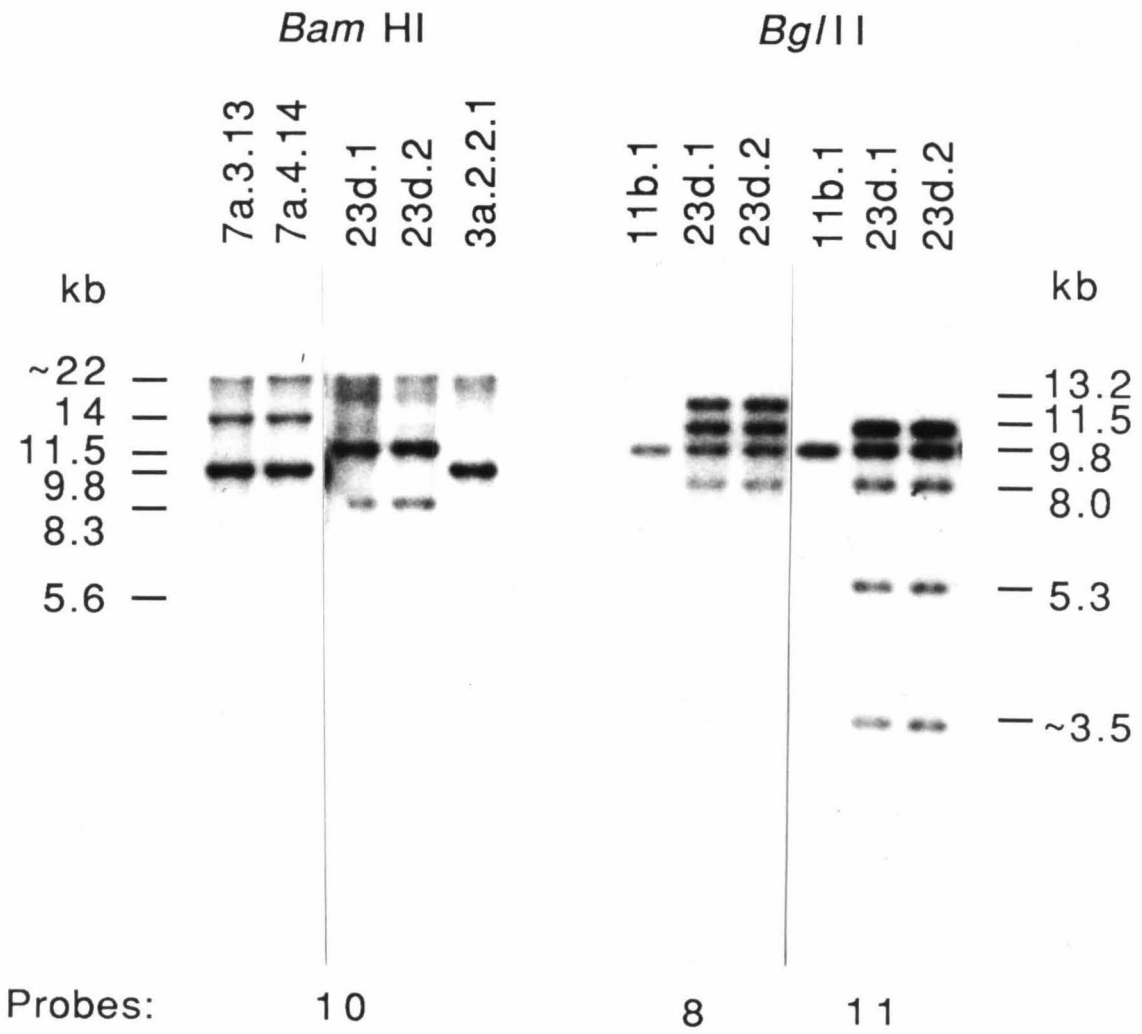
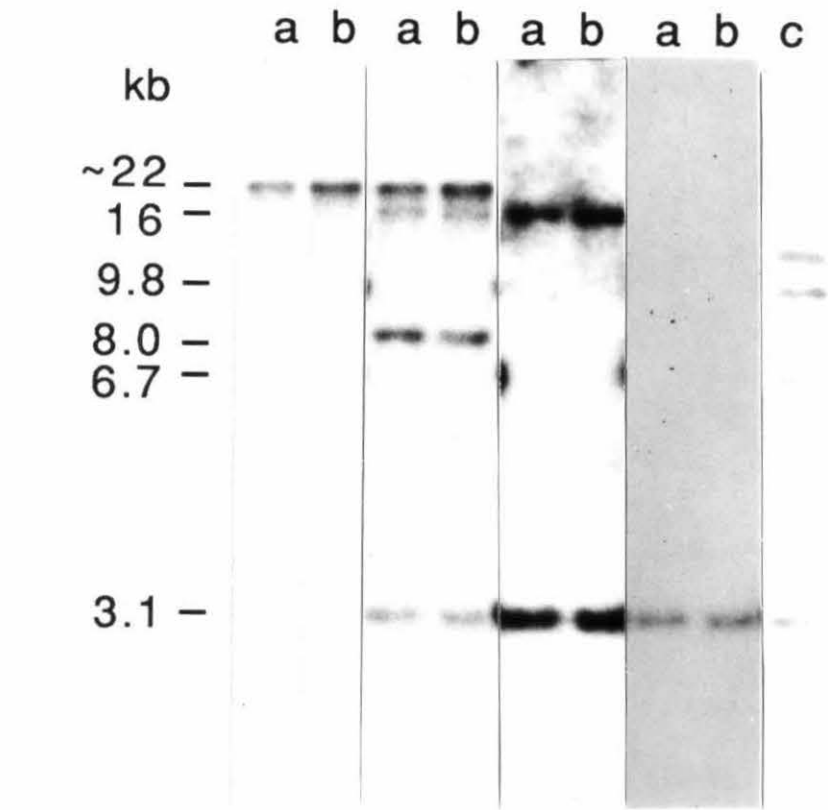

Bam HI
Bgl II
7a.3.13
7a.4.14
23d.1
23d.2
3a.2.2.1
11b.1
23d.1
23d.2
11b.1
23d.1
23d.2
kb
~22
14
11.5
9.8
8.3
5.6
kb
13.2
11.5
9.8
8.0
5.3
~3.5
Probes:
10
8
11

Figure 12. Evidence for a junction fragment in the 3013 fragment in plants 10a.6.1 (lane a) and 10a.6.2 (lane b). A *Bam*HI genome blot is shown hybridized with probe 8 (contaminated with probe 1) and flanking probes. For size standards, lane c contains whAt3013 digested with *Bam*HI and pMON410.1 (wh) digested with *Eco*RI. The expected 9.8 kb fragment containing the ADH gene is absent.



