The DNA of *Arabidopsis thaliana*

Leslie S. Leutwiler*, Barbara R. Hough-Evans and Elliot M. Meyerowitz

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

**Summary.** *Arabidopsis thaliana* is a small flowering plant of the mustard family. It has a four to five week generation time, can be self- or cross-pollinated and bears as many as 10^6 seeds per plant. Many visible and biochemical mutations exist and have been mapped by recombination to one of the five chromosomes that comprise the haploid karyotype. With the experiments reported here we demonstrate that *Arabidopsis* has an extraordinarily small haploid genome size (approximately 7 x 10^7 nucleotide pairs) and a low level of cytosine methylation for an angiosperm. In addition, it appears to have little repetitive DNA in its nuclear DNA, in contrast to other higher plants.

**Introduction**

Plants have large differences in their nuclear DNA content. Within angiosperms there is a nearly thousand-fold range of variation, and there appears to be no correlation between genome size and organisal complexity (Bennett and Smith 1976). Most of the angiosperms currently used in molecular genetic studies have large genomes; correlating with these large (more than 10^9 nucleotide pairs) genomes is a large fraction of DNA that is repeated many times, with individual repetitive elements dispersed throughout the genome (Flavell 1980). Both of these properties, large haploid genome and high fraction of dispersed repetitive DNA, create difficulties in the standard procedures of molecular genetics: a large genome requires that large genomic libraries be screened to obtain any individual gene, and collection of overlapping clones to define a chromosomal region is greatly hindered by dispersed repetitive elements.

*Arabidopsis thaliana* is a small flowering plant of the mustard family with a short generation time (four to five weeks), the ability either to cross- or self-pollinate, and a haploid chromosome number of only five. It has been used in a number of experiments in classical and biochemical genetics (Rédei 1970). Numerous visible and biochemical mutations are known, and many of these have been mapped by recombination to positions on the five linkage groups (Koornneef et al. 1983). It is susceptible to infection by *Agrobacterium tumefaciens*; there thus exists a potential for using Ti plasmid as a means of achieving DNA-mediated transformation of *Arabidopsis* (Aerts et al. 1979). Microspectrophotometric measurements of nuclear DNA content in this plant indicate that it has a small haploid genome (Bennett and Smith 1976). Thus, *Arabidopsis* has several advantages as an organism for use in molecular genetic studies of higher plants. As a preliminary to a number of such studies, we here report experiments that characterize the composition and organization of *Arabidopsis* DNA, including measurements of kinetic complexity, genome size, and nucleotide composition. We find that *Arabidopsis* has an extraordinarily small genome (approximately 7 x 10^7 nucleotide pairs) and little repetitive DNA. This confirms its potential as an organism for use in molecular genetic experimentation, and provides basic information as a foundation for such studies. In addition, it establishes a new minimum genome size for a higher plant.

**Materials and methods**

*Biological material. Seeds of Arabidopsis thaliana* strain Columbia were obtained from Dr. A. Kleinhofs, Program in Genetics, Washington State University, Pullman, WA 99164. Wheat germ (Fisher Natural – not toasted) was obtained from a local market.

*DNA isolation. Arabidopsis* seeds were germinated on a mixture of sterile soil and peat (mixed 3:1) and grown under constant illumination (7000 lux) at 25°C and 60% relative humidity. Approximately 10 g of five week old plants were harvested, rinsed with water, and ground in a mortar and pestle with an equal weight of glass beads in the presence of liquid nitrogen. Two grams of the powder were added to 10 ml 150 mM Tris-HCl, pH 8.5; 100 mM EDTA; 2% N-lauroyl sarcosine, sodium salt; 0.1 mg/ml proteinase K and incubated at 37°C with gentle stirring for 30 min. The residue was removed by centrifugation and re-extracted twice; the supernatant was ethanol precipitated to remove a fluorescent compound which interfered with the ethidium bromide fluorescence of subsequent steps. Following centrifugation of the ethanol precipitate, the pellets were resuspended in 10 mM Tris-HCl, pH 8.0; 1 mM EDTA (TE) and banded to equilibrium in CsCl density gradients in the presence of 1 mg/ml ethidium bromide. The bands were collected, the ethidium bromide was removed by butanol extraction, and the DNA ethanol precipitated. The pellets were resuspended in 1 ml TE, loaded on a 1 M NaCl solu-
tion and centrifuged in a SW 50.1 rotor at 40,000 rpm for 5 h in order to remove any residual RNA. The DNA pellet was resuspended in TE, and the DNA concentration was measured by absorption at 260 nm minus that at 320 nm to correct for light scatter.

Wheat germ DNA was prepared essentially according to the above procedure. Wheat germ (0.5 g) was ground in an ice cold mortar and pestle with an equal weight of glass beads in the presence of 3 ml of proteinase K (10 mg) in 150 mM Tris-HCl, pH 8.5, 100 mM EDTA. Following grinding, the slurry was made up to 10 ml with N-lauroyl sarcosine (to a final concentration of 2%) in the Tris-EDTA mixture and incubated at 37°C with gentle stirring for 30 min. The remaining procedure for DNA isolation was the same as that for Arabidopsis DNA, except the initial ethanol precipitation step was omitted.

