The centromere region of *Arabidopsis thaliana* chromosome 1 contains telomere-similar sequences

Eric J.Richards*, Howard M.Goodman and Frederick M.Ausubel
Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114 and
Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

Received March 4, 1991; Revised and Accepted April 30, 1991 EMBL accession nos X58101–X58106 (incl.)

ABSTRACT

We describe the structure of an *Arabidopsis thaliana* genomic clone containing two classes of repetitive DNA elements derived from the centromere region of chromosome 1. One class is comprised of tandem arrays of a highly reiterated repeat containing degenerate telomere sequence motifs. Adjacent to these telomere-similar repeats we found a dispersed repetitive element reiterated approximately five times in the *A. thaliana* genome. The nucleotide sequence of the dispersed repeat is unusual, being extremely AT-rich and composed of numerous, overlapping repeat motifs.

INTRODUCTION

The DNA sequences at the extreme chromosomal termini of many, perhaps all, eukaryotes is composed of tandem repeats of short, G-rich sequence blocks (described by the loose consensus 5'[T/$A_{1-4}G_{1-8}$]3' [as read towards the chromosomal terminus]) (reviewed in 1). DNA sequences which are similar to these telomeric repeats have also been found at various non-telomeric locations within the genome of many organisms.

Telomere-similar DNA is frequently found in telomere-flanking regions. For example, most *Saccharomyces cerevisiae* telomere-flanking regions contain short (50–150 bp) tracts of the telomere-similar sequence, $(TG_{1-3})_n$, separating tandemly arranged units of the moderately repeated elements, X and Y' (2). Similarly, in the protozoan *Plasmodium berghei*, approximately 150 bp of degenerate telomere-similar sequences are found embedded within a 2.3 kb tandem repeat located in subtelomeric regions (3, 4). Degenerate telomere motifs have also been noted in telomere-flanking regions in Trypanosomes (5).

Not all telomere-similar sequences are located close to chromosomal termini, however. Some of the earliest characterized centromeric satellite DNAs were found to be composed of tandemly repeated variants of TTAGGG (6, 7). More recently, *in situ* hybridization experiments, using a human telomere repeat $(TTAGGG)_n$, have revealed the presence of telomere-similar sequences at the centromere of human chromosome 2 and the centromeres of several vertebrate chromosomes (8, 9).

The presence of telomere-similar sequences at the centromere is not restricted to animal chromosomes. In this report, we describe the structure of repetitive DNA elements from the flowering plant, *Arabidopsis thaliana*, which contain telomere-similar sequences but which lie in the centromere region of chromosome 1. In addition, we report the structure of an unusual repetitive element which resides next to an array of these telomere-similar repeats.

METHODS

Plasmids and probes

Four cloning/sequencing vectors were used in this study: pSDC13 (10), pUC12 (11), M13mp18 (12) and pBluescript KS– (Stratagene). *A. thaliana* telomeric probes were generated by radiolabeling the 0.4 kb *Pst*I-*Bam*HI insert of pAtT4 (13) which contained tandem repeats of the *A. thaliana* telomeric sequence, 5'[TTTAGGG]-3'.

DNAs were prepared by standard procedures (14). Radioactive probes were made by the random priming method using kits purchased from Boehringer Mannheim Biochemicals.

Source of enzymes

Restriction enzymes and calf intestinal phosphatase were purchased from New England Biolabs or Boehringer Mannheim Biochemicals. New England Biolabs was the source for *Bal*31 nuclease. T4 DNA ligase was purchased from New England Biolabs or United States Biochemicals.

Southern hybridization

Total genomic DNA from the Landsberg and Columbia ecotypes were prepared as described in Ausubel et al. (14). *A. thaliana* nuclear DNA was prepared by the protocol outlined in Olszewski et al. (15).

Exonuclease digestion of *A. thaliana* nuclear DNA was performed using 0.25 U/ml of *Bal*31 nuclease at 30°C at a DNA concentration of 10 μg/ml in 12 mM $CaCl_2$, 24 mM $MgCl_2$, 0.2 M NaCl, 20 mM Tris-HCl (pH 8.0), 1 mM EDTA, 100 μg/ml BSA.