Probes: 10 8 1 4 8

CHAPTER THREE

An RFLP Linkage Map to Facilitate Gene Cloning in *Arabidopsis thaliana*

(Submitted for publication)

An RFLP Linkage Map to Facilitate Gene Cloning in *Arabidopsis thaliana*

Caren Chang, John L. Bowman, Arthur W. DeJohn *, Eric S. Lander †,
and Elliot M. Meyerowitz

Division of Biology 156-29
California Institute of Technology
Pasadena, California 91125

* Current address:
Department of Geology
Humboldt State University
Arcata, California 95521

† Whitehead Institute for Biomedical Research
9 Cambridge Center
Cambridge, Massachusetts 02142
and
Harvard University
Cambridge, Massachusetts 02138

ABSTRACT

We have constructed a restriction fragment length polymorphism linkage map for the nuclear genome of the flowering plant *Arabidopsis thaliana*. The map contains 90 randomly distributed molecular markers and is physically very dense; over 50% of the genome is within 2 centimorgans, or 270 kilobase pairs, of the mapped DNA fragments. The map was based on the meiotic segregation of markers in two different crosses. The RFLP linkage groups were integrated with the five classically mapped linkage groups by virtue of mapped mutations included in these crosses. Markers consist of both cloned *Arabidopsis* genes and random low copy number genomic DNA clones which are able to detect polymorphisms with the restriction enzymes *Eco*RI, *Bgl*II, and/or *Xba*I. These cloned markers can serve as starting points for chromosome walking, allowing for the isolation of *Arabidopsis* genes of known map location. The restriction fragment length polymorphism map also can associate clones of unknown gene function with mutant phenotypes, and vice versa.

INTRODUCTION

Cloning of genes known only by mutant phenotype and genetic map position can be accomplished by successively isolating overlapping clones from a DNA library until the desired locus is reached (1, 2). This approach, known as chromosome walking, is facilitated by having starting clones in close proximity to the endpoints of interest. In order to provide such starting points in the flowering plant *Arabidopsis thaliana*, we have constructed a genetic linkage map of the *Arabidopsis* nuclear genome consisting of 90 molecular markers that are distinguished on the basis of restriction fragment length polymorphisms (RFLPs). RFLPs are codominant, typically neutral, genetic markers. They are provided in abundance by differences between allelic DNA sequences that result from nucleotide substitution, rearrangement, insertion or deletion. Accordingly, they can be used to construct detailed genetic linkage maps, follow inheritance of genetic diseases and other heritable traits and examine variation between and within populations (3, 4, 5). This RFLP linkage map consequently enhances the utility of *Arabidopsis* for experiments in plant molecular genetics.

Previously recognized attributes of *Arabidopsis* include an extremely small genome, rapid generation time, fecundity, ease of growth, and an extensive background of classical genetic analysis (6). Chromosome walking in *Arabidopsis* should be facilitated by the small genome and virtual absence of dispersed repetitive DNA sequences (an average spacing of 125 kb between repeats as compared to 1.4 kb in tobacco and 0.3 kb in pea (6)). The end of a chromosomal walk in *Arabidopsis* can be determined by complementation of the mutant phenotype via transformation (7). Complementation has been demonstrated for an *Arabidopsis* ADH null mutant transformed with the *Arabidopsis* ADH gene (C. Chang and E.M. Meyerowitz, unpublished work). Numerous mutations affecting a range of developmental processes, various enzyme activities, and hormone synthesis and response have been isolated in *Arabidopsis*, and approximately 80 of these mutations have been

ordered into a genetic linkage map (8, 9). The RFLP map is consistent with this map, and the two have been partly integrated.

Since the *Arabidopsis* genome consists of about 500 centimorgans (cM) and about 70,000 kilobases (kb) (10), the 90 markers now mapped are at an average genetic spacing of about 5.6 cM and an average physical spacing of 780 kb.

MATERIALS AND METHODS

Cloned *Arabidopsis* Genes. Cloned *Arabidopsis* genes used as RFLP markers were kindly provided as follows: chalcone synthase (λCHS2) from R. Feinbaum and F. Ausubel, nitrate reductase (λAt–24) (11) from N. Crawford, actin (pAtc4) from R. Ferl, phytochrome (λPh.Ara.1) from R. Sharrock, C. Gatz, and P. Quail, acetolactate synthase (pGH1) (12) from G. Haughn and C. Somerville. Cloned genes from our laboratory were: λfAt3012 (alcohol dehydrogenase) (13), sAt2105 (12S seed storage protein) and nAt1511 (small seed RNA-coding) (P. Pang, R. Pruitt and E. Meyerowitz, in preparation). The remaining mapped fragments were chosen from a genomic DNA library.

***Arabidopsis* Strains.** The ecotypes Niederzenz (Nd-0) (14), Landsberg (La-0) (15), and Columbia (Col-0) (14), and the mutations, *an* (16), *gl-1*(16), *tz* (17), *er* (18), *ap-2* (19), and *clv-1* (7) have been previously described. To perform crosses, flowers of the female recipient were hand-emasculated and anthers from the donor were applied to the stigmatic surface of the recipient.

Library Screening. The *Arabidopsis* (Col-0) genomic lambda library was screened using standard techniques (20).

Isolation of DNA. Rapid lambda DNA preparations followed the procedure described in reference 21, except that the clear lysate was incubated with 1 μl DNase (1 mg/ml) and 1 μl RNAse (1 mg/ml) for 20 minutes at 37°C, followed by incubation with 5 ul diethylpyrocarbonate for 5 minutes at room temperature, prior to treatment with $NaDodSO_4$.

Whole plant DNA was extracted from 3-6 week old plants by either of two methods described in ref. 22 and ref. 23. The procedure in ref. 23 was modified as follows. Stirring of the lysis mixture and the first phenol-chloroform extraction were omitted. Instead, after adding the $NaDodSO_4$, the mixture was vortexed and incubated at 65°C for 10 minutes. Two extractions were performed using a 2:1 phenol-chloroform ratio. The

DNA was precipitated in 1M KOAc and one volume of isopropanol at room temperature. The DNA was centrifuged (2.55 x 10^5 g for 20 hours) in a CsCl gradient and the CsCl was removed by ethanol precipitation of the DNA.