Chloroplast DNA isolation. Whole plants (8.75 g) were dark adapted 12-18 h and homogenized in 40 ml buffer A (0.3 M mannitol, 0.05 M Tris, 0.003 M EDTA, 0.001 M mercaptoethanol, 0.1% bovine serum albumin, pH 8.0 – Kolodner and Tewari 1975) with three 5-second bursts at high speed in a Waring blender. Following filtration through 4 layers of Miracloth, the supernatant was layered on 2 M sucrose: 80% Percoll: 60% Percoll: 40% Percoll in buffer A (0.6:1:1:1 v/v) and centrifuged at 8,000 rpm in a Sorvall HB-4 rotor for 30 min. The chloroplast band at the 40%–60% Percoll interface was collected, diluted in buffer A and pelleted in a HB-4 rotor at 2,000 rpm for 5 min. The pellet was solubilized in 5 ml 0.05 M Tris-HCl, pH 8.0, 0.02 M EDTA containing 2% N-lauroyl sarcosine and 100 μg/ml proteinase K, incubated at 37°C 20 min and phenol extracted twice. The aqueous phase was re-extracted with chloroform: 1% isooamyl alcohol and ethanol precipitated. The DNA was further purified on a CsCl equilibrium density gradient.

Preparation of labeled probes. A. thaliana total DNA or Drosophila melanogaster embryo DNA was nick translated (Rigby et al. 1977) with [3H]-deoxycytidine triphosphate, and a lambda clone containing a chloroplast DNA insert (ZbAt003) was nick translated with [32P]-deoxycytidine triphosphate. The nick translation mixture was incubated at 13°C for 45 min and the DNA purified by passage over a Bio-Gel P-60 (100–200 mesh; Bio-Rad) column.

The nick translation yields a mixture of fragments ranging in size from greater than 1700 base pairs (bp) down to less than 154 bp. Therefore, it was necessary for the reassociation measurements to isolate single-stranded DNA fragments longer than 500 bp so that they (1) could be sheared uniformly with the unlabeled DNA and (2) were long enough to reassociate under the hybridization conditions. DNA from the plasmid pBR325 (Bolivar 1978) was digested with HinfI and used as a standard for size measurements. In order to make the nick translated DNA single-stranded the DNA probe was made 0.1 N in NaOH prior to loading. Following electrophoresis through 2% agarose gels to separate the different size DNA strands, a slot was cut in the gel perpendicular to the electric field in the track containing the nick translated DNA and at the location for fragments of 500 bp. A piece of Whatman DE 81 paper was inserted and electrophoresis continued until all activity present initially in the gel on fragments greater than 500 bp could be found on the DE 81 paper. The piece of DE 81 paper was then rinsed three times with 0.1 M Tris-HCl, pH 7.5, 0.1 M NaCl and the DNA eluted with 0.1 N NaOH.

Shearing of DNA. Unlabeled Arabidopsis (and Drosophila) DNA and labeled tracer DNA of the correct size (described above) were mixed before shearing in a Virtis homogenizer as described by Britten et al. (1974). Ten ml of the DNA mixture in TE was diluted with 20 ml glycerol and sheared in a Virtis homogenizer at 60,000 rpm in a dry ice-ethanol bath for 30 min. The average single-strand length of the DNA was determined by electrophoresis through a 2% agarose gel using HinfI digested pBR325 DNA as size standards. The unlabeled DNA was visible by UV fluorescence while the [32P]-labeled DNA was visualized by autoradiography of the gel. Both procedures indicated the DNA had been sheared to an average single-strand length of 375 nucleotides. After shearing the DNA was ethanol precipitated and the precipitate subsequently dissolved in a small volume of TE. This solution was layered on a Chelex-100 (200–400 mesh, sodium form; Bio-Rad) column in order to remove heavy metal ions; the fractions containing DNA were combined and reprecipitated with ethanol. The precipitate was resuspended in 0.4 M sodium phosphate buffer pH 6.8 (PB) and the DNA concentration determined.

DNA reassociation. Reaction mixtures were prepared and analyzed essentially according to Britten et al. (1974). DNA samples ranging in concentration from 3.33 μg/ml to 1.375 μg/ml in either 0.12 M or 0.4 M PB were sealed in capillaries or ampoules, denatured by boiling for 1 min, and reassociated in a 60°C water bath for the required incubation times. After reassociation to the appropriate Cot, the samples were frozen in dry ice-acetone and kept frozen at –20°C.

Reassociated DNA was separated from single-stranded DNA by hydroxylapatite (Bio-Rad, DNA Grade Bio-Gel HTP) chromatography in water-jacketed columns. DNA samples reassociated in 0.12 M PB were thawed and loaded directly on the column (2 ml bed volume), while those samples reassociated in 0.4 M PB were diluted to 0.12 M PB before loading. Single-stranded DNA was eluted in five to six 1 ml fractions with 0.12 M PB at 60°C; double-stranded DNA bound to hydroxylapatite was denatured by heating the column to 98°C and eluted in five 1 ml fractions with the same buffer. Reassociation of unlabeled DNA was determined by measuring the A260 nm for each fraction and correcting for light scattering at 320 nm. Reassociation of labeled DNA was determined subsequently by counting each fraction in Aquasol-2 (New England Nuclear) in a Beckman LS-250 scintillation counter. Tritium and [32P]- cpm were measured simultaneously by counting in a half-tritium channel and a channel spanning the [14C]- [32P] energy range respectively.

Cot values were calculated for each sample by multiplying the DNA concentration (moles nucleotide per liter) by time (sec). Reassociation data were fit by the equation:

\[ \frac{C}{C_0} = \frac{1}{1 + kC_0 t} \]

where \( C \) = concentration of single-stranded DNA at time \( t \), \( C_0 \) = original DNA concentration, and \( k \) = second order reassociation rate constant. The results were analyzed by a non-linear least-squares computer program (Pearson et al.