Southern blots were prepared using nylon membranes (GeneScreen, NEN) and the UV cross-linking protocol of Church

* To whom correspondence should be addressed at Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA

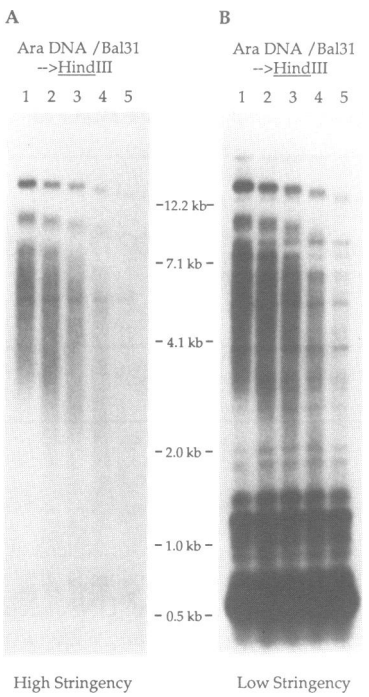

Figure 1. *A. thaliana* telomeric repeats cross-hybridize to non-telomeric sequences at reduced stringencies. *A. thaliana* nuclear DNA was treated with *Bal*31 for 0 (lane 1), 5 (lane 2), 15 (lane 3), 30 (lane 4), or 50 minutes (lane 5) and subsequently digested with *Hin*dIII. The DNAs were then size-fractionated by electrophoresis through a 0.8% agarose gel and transferred to a nylon membrane. The membranes were probed with radiolabeled pAtT4 insert DNA and washed at either high stringency ((A) 0.1×SSC, 0.1% SDS @ 60°C) or low stringency ((B) 2×SSC, 1% SDS @ 60°C).

and Gilbert (16). Hybridization were carried out in 0.5 M $NaHPO_4$ (pH 7.2), 7% SDS, 1 mM EDTA, 1% BSA (16). Hybridization temperatures and wash conditions are noted in the appropriate figure legends.

Construction and screening of genomic libraries

The pAtT12 genomic clone was obtained from a size-selected library constructed to isolate large telomeric inserts. *A. thaliana* nuclear DNA (2 μg) was digested with *Bal*31 nuclease to remove approximately 30 bp from the termini of the chromosomal fragments (DNA concentration = 10 μg/ml, enzyme concentration = 0.25 U/ml, 30°C, 60 seconds). The cloning vector pSDC13 was prepared by sequential digestion with *Hin*cII and *Bam*HI, and treated with calf intestinal phosphatase to remove 5′ phosphates. The nuclease-treated DNA was then ligated to 0.6 μg of prepared vector using T4 DNA ligase (DNA concentration = 130 μg/ml, enzyme concentration = 20,000 U/ml, 22°C). The ligation products were digested with *Bam*HI and size-fractionated by agarose gel electrophoresis (0.7%, low melting agarose). Linear vector-genomic DNA chimeras of 11−15 kb were purified and circularized with T4 DNA ligase (DNA concentration = 0.5 μg/ml, enzyme concentration = 130 U/ml, 4°C). The circularized DNAs were then transformed into K802*recA* using the protocol developed by Michael Scott (pers. comm.).

Clones cross-hybridizing with the pAtT4 insert were identified by colony hybridization (17). Colonies were grown on Colony/Plaque Screen (NEN) filters and lysed by autoclaving the membranes for 1 minute. Hybridization was carried out at

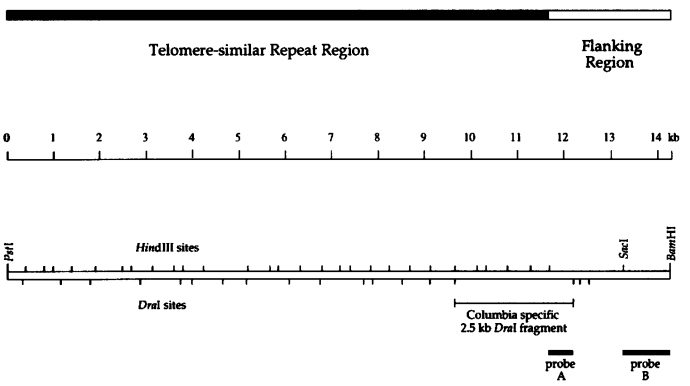

Figure 2. Molecular organization of telomere-similar clone pAtT12. The 14.3 kb insert of pAtT12 contains two domains: an 11.7 kb region containing highly reiterated telomere-similar DNA sequences, and a 2.6 kb 'flanking region' devoid of telomere-similar sequences. The restriction map of the insert, shown at the bottom of the figure, illustrates the high density and periodic distribution of *Hin*dIII and *Dra*I sites in the telomere-similar region. The insert was cloned into pSDC13 as a *Bam*HI to blunt end fragment; the *Pst*I site shown at the left end of the insert is derived from the vector polylinker. The hybridization probes (A and B) and the 2.5 kb Columbia-specific *Dra*I fragment used in the RFLP mapping experiments are shown at the lower right.