Genome Blot Hybridizations. 0.5-2.0 μg of whole plant DNA was digested with 10-20 units of restriction enzyme and 1 mM spermidine (Sigma) for 2-4 hours and then subjected to gel electrophoresis in 0.8% agarose gels at 20-40 volts until the Orange G (Sigma) loading dye migrated approximately 13 cm. The digested DNA was loaded into 5 x 1.5 x 5 mm wells. Gels were blotted and the DNA cross-linked to Hybond-N filters (Amersham) according to the manufacturer's instructions. ^{32}P-labeled DNA probes (2-3 x 10^8 cpm/μg) were produced by nick-translation (24) of lambda clone DNA (21). Filters were pre-hybridized for 1-10 hours and hybridized (10^6-10^7 cpm/ml) for 18-48 hours at 65°C in 5x SSPE (900 mM NaCl, 50 mM $NaH_2PO_4 \cdot H_2O$, 40 mM NaOH, 5 mM Na_2EDTA), 5x Denhardt's solution, 0.5% $NaDodSO_4$, and 20 μg/ml denatured salmon sperm DNA. Filters were washed at 65°C in 2x SSPE for 30 minutes, 2x SSPE with 0.1% $NaDodSO_4$ for 30 minutes, and 0.1x SSPE for 15 minutes. Pre-flashed Kodak XAR-5 film was exposed to filters for 1-7 days at -70°C with an intensifying screen. Filters were re-used up to 30 times after removing previous probe by submerging filters in boiling 0.1% $NaDodSO_4$.

Linkage Analysis. Multi-point linkage analysis was performed using the MAPMAKER computer program (25, 26). Initially two-point linkage analysis was performed to determine the maximum likelihood recombination fraction and the associated LOD score for each pair of loci. The LOD score reflects the degree of support for linkage (27), and is defined as the log10 of the ratio of the probability that the data would arise if the loci are actually linked at a given recombination fraction divided by the probability that the data would arise if the loci are unlinked. Pairs of loci were considered linked if the LOD score exceeded 3.0. Once the linkage groups were assembled, three point and subsequently n-point linkage analysis was performed on each linkage group.

RESULTS

Construction of the *Arabidopsis* RFLP map involved three steps: (i) collecting and identifying potential RFLP markers; (ii) performing the genetic crosses; (iii) determining the linear order of RFLP loci and the map distances between them by examining segregation of the RFLPs.

Obtaining RFLP Markers. Cloned inserts for the first step should be large and have a low copy number in the genome in order to increase the probability of detecting polymorphisms while at the same time avoiding RFLPs at multiple loci. Both these criteria are met by large genomic clones of *Arabidopsis* due to the low amount of dispersed repetitive sequences. Our primary source of RFLP clone candidates was a collection of random low copy number clones that were obtained by screening (20) an *Arabidopsis* total genome lambda library (22) (estimated average insert size of 12.5 kb). The probe was total *Arabidopsis* DNA, and an internal lambda vector DNA sequence (to detect non-recombinant phage), that had been ^{32}P-labeled by nick-translation (24). In agreement with the proportion predicted by genome studies (10, 22), 60% of the screened recombinant clones gave either a weak or undetectable signal, presumably because they represent sequences of low copy number in the nuclear genome. These clones were isolated and rescreened by either plaque screen or clone DNA blot (20), resulting in the isolation of 280 low copy number clones. We also used eighteen single copy *Arabidopsis* clones (previously isolated from the same DNA library (22)), and eight cloned *Arabidopsis* genes (see Materials and Methods).

To screen plant DNA for the presence of RFLPs, the clones were used as ^{32}P-labeled probes in hybridization to gel blots containing genomic DNA of three different *Arabidopsis* strains, digested separately with each of three different restriction enzymes (Fig. 1). The strains used were: (i) Niederzenz (Nd-0); (ii) a triply marked line (which we designate as C) in a background that is largely Columbia (Col-0) carrying the recessive visible mutations

angustifolia (*an*), *glabra-1* (*gl-1*), and *thiazole requiring* (*tz*); and (iii) Landsberg (La-0) carrying the recessive visible mutation *erecta* (*er*). (Nearly all the available and mapped mutations in *Arabidopsis* have been induced in Landsberg *erecta* (8).) The restriction enzymes used were *Eco*RI, *Xba*I, and *Bgl*II. (*Hin*dIII, *Sac*I, and *Xho*I were included initially, but appeared to yield a lower frequency of RFLPs.) We screened all three lines with 202 clones for RFLPs that are revealed by at least one of the three enzymes; Nd-0 and C were screened with 23 additional clones, and C and La-0 *er* were screened with 3 additional clones. *Eco*RI, *Xba*I, and *Bgl*II were found to be equally useful for RFLP detection. The overall frequency of clones that detected RFLPs with at least one of the three enzymes for at least two pairs of lines was 54%; the frequency was 37% between Nd-0 and C, 46% between Nd-0 and La-0 *er* and 39% between C and La-0 *er*. The frequency of the clones detecting fragments of different lengths in all three strains was 15%. Only those RFLPs which appear to be distinctly resolved codominant alleles have been included in these figures. Not counted are RFLPs of the dominant/null allele type in which there are new (or absent) fragment(s) in one of the lines with no apparent alterations in any of the other fragment lengths. These polymorphisms probably did not result from sequence duplications since polymorphisms were not revealed by all three enzymes. Most likely, they are caused by the presence of a common restriction site within, and a polymorphic restriction site outside of, the probe sequence if in one line the site is far enough to produce an undetected restriction fragment that is larger than the mean size of the isolated DNA. Such polymorphisms, accounting for 12% of all RFLPs, were not mapped. A survey of all RFLPs indicated that Nd-0 and La-0 *er* differ in approximately 1.4% of the nucleotides in their low copy genomic DNA, C and Nd-0 in approximately 1.3% in their low copy genomic DNA, and C and La-0 *er* in approximately 1.1% of their low copy genomic DNA. In this calculation we assumed that: (i) there is one nucleotide substitution per RFLP, (ii) there are four restriction sites detected by a probe per enzyme, and (iii) all sites are non-

overlapping. The above results suggest that a cross between any two of the three strains is suitable for RFLP mapping.

