A highly repeated DNA sequence in *Arabidopsis thaliana*

Jose M. Martinez-Zapater, Mark A. Estelle, and Chris R. Somerville
MSU-DOE Plant Research Laboratory, Michigan State University, East Lansing MI 48824, USA

**Summary.** Three members of a family of highly repeated DNA sequences from *Arabidopsis thaliana* have been cloned and characterized. The repeat unit has an average length of 180 bp and is tandemly repeated in arrays longer than 50 kb. This family represents more than one percent of the *Arabidopsis* genome. Sequence comparisons with tandemly repeated DNA sequences from other Cruciferae species show several regions of homology and a similar length of the repeat unit. Homologies are also found to highly repeated sequences from other plant species. When the sequence CCGG occurs in the repeated DNA, the inner cytosine is generally methylated.

**Key words:** Repetitive DNA – *Arabidopsis*

**Introduction**

A large fraction of the genome of higher eukaryotes is composed of families of repeated DNA sequences. Individual members of a family can be dispersed throughout the genome or clustered in long arrays of tandem repeats. Tandemly repeated DNA can sometimes be isolated as satellite bands in density gradients and has also been referred to as satellite DNA. Analysis of DNA reassociation kinetics usually resolves this DNA as a rapidly reassociating fraction mainly composed of a unit DNA sequence tandemly repeated thousands of times. Where cytological analysis has been carried out, this DNA has always been found associated with heterochromatic regions of the chromosome (see John and Miklos 1979; Brutlag 1980; Singer 1982 for reviews).

Remarkable variation is observed between highly repeated DNA sequences from different species, indicating a rapid rate of evolution. Sequence comparisons between different families of repetitive DNA from the same species indicate that this sequence divergence occurs through random mutation and some unknown amplification mechanism (Brutlag 1980; Singer 1982). There is no experimental evidence regarding the function of satellite DNA. However, the association with heterochromatin suggests that it may be important for chromosome structure. The observation that in some organisms satellite DNA and heterochromatin are drastically reduced in nuclei with somatic functions is consistent with the idea that satellite DNA has some role in germ line processes (i.e. meiotic chromosome pairing, recombination and evolutionary processes) (Bostock 1980). On the other hand it has also been suggested that repeated DNA has no functional significance and simply accumulates in the genome because it does not produce any phenotypic disadvantage (Doolittle and Sapienza 1980; Orgel and Crick 1980).

In general, plant species may show very different nuclear DNA contents, even within the same family. The fraction of the genome corresponding to repeated DNA increases with increasing DNA content. Thus, variation in DNA content between species seems to be due mainly to variation in the amount of repeated DNA (Flavell et al. 1974; Flavell 1982). Highly repetitive DNA has been studied in several plant species at the molecular level: rye (Bedbrook et al. 1980), wheat and barley (Dennis et al. 1980), *Scilla* (Deumling 1981), maize (Peacock et al. 1981; Viotti et al. 1985), mustard (Capesius 1983), broad bean (Kato et al. 1984) and radish (Grellet et al. 1986). This DNA is composed of tandemly repeated sequences of different complexities. “In situ” hybridization analyses performed in a number of monocot species using labeled satellite DNA have always revealed an association with heterochromatic regions of the chromosome, including telomeres (Bedbrook et al. 1980; Deumling 1981; Deumling and Greilhuber 1982), centromeres (Dennis et al. 1980; Hutchinson and Lonsdale 1982), interstitial blocks (Dennis et al. 1980), maize knob heterochromatin (Peacock et al. 1981) and B chromosomes of maize (Dennis et al. 1980).

*Arabidopsis thaliana*, with a genome size of 70,000 Kb, has the lowest DNA content known among higher plant species (Leutwiler et al. 1984). Repeated DNA sequences represent about 25% of its genome, which is also a relative low proportion by comparison with other plant species (Meyerowitz and Pruitt 1985). About one half of this repeated DNA behaves as a rapidly reassociating fraction (Leutwiler et al. 1984) and the other half corresponds to middle repetitive sequences, in which the major component is rDNA (Pruitt and Meyerowitz 1986). Therefore, the relative simplicity of its repeated DNA plus the suitability of *Arabidopsis* for molecular studies (Meyerowitz and Pruitt 1985) make it a good system for the study of the structure and function of repetitive sequences. Here we report the cloning and sequence characterization of several elements from one family of repeated DNA sequences that represents more than 1% of the *Arabidopsis* genome.
Materials and methods

Plant material and growth conditions. Plants of Arabidopsis thaliana (L.) Heynh. Columbia wild-type were grown in continuous fluorescent illumination (150 μE m⁻² s⁻¹) at 22°C on a perlite:vermiculite:sphagnum (1:1:1) mixture irrigated with mineral nutrients (Somerville and Ogren 1982).

DNA extraction and molecular cloning. Total DNA was purified as described by Leutwiler et al. (1984) from leaves of 3–4 weeks old Arabidopsis rosettes. To clone the repeated fragments, we fractionated a total HindIII digestion of cellular DNA in a 1% low melting agarose gel. Fragments corresponding to approximately 180 and 360 bp were isolated from the gel and ligated into the HindIII site of pUC19. The ligation mixture was used to transform E. coli JM83 to ampicillin resistance. Two plasmids pAS1 and pAS2 containing the 180 bp fragment and one pAL1 containing the 360 bp fragment were retained from among the ampicillin resistant colonies.

Hybridization analysis. Southern hybridizations were carried out for 24 h at 42°C in 50% formamide and 5 x SSC and washed for 1 h at 65°C in 0.1 x SSC. For all hybridizations the probe was a 178 bp HindIII fragment from pAS1 representing one repeat unit. This fragment was purified by fractionation on a low melting agarose gel, and separated from the melted agarose by NACS (BRL) chromatography. After purification, it was labelled by nick-translation using ³²P-deoxyctydine triphosphate (Rigby et al. 1977). For dot blots, defined amounts of total cellular DNA and EcoRI restricted pASI plasmid DNA, were denatured in 0.4 N NaOH and neutralized with an equal volume of 2 M NH₄COOH. DNA samples were then applied to nitrocellulose filters and baked at 80°C for 2 h. These filters were hybridized under the same conditions as described for Southern blots. After autoradiography the number of cpm bound to the filter at each concentration of pASI and cellular DNA was determined by counting the appropriate region of the filter by liquid scintillation counting. Comparison of the radioactivity bound within the linear range of response was used to estimate the relative amount of the 180 bp sequence per haploid genome.

DNA sequencing. Sequencing was carried out following a modification of the dideoxynucleotide chain-terminator method of Sanger et al. (1977). Inserted DNA subcloned in the HindIII site of M13mp18 in both orientations was used as a template for the sequencing reaction using the 15mer universal primer (New England Biolabs) and ³²S-deoxythioadenosine triphosphate (Biggin et al. 1983). Reaction products were fractionated on an 8% (w/v) acrylamide buffer gradient gel following the method of Biggin et al. (1983). Sequence comparisons were made using previously described computer programs (Pustell and Kafatos 1984).

Results

Detection and isolation of the repeated element

Total restriction of Arabidopsis thaliana DNA with HindIII showed, after fractionation on a 1% agarose gel and ethidium bromide staining, two prominent bands correspond-