60°C using 1 M NaCl, 10% dextran sulfate, 1% SDS, 25 μg/ml tRNA and a probe concentration of 10^5 cpm/ml. The filters were washed at 60°C with 0.2×SSC, 0.1% SDS, a stringency which apparently allows the degenerate telomere-similar sequences to by recognized by the telomere repeat probe (compare to 'high stringency' wash [0.1×SSC, 0.1%SDS @ 60°C] in Figure 1A).

Restriction mapping of the pAtT12 insert was performed by the endlabeling/partial digestion protocol of Boseley et al. (18).

DNA sequencing

The ≈500 bp and ≈180 bp *Hin*dIII fragments from the pAtT12 insert were prepared for sequencing by subcloning into either pUC12 or M13mp18 using standard procedures. The following subclones were sequenced: pAtT20 (527 bp), pAtT24 (491 bp), pAtT25 (483 bp), pAtT28 (183 bp) and pAtT29 (183 bp).

The pAtT12 flanking region was subcloned into pBluescript KS− to generate pAtT27. Deletion derivative of pAtT27 suitable for sequencing were made using the nested exoIII/nuclease S1 deletion procedure (14).

Dideoxy sequencing reactions were carried out, on both single- and double-stranded templates, using Sequenase enzyme and kits purchased from United States Biochemicals. Some regions were sequenced using oligonucleotide primer made on a Biosearch DNA synthesizer (New Brunswick Scientific).

RFLP mapping

The RFLP mapping was done as described in Nam et al. (19). Briefly, inter-ecotype crosses between wild type Columbia and various Landsberg marker lines were conducted. Genomic DNA was prepared from individual, phenotyped F2 plants (represented by pooled F3 progeny) and Southern blots of these DNAs were prepared. The blots were then hybridized with a collection of anonymous cosmid clones and selected clones of interest. The resulting hybridization patterns were analyzed and segregating RFLP alleles noted. The likely orders and map positions of the RFLPs were determined using the MAPMAKER program (20).

```
Plurality: 1    AAGCTTTGAGAATCAAGAAGAAGCTTTG.TAGCGGATTTGTAGTCAAATATGACTAGATGCTTTGTGTATTATTGAGCATAAGAACTAGAACCGCAACCA  100
pAtT20:                          .....G.G.....................................................................T......G
pAtT24:         ...................TG.....AT.........A...C.......T.................CG.....................................
pAtT25:                          ........C.....................T.....................................A...........

Plurality: 101  GATTCCGGAAGCCTAAAGTAGGATTTTGGTTTTAAAGTTTGGGATCTATGGTTTAAGTTTTTTTGGGTTTAGGTTTAAAGGTTTAGGGTTTAGGTTTAAG  200
pAtT20:         ..........................A.......G.....................................................................
pAtT24:         .C....AA...-.....................A...........GA.........----------------------------T..............
pAtT25:         ..............................................T....-----------------------------------------..

Plurality: 201  GGTTTAGGTTTAAGGGTTTAGGGTTAAGAGTTTATGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTT..GGGTAGGAT  300
pAtT20:         ..................................................G..........A.........................TT.....A...
pAtT24:         ........G...C.A.............G.....G......A.....T.........T..........T..A...TTA.....------.......
pAtT25:         A..............A.................T.............C.....T.........A.........A....G.........AG..T.T....

Plurality: 301  TTATGGTTTGGGGATTTA.GGTGTAGAGTTTAGGTTTAGTGTTTAGGAGCAAATATACTAT.A..CTTTGAAAAGCAAGAAGAATCTTGGTTAGCATTTC  400
pAtT20:         ..................T.....................................A.TAT...A.G.....................................
pAtT24:         .....T..A...--....A..-------.....................G.....------------...A.....T...C.C......TT.A...G...T
pAtT25:         ..TA.....T........G.............................T........G...G.TA...A..G.............G........G....
pAtT28:                                                                  .AG.......G.............G.............
pAtT29:                                                                  .AG.......G.............G.............

Plurality: 401  GGAGTGAAATATGACTAGATGTCATGTGTATGATTGA..ATATAAAACTAGAACCTCAACGAGATCCCGAAAAGTAAAGTAGT.CTTCCTTGTTATACGA  500
pAtT20:         A..............C.....................---...G.................C.....T.................G...........A.T..
pAtT24:         ...A.C......CT.A.....AT.G...G.........-GT.T..GGG...A....A....A..GG.TATG..CC....T..-.CT....G.....G.A..
pAtT25:         ..................................AT.........T..T.................................G.................
pAtT28:         ..............G.....................--.....................................G...........A.............
pAtT29:         ..............G.....................---.T.....................................G...........A.............

Plurality: 501  TTCAAAAC..CAAAAACTCATAAGAACTT.GGCTTCACCATCAAAGCTT
pAtT20:         ........AT.................T..C.....................
pAtT24:         .C.......--T.....-.....C.G....A....A..T...........
pAtT25:         .....G..--T..............T...T.A.....................
pAtT28:         ........--.........GT.........T..G.......G.........
pAtT29:         ........--.........GT.........C..G.....................
```