Genetic Crosses. Our RFLP map is based on two crosses, each including a set of visible mutations for the purpose of positioning RFLP linkage groups relative to the five linkage groups established by mapping morphological and biochemical mutations. One of the two crosses was between Nd-0 and the triple mutant line C marked on chromosomes 1, 3, and 5. The other cross was between Nd-0 and La-0 marked with the recessive visible mutations *clavata-1* (*clv-1*), *er*, and *apetala-2* (*ap-2*) on chromosomes 1, 2, and 4, respectively. For each cross, the resulting F_1 plants were allowed to self-pollinate to produce F_2 individuals. F_2 plants were allowed to self-pollinate to each produce a pool of F_3 plants. Although segregation of RFLPs is present in F_2 individuals, the F_3 plants enabled us to score the genotypes of F_2 plants for the recessive visible markers. DNA was prepared from pools of 15 or more F_3 plants per F_2 individual, providing a greater amount of tissue, and thus DNA, than would have been provided by a single F_2 plant.

Segregation of RFLPs. To examine the segregation of RFLPs, DNA probes were hybridized to genome blots containing digested F_3 DNAs of both crosses. An example for two probes is shown in Fig. 2. For each of the enzymes *Eco*RI, *Xba*I, and *Bgl*II, a set of genome blots was prepared using the F_3 DNAs of each cross. RFLP probes were hybridized sequentially to filters of the appropriate digest and cross. Each F_3 pool was scored with respect to the parental DNA restriction patterns as being homozygous for one of the parental types, heterozygous, or unscorable for technical reasons. Using the Nd-0 x C cross alone, for which 106 different F_3 pools were scored, 41 RFLPs were mapped. Using the Nd-0 x La-0 cross alone, for which 119 different F_3 pools were scored, 25 RFLPS were mapped. In order to align linkage groups of clones mapped in the separate crosses, 24 additional markers were mapped in both crosses. The resolution afforded by 225 F_3 pools is 0.22 cM, i.e. one recombinant divided by 450 chromatids, which represents on average 30 kb in the *Arabidopsis* genome.

All but two of the clones appeared to detect single loci. Clone nAt1511 (which contains a gene encoding an abundant seed-specific mRNA) hybridized to two pairs of segregating RFLP alleles; one RFLP was within the cloned copy and the other was within an uncloned homologous sequence whose existence was indicated by a comparison between genome blots and the restriction map of the clone. We mapped both RFLPs and found them to be closely linked. Probe 281 hybridized to two polymorphic fragments that mapped to different chromosomes. Each probe may hybridize to nonadjacent members of a gene family, or may contain two unrelated sequences which were ligated together during the library construction. When such markers are used to initiate a chromosome walk, a clone specific to the RFLP locus of interest must be obtained.

Linkage Analysis. The genetic map was constructed using the MAPMAKER computer program (25, 26) to analyze the combined data from the two crosses. Two-point linkage analysis was first performed to determine the maximum likelihood recombination fraction and the associated LOD score (27) for each of the 4560 pairs of the 96 loci. Pairs of loci were considered linked when the LOD score exceeded 3.0. (The traditional threshold for declaring linkage in human genetics is a LOD score of 3.0. Since the *Arabidopsis* genome is roughly seven times smaller in genetic length than the human genome, the use of a LOD score threshold of 3.0 in this study is quite conservative.) The 96 loci fell into five linkage groups, which were assigned to the five chromosomes by virtue of the visible markers with previously assigned genetic positions.

Three-point linkage analysis was next used to narrow down possible genetic orders for the loci (a-b-c, a-c-b, b-a-c). For each order, we computed the likelihood that the data would have arisen given the maximum likelihood three-point map for that order. A particular three-point order was ruled out if some alternative order for the loci was at least 100,000 times more likely to have given rise to the data. For each linkage group, we determined all genetic orders which contained no sub-orders that had been ruled out by the

three-point analysis. Typically, some 20-50 genetic orders for the linkage group met this criterion.

Finally, full n-point linkage analysis was performed for each remaining genetic order to compute the most likely recombination fractions and the chance that the map would have given rise to the data. Genetic orders were discarded which were at least 1,000-fold less likely to give rise to the data.

A single genetic order emerged for each chromosome, which was unique up to the possible uncertainty in the order of certain pairs of loci separated by less than 3 cM. The final map, shown in Figure 3, displays distances in Kosambi (28) centimorgans and indicates any order that is supported by odds of less than 1000:1. As a confirmation, the entire process was repeated using just the data for the Nd-0 x C cross alone and the Nd-0 x La-0 cross alone; the resulting genetic orders and maps were consistent with the map shown in Figure 3.

The orientation of RFLP chromosomes 1 and 5 with respect to the standard map has been determined using two visible markers on each chromosome: *an* and *clv-1* on chromosome 1, and a recessive visible mutation *pistillata* from a third cross (Bowman, J. L. & Meyerowitz, E. M., work in progress) and *tz* on chromosome 5. Chromosomes 3 and 4 are shown in the orientation most likely to agree with the standard map based on the map position of an acetolactate synthase (ALS) mutation, *chlorsulfuron resistance*, on chromosome 3 (29, G.W. Haughn and C. Somerville, unpublished work) for which we have mapped the ALS gene clone, and on the distal position of *ap-2* on chromosome 4. Further crosses are in progress to determine unequivocal orientation for chromosomes 2, 3, and 4 (J.L. Bowman and E.M. Meyerowitz, work in progress).

DISCUSSION

We have constructed an RFLP linkage map of 90 molecular markers that can be integrated with the standard linkage map of approximately 80 mutation markers. The RFLP map contains linkage groups of 144, 80, 93, 62 and 121 cM for a total of 501 cM, which corresponds well with the standard genetic map showing linkage groups of 126, 51, 91, 69 and 100 cM for a total of 437 cM (9). The positions of the visible markers that are common to both maps are also in essential agreement; on the standard mutation map, *an* and *clv-1* are on chromosome 1 at positions 0 and 114.6 cM, respectively, *er* is on 2 at 15.9, *gl-1* is on 3 at 40.9, *ap-2* is on 4 at 58.6, and *tz* is on 5 at 82.4. The RFLP map is also consistent with the cytological karyotype (30, 31). The positions of centromeres in the RFLP map might be determined by the use of existing telotrisomic lines (8). One reason this is of interest is that the C-banded heterochromatin associated with the centromeres is likely to present a barrier to chromosome walks across centromeres.

RFLP mapping of cloned genes (which have been obtained by other means) may associate the genes with mapped loci that are responsible for mutant phenotypes. For example, the RFLP map position of the chalcone synthase gene clone suggests a correlation between this gene and the pigment mutation *transparent testa-4* (9); and the map position of the nitrate reductase gene clone suggests a correlation between this gene and the mutation *chlorate resistance-3* (9) in which nitrate reductase activity is reduced. Thus the combined maps have the potential to provide information on clones having unknown gene function, or to associate mutant phenotypes with gene products.

The RFLP map should facilitate cloning of genes known only by mutant phenotype by providing starting points for chromosome walking. The labor involved in chromosome walking depends upon the physical distance between the starting clone and the desired gene. While an actual correlation of physical distance and genetic distance has yet to be established for specific regions of the genome, a general measure is 140 kb per cM (the

genome size in kilobases divided by the RFLP map genome size in centimorgans). In terms of averages, the mean distance between RFLP markers is 780 kb, and therefore, the mean walking distance to reach an arbitrary gene is 195 kb. A more descriptive measure of the map's usefulness than the average spacing between markers is to examine the percent of the genome (the genome being defined by the RFLP map) within a given distance from any RFLP marker (Fig. 4). This calculation shows that over half of the genome lies within only 270 kb of an RFLP marker, and if one is willing to walk about three times that distance, then one has access to 90% of the genome. The required walking distances can be reduced by increasing the density of RFLP markers on the map, a job which is underway (J.L. Bowman, S. Kempin, and E.M. Meyerowitz, work in progress). This is not difficult since, in contrast with nonmolecular markers, the number of RFLP markers that are segregated in a single cross is essentially unlimited, i.e. restricted only by the number of RFLP probes that are identified. The distances between consecutive loci closely fit the negative exponential distribution that would be expected from randomly picking points along linear chromosomes, which indicates that current gaps in the map are likely to be filled. Since all 96 loci in this study are linked in the RFLP map, it is expected that the vast majority of the *Arabidopsis* genome is linked to the map. Improved methods of walking and jumping (32, 33) will shorten the time necessary to walk larger distances. Furthermore, a library of ordered cosmid clones for the entire *Arabidopsis* genome is being constructed (34). (The mapped RFLP markers themselves represent 1.6% of the *Arabidopsis* genome, assuming 12.5 kb of genomic DNA per marker.) Chromosome walks (or an ordered library) will provide information on the true physical distances that are represented by a centimorgan in different genomic regions.

With the Ti plasmid-mediated transformation procedure presently in use, clones along a walk can be assayed for the ability to complement mutant phenotypes (7). In conclusion, the outlook for cloning *Arabidopsis* genes defined only by mutant phenotype now seems promising.

REFERENCES

1. Bender, W., Spierer, P., and Hogness, D.S. (1983). *J. Mol. Biol.* 168, 17-33.

2. M. Steinmetz *et al.*, (1982). *Nature* 300, 35-42.

3. Botstein, D., White, R., Skolnick, M., and Davis, R. W. (1980). *Am. J. Hum. Genet.* 32, 314-331.

4. Helentjaris, T. (1987). *Trends in Genet.* 3, 217-221.

5. Burr, B., Evola, S.V., Burr, F.A., and Beckman, J.S. (1983). *Genetic Engineering; Principles and Methods*, Vol. 5 (Setlow, J.K. and Hollaender, A., eds.), Plenum Press, New York, pp. 45-59.

6. Meyerowitz, E.M. and Pruitt, R.E. (1985). *Science* 299, 1214-1218.

7. Lloyd, A.M., Barnason, A.R., Rogers, S.G., Byrne, M.C., Fraley, R.T., and Horsch, R.B. (1986). *Science* 234, 464-466.

8. Koornneef, M., van Eden, J., Hanhart, C.J., Stam, P., Braaksma, F.J., and Feenstra, W.J. (1983). *J. Hered.* 74, 265-272.

9. Koornneef, M. (1987). *Genetic maps 1987: A compilation of linkage and restriction maps of genetically studied organisms* (O'Brian, S.J., ed.), Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, pp. 742-745.

10. Leutwiler, L.S., Hough-Evans, B.R., and Meyerowitz, E.M. (1984). *Mol. Gen. Genet.* 194, 15-23.

11. Crawford, N., Campbell, W.H., and Davis, R. (1986). *Proc. Natl. Acad. Sci. U.S.A.* 83, 8073-8077.

12. Haughn, G.W., Smith, J., Mazur, B., and Somerville, C. (1988). *Mol. Gen. Genet.* 211, 266-271.

13. Chang, C. and Meyerowitz, E.M. (1986). *Proc. Natl. Acad. Sci. U. S. A.* 83, 1408-1412.

14. Kranz, A.R. (1978) *Arabidopsis Information Service* 15, 118-139.

15. Robbelen, G. (1965). *Arabidopsis Information Service.* 2, 36-47.

16. Koornneef, M., Dellaert, L.W.M., and van der Veen, J.H. (1982). *Mut. Res.* 93, 109-123.

17. Rédei, G.P. (1962). *Genetics* 47, 979.

18. Rédei, G.P. and Hirono, Y. (1964). *Arabidopsis Information Service* 1, 9-10.

19. Koornneef, M., De Bruine, J.H., and Goettsch, P. (1982). *Arabidopsis Information Service* 17, 11-18.

20. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).

21. Davis, R.W., Botstein, D. and Roth, J.R. (1980). *Advanced Bacterial Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).

22. Pruitt, R.E. and Meyerowitz, E.M. (1986) *J. Mol. Biol.* 187, 169-183.

23. Shure, M., Wessler, S. and Fedoroff, N. (1983) *Cell* 35, 225-233.

24. Rigby, P.W.J. , Dieckmann, M., Rhoades, C., and Berg, P. (1977). *J. Mol. Biol.* 113, 237-251.

25. Lander, E.S., Green, P., Abrahamson, J., Barlow, A., Daly, M., Lincoln, S. and Newberg, L. (1987). *Genomics* 1, 174-181.

26. Lander , E.S., and Green, P. (1987). *Proc. Natl. Acad. Sci. U.S.A.* 84, 2363-2367.

27. Ott, J. (1985). *Analysis of Human Genetic Linkage* (John Hopkins University Press, Baltimore).

28. Kosambi, D.D. (1944). *Ann. Eugen.* 12, 172-175.

29. Haughn, G.W. and Somerville, C. (1986). *Mol. Gen. Genet.* 204, 430-434.

30. Schweizer, D., Ambros, P., Gründler, P., and Varga, F. (1987). *Arabidopsis Information Service* 25, 27-34.

31. Sears, L.M.S. and Lee-Chen, S. (1970). *Can. J. Genet. Cytol.* 12, 217-223.

32. Collins, F.S., Drumm, M.L., Cole, J.L., Lockwood, W.K., VandeWoude, G.F., and Iannuzzi, M.C. (1987). *Science* 235, 1046-1049.