Figure 3. DNA sequence of repetitive elements from the telomere-similar region of pAtT12. The DNA sequence of three ≈500 bp telomere-similar repetitive elements (pAtT20, pAtT24 and pAtT25) are aligned with the sequence of two interspersed ≈180 bp elements (pAtT28 and pAtT29). Nucleotide identities to the consensus (plurality) are denoted by dots, and difference are shown in the corresponding lines. Dashes denote deletions. The simple-sequence domain containing degenerate telomere motifs is located from nucleotides 150 to 350.

RESULTS

A. thaliana telomeric repeats cross-hybridize with non-telomeric sequences at reduced stringency

We have previously described the isolation of an *A. thaliana* telomere clone, pAtT4, containing tandem repeats of the sequence 5′[TTTAGGG]3′ (13). The pAtT4 insert hybridized to both telomeric and non-telomeric restriction fragments under reduced stringency conditions. Southern blots of nuclease *Bal*31-treated, *Hin*dIII-digested *A. thaliana* nuclear DNA were probed with radiolabeled pAtT4 and washed at different stringencies. Under high stringency conditions (Figure 1A) the pAtT4 probe recognized dispersed, exonuclease-sensitive bands characteristic of telomeric restriction fragments. A few faint bands, which were insensitive to the exonuclease treatment, were also detected by the telomeric probe under high stringency conditions. In contrast, when the stringency of hybridization was reduced, a different hybridization pattern was seen. The *Bal*31-sensitive telomeric signals were still detected, but the majority of the hybridization signal corresponded to several discrete, exonuclease-insensitive bands (Figure 1B). Most prominent among these exonuclease-insensitive signals was a band corresponding to *Hin*dIII restriction fragments of approximately 500 bp.

These data indicate that the *A. thaliana* genome contains sequences resembling telomeric repeats which are not located at the chromosomal termini. Since the most extensive nuclease *Bal*31 digestion (Figure 1, lane 5) removed approximately 2.5 kb from the chromosome ends but did not alter the electrophoretic migration of the cross-hybridizing bands, we conclude that most of the exonuclease-insensitive telomere-similar sequences are located farther than 2.5 kb from the chromosomal termini. The fact that the telomere-similar sequences were only significantly detected under reduced stringency conditions suggests that these

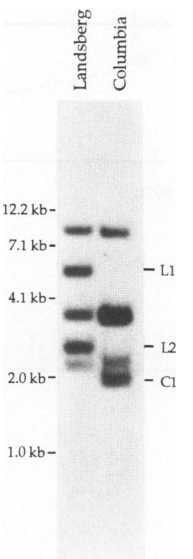

Figure 4. The flanking region of pAtT12 contains a moderately repetitive element. Total genomic DNA from the *A. thaliana* ecotypes Landsberg and Columbia were digested with *Dra*I and size-fractionated by electrophoresis through a 1% agarose gel. The DNAs were then transferred to a nylon membrane and hybridized with radiolabeled probe B (a 1 kb *Sac*I fragment from the right edge of the pAtT12 insert; the *Sac*I sire which lies directly adjacent to the *Bam*HI site in the vector polylinker was used for convenience). The filter were washed in 0.2×SSC, 0.1×SDS @ 60°C. The polymorhpic bands corresponding to the RFLP markers L1, L2 and C1 are shown. As described in the text, the L2 and C1 pair of polymorphic bands behaved as alleles and mapped to chromosome 1 at the location identified by the probe A RFLP. The L1 allele mapped to the central region of chromosome 5. The dark ≈3.5 kb band in the Columbia lane is a doublet which was not resolved on this gel.

sequences contain either short stretches of telomeric repeats or repeats which are similar but not identical to the [TTTAGGG] telomeric motif.