33. Poustka, A., Pohl, T.M., Barlow, D.P., Frischauf, A.-M., and Lehrach, H. (1987). *Nature* 325, 353-355.

34. Hauge, B., Fritze, C., Nam, H.-G., Paek, K.-H. and Goodman, H.M. (1987). Abstract # 32, *Third International Meeting on Arabidopsis*, Michigan State University.

Figure 1. Autoradiogram of a typical genomic blot to screen for RFLPs between different parental lines. The figure shows several RFLPs detected by clone 305 hybridized to Nd-0 (N), C (C), and La-0 *er* (L) DNA digested with each of the enzymes *Eco*RI, *Xba*I, and *Bgl*II.

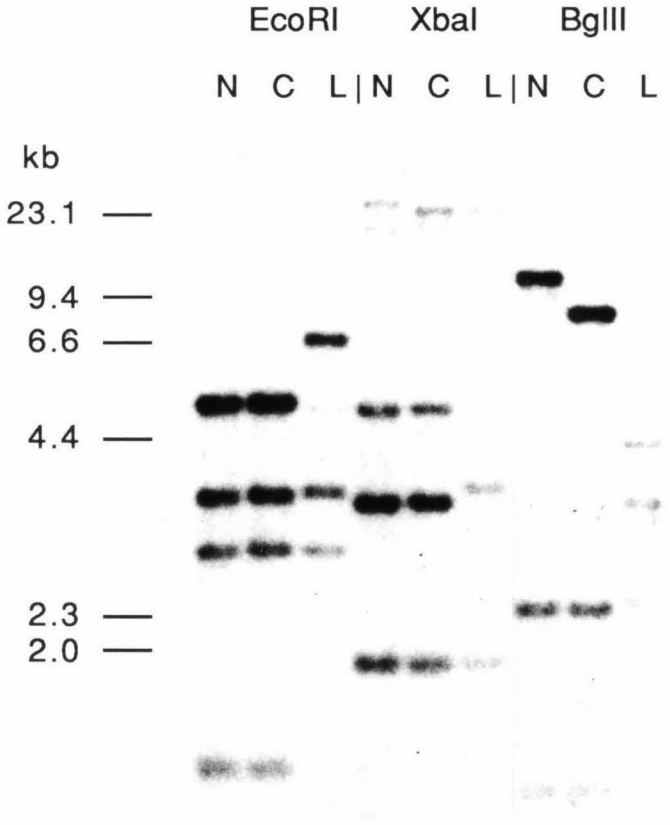

EcoRI
XbaI
BglII
N C L | N C L | N C L
kb
23.1
9.4
6.6
4.4
2.3
2.0

Figure 2. Segregation of RFLPs in pooled F_3 progeny is shown in autoradiograms of the same genome blot hybridized sequentially with two unlinked RFLP probes. The blot contains genomic DNA samples digested with *Eco*RI. In (**A**) the blot is probed with clone 322, and in (**B**) with clone 214. The first and second lanes contain DNA of the parental lines Nd-0 (N) and C (C) respectively. The 18 lanes to the right contain DNA of pooled F_3 progeny derived from the cross between Nd-0 and C, scored as either N, C or H (heterozygous).

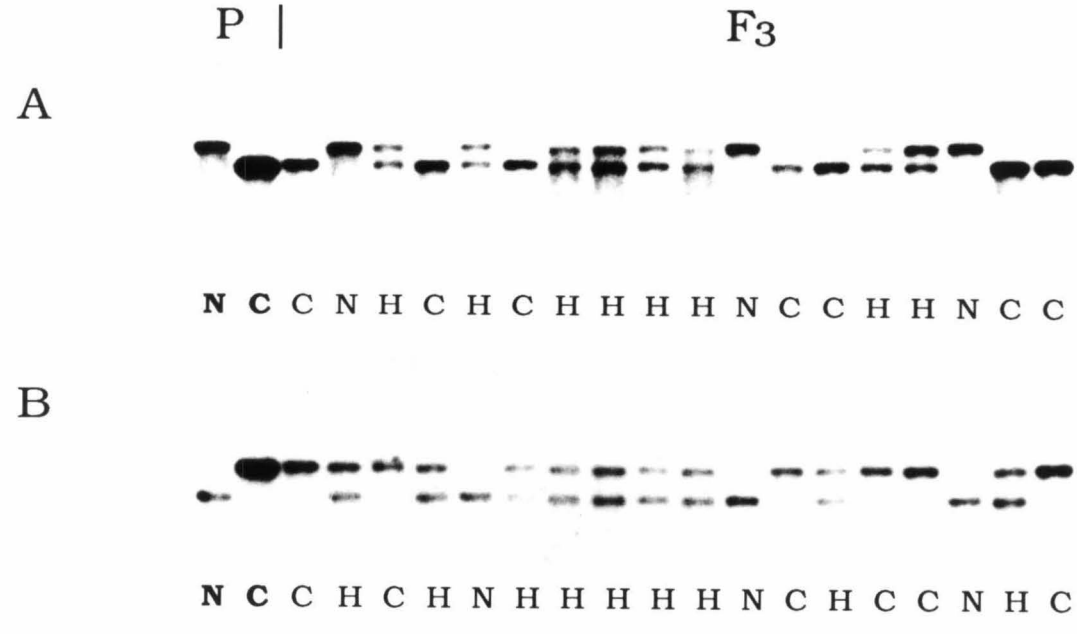

P
F3
A
N C C N H C H C H H H H N C C H H N C C
B
N C C H C H N H H H H H N C H C C N H C

Figure 3. An RFLP linkage map of the five chromosomes of the *Arabidopsis thaliana* genome. Map distances are shown in centimorgans. (We arbitrarily assigned a position of zero to the top-most marker on each chromosome.) Markers are designated by a clone number for the random low copy clones, and by a clone name for the known genes. An *a* or *b* following the marker name indicates the particular RFLP alleles of the clone used to map the locus. Markers scored in both crosses are indicated by (▭). Markers scored only in the cross between Nd-0 and La-0 are indicated by (▬). The remaining markers were scored only in the cross between Nd-0 and C. Markers placed in an order which is less than 1000 times more likely to have given rise to the data than any alternative order are bracketed to the right of the map distances; the numbers by the brackets are the $\log_{10}$ of likelihood ratio of the order shown to the next most likely alternative order. In three separate cases (115, 254, and 506), the order of markers could not be resolved, i.e., the likelihood for any of the possible intervals was virtually the same; the approximate positions of these markers are indicated to the left of the other markers.

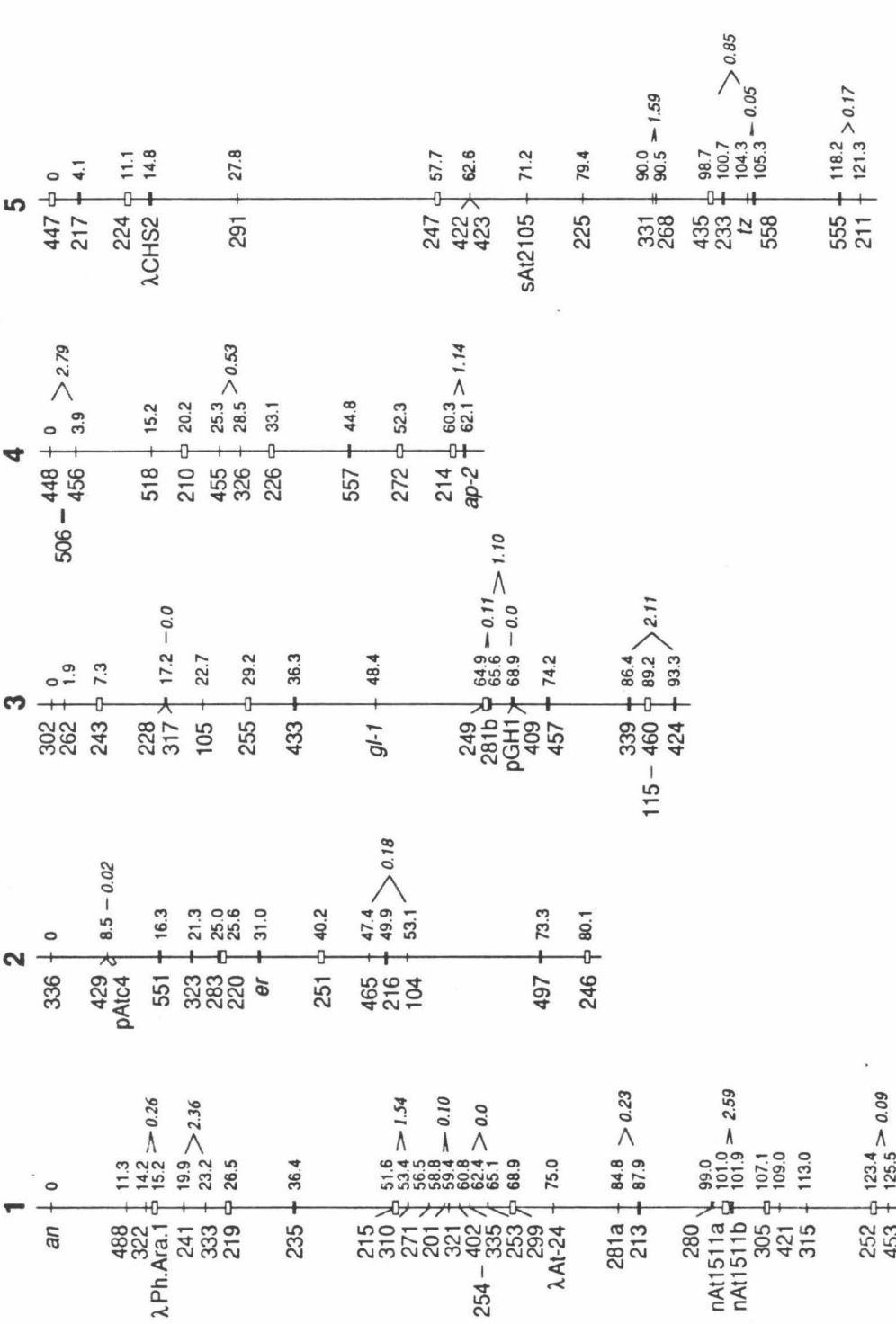
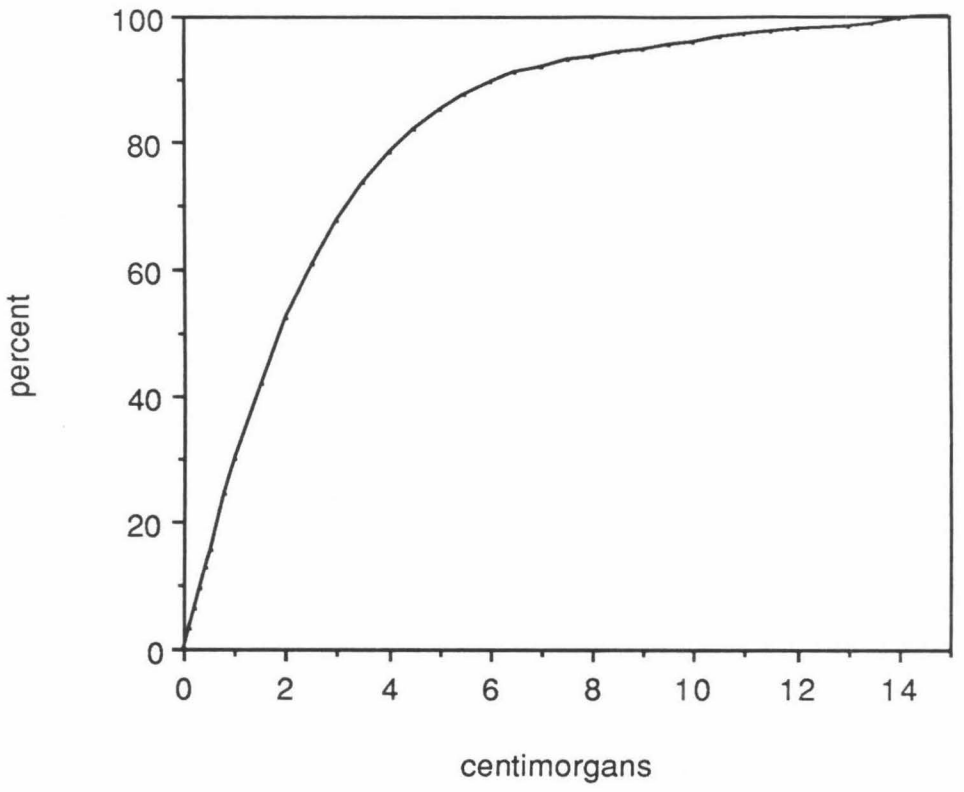

1
2
3
4
5
an
λPh.Ara.1
λAt-24
nAt1511a
nAt1511b
clv-1
jAt3012
pAtc4
er
gl-1
pGH1
ap-2
λCHS2
sAt2105
tz

Figure 4. A graph depicting the completeness of the RFLP linkage map as measured by the percent of the genome within a given map distance (cM) of an RFLP marker. To avoid assuming lengths for the ends of the chromosomes, the genome is considered here to be the sum of the RFLP chromosome lengths, where the two ends of each chromosome are defined by the outermost RFLP markers.