Isolation and characterization of an *A. thaliana* telomere-similar sequence clone

To determine the nature of the telomere-similar sequences we isolated several clones containing these sequences from an *A. thaliana* genomic library (Columbia ecotype) using the insert of the telomeric clone pAtT4 as a hybridization probe. We analyzed the structure of one of these clones, designated pAtT12, in detail.

A restriction map of the pAtT12 insert is shown in Figure 2. Southern blot experiments indicated that the 2.6 kb region residing at the right end of the insert as drawn in Figure 2 (referred to as 'flanking region') does not contain telomere-similar sequences (data not shown). The remaining 11.7 kb of the 14.3 kb insert is comprised of many short *Hin*dIII and *Dra*I restriction fragments, most of which cross-hybridized with the telomeric probe. The high density and relatively even spacing of *Hin*dIII and *Dra*I restriction sites in the telomere-similar region (see Figure 2) suggests this region is composed of short repeat elements arranged in tandem. The majority of *Hin*dIII sites are spaced approximately 500 bp apart; however, there are several smaller intervals of ≈180 bp and a larger 1 kb interval. The 1 kb and ≈500 bp *Hin*dIII fragments cross-hybridized with the telomeric probe, while the ≈180 bp *Hin*dIII fragments did not (data not shown).

To elucidate the structure and organization of sequences in the telomere-similar region of pAtT12, we determined the DNA sequence of three ≈500 bp and two ≈180 bp *Hin*dIII restriction fragments from this region (Figure 3). The larger *Hin*dIII fragments range in size from 483 to 527 bp. These sequences display 79 to 89% similarity with each other, and contain a 150 to 200 bp simple-sequence domain (nucleotides 150–350 in Figure 3) which resembles the $[TTTAGGG]_n$ structure of *A. thaliana* telomeres. The variable length of the simple-sequence region accounts, in part, for the different sizes of the repeat units. Several identical matches to the telomeric motif are present in the simple-sequence region but most of the domain is composed of imperfect telomere-similar repeats. The degenerate nature of the telomere-like repeats explains why the telomeric clone pAtT4 hybridizes strongly to these repeats only under low stringency conditions. The ≈180 bp *Hin*dIII fragments are closely related to the ≈500 bp element but lack the telomere-similar domain.

The DNA sequence and restriction mapping data indicate that the pAtT12 insert is primarily composed of many copies of related telomere-similar repeats. The irregularities in the spacing of the *Hin*dIII sites seen in Figure 2 are the result of the variable size of the larger ≈500 bp repeat units as well as the interspersion

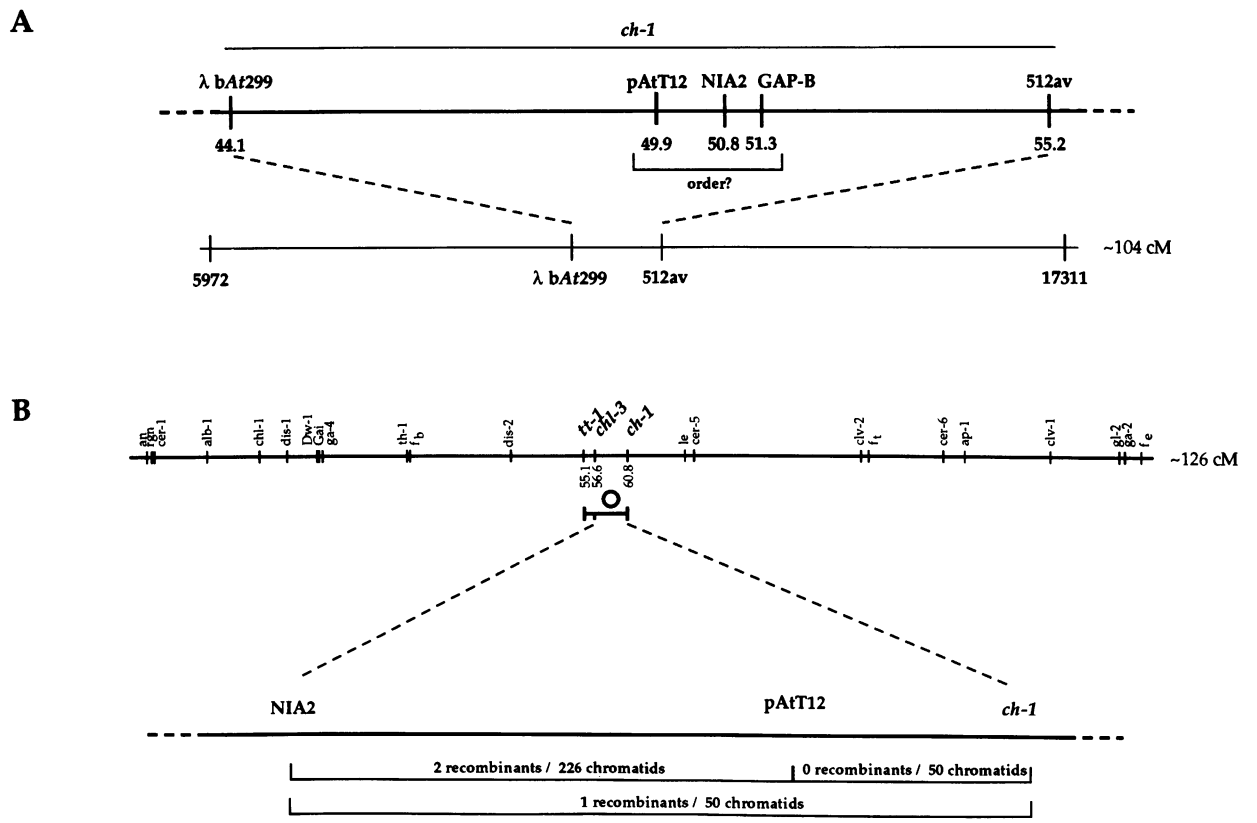

Figure 5. The pAtT12 insert was derived from the centromere region of chromosome 1. An RFLP map of *A. thaliana* chromosome 1, with the two most distal markers and a magnified view of the centromere region, is shown in A. The positions of the markers is given in cM. The brackets that group pAtT12, NIA2 and GAP-B markers indicates that the order of these markers is still uncertain at present. Similarly, the line under the morphological marker *ch1* indicates that, given the present data, this marker can be placed in any position in this region. B shows the classical genetic map of chromosome 1 (33) correlated with the corresponding region of the RFLP map as described in the text. The centromere maps between *tt1* and *ch1* (denoted by the bracket) and is more closely linked to the latter marker (indicated by the position of the circle). The approximate position of pAtT12 relative to *ch1* and NIA2/*chl3* is shown based on the number of infrequent recombination events which separated these markers as scored in the F2 progeny of a Landsberg×Columbia cross.

of the ≈180 bp variant repeats which lack the telomere-similar domain.

The pAtT12 insert most likely represents the organization of a considerable fraction of telomere-similar sequences in the *A. thaliana* genome. The size of the *Hin*dIII repeats corresponds to the prominent *Hin*dIII band at approximately 500 bp seen on genomic Southern blots (Figure 1B). The other exonuclease-insensitive *Hin*dIII bands may represent repeat multimers or related repeats of different size.

Chromosomal location of pAtT12 insert sequences

Highly reiterated tandem repeats are frequently located in constitutively condensed regions of the chromosome called heterochromatin (reviewed in 21). *A. thaliana* chromosomes contain heterochromatic regions around all five centromeres and some of the telomeres (22), features common to most higher eukaryotic organisms. These considerations prompted us to determine the chromosomal location of the telomere-similar repetitive DNA array contained in the pAtT12 insert.

Chromosomal localization of the pAtT12 insert sequences was accomplished by utilizing the RFLP mapping tools recently developed for *A. thaliana* (19). RFLP alleles recognized by low-copy number probes from the flanking region of the pAtT12 insert were genetically mapped in segregating F2 populations of an inter-ecotype cross (Landsberg×Columbia). The first hybridization probe used was a 470 bp *Hin*dIII-*Dra*I fragment (probe A, see Figure 2) which hybridized to a 2.5 kb *Dra*I fragment present only in genomic DNA from the Columbia ecotype parent. The