100
80
60
40
20
0
percent
0
2
4
6
8
10
12
14
centimorgans

CONCLUSION

This thesis examines several genetic aspects within the developing area of *Arabidopsis* molecular genetics as part of a larger effort to learn about the molecular, physiological, and biochemical processes of higher plants. Using *Arabidopsis* as a model system, the work described in this thesis examines and provides approaches for increasing our understanding of gene function in plant growth and development. One approach facilitates the study of gene expression and the other allows for the isolation of a wide range of genes. The latter is related to the unusual size and organization of the *Arabidopsis* genome. In addition, experiments in this thesis add to our knowledge of genetic structure and gene regulation in flowering plants.

The alcohol dehydrogenase (ADH) gene was one of the first genes to be cloned in *Arabidopsis* and the first to be sequenced. Two general principles concerning *Arabidopsis* genes were introduced by the cloning and characterization of this gene. The first is that genes can be cloned from an *Arabidopsis* DNA library using heterologous gene probes. (For ADH, this reflects the conservation between monocots and dicots which are thought to have diverged 150 million years ago.) Thus genes initially isolated in *Arabidopsis* will be useful for obtaining the homologous genes in other plants of interest. Second, *Arabidopsis* genes contain fewer and smaller introns compared with the introns of homologous genes in other experimental plants. In addition, whereas other plants may contain several genes in a gene family, genes in *Arabidopsis* are present in fewer copies. In general, the dozen or so genes recently isolated in *Arabidopsis* (mostly using heterologous probes) have upheld these observations made with the ADH gene. The smaller gene size and the reduced number of genes alone cannot account for the relative difference in genome size between *Arabidopsis* and other higher plants.

DNA sequence analysis showed that the ADH gene contains sequences characteristic of expressed eukaryotic genes in terms of the TATA box, polyadenylation signal, and intron structure. From the deduced protein sequence, *Arabidopsis* ADH is highly conserved with ADH in monocots and mammals, but not in fungi or insects. Part of the 5'

DNA sequence homology between the maize ADH1 and the *Arabidopsis* ADH genes, found in Chapter One, has since been discovered to be a binding site for two proteins in maize upon induction of ADH1 (Ferl, R.J., and Nick, H.S., *J. Biol. Chem.* 262, 7947-7950, 1987). This sequence is likely to be involved in regulation of ADH gene expression in *Arabidopsis*. Tissue- and stage-specific ADH activity was identified using histochemical staining. ADH is constitutively expressed in developing embryos, and is inducible in root tips (as in maize) based on tissue staining and native protein gels. Levels of both ADH message and activity were greater in the Bensheim ecotype than in Landsberg *erecta*.

ADH gene fusions can potentially be used to isolate mutations in trans-acting regulators of gene expression based on counter-selection of ADH activity. Promoters of tissue- or stage- specific genes can be fused to the structural portion of the ADH gene and then the constructs can be transformed into an ADH null background (the present transformation efficiency is sufficient for this approach). After mutagenesis of a transformed line carrying several copies of the introduced DNA, mutants in trans-acting genes can be selected with allyl alcohol. This approach is presently being taken with the seed-specific genes because allyl alcohol selection has been established at the seed level. The fact that seeds with low levels of ADH can be selectively eliminated, as shown in Chapter Two, suggests that mutations that do not completely abolish expression can be isolated.

Agrobacterium-mediated transformation of *Arabidopsis* was shown to be effective for assaying complementation of a mutant allele by a genomic clone carrying the wild-type gene. The work presented here is the first demonstration of this in a higher plant. The transferred DNA is stable and inherited in a Mendelian fashion. However, DNA rearrangements and factors such as genomic location of the introduced DNA were found to affect gene expression. Transformation and complementation of a mutant phenotype show that clones can be assayed for complementation in a chromosome walk, and that large pieces of contiguous DNA are transferred by this procedure. Although the compactness of

the *Arabidopsis* genome is well-suited for shot-gun complementation, the efficiency of transformation is presently much too low to make such an approach practical.

Mapping of restriction fragment length polymorphisms (RFLPs) in *Arabidopsis* has revealed several principles of genetic organization of this plant, and will continue to be valuable for different kinds of analyses. For example, through mapping it was established that random low copy clones in an *Arabidopsis* genomic DNA library are randomly distributed in the genome. This information concerning the organization of the *Arabidopsis* genome was not previously known, and it suggests that low copy sequences in a genomic DNA library are likely to be equally represented (which is important for chromosome walking). It also suggests that there are no large genomic regions for which the relation of physical map distance, measured in kilobase pairs, and meiotic map distances, as measured in centiMorgans, is greatly different from the average. To examine the evolution of plant genomes, linkage of RFLP markers in *Arabidopsis* can be compared with that in other Brassicaceae. RFLPs are also a way to observe DNA sequence variations between different individuals; the three *Arabidopsis* ecotypes in Chapter Three were found to diverge approximately equally from each other. Using RFLP mapping, cloned genes (with or without an identified function) have been associated with previously isolated and mapped mutations as shown for chalcone synthase and nitrate reductase.

The RFLP map is a powerful tool for isolating a wide range of genes. The distinct advantage of chromosome walking is that one can start with only a mutant phenotype that can be mapped relative to the RFLP map. The range of obtainable genes is not limited to genes that either have been cloned in heterologous systems, are differentially or abundantly expressed, or for which the corresponding protein is isolated or identifiable. Thus, genes not previously cloned, and especially genes that encode rare messages for (presently) unknown functions, can be obtained. Chromosome walking is facilitated in *Arabidopsis* due to the small genome and lack of dispersed repeats. These features are not known for any other higher plant. The use of the RFLP map for gene cloning is in progress in both

the Meyerowitz laboratory and in other laboratories. In the Meyerowitz laboratory, overlapping clones of a cosmid library are being isolated in walks toward the genes responsible for the following mutations: homeotic floral mutations *apetala*-2, *apetala*-3, and *pistillata*, a meristic floral mutation called *agamous*, and a mutation in phytohormone response called *ethylene resistance*. The isolation of these genes and others will represent a significant opportunity to study plant growth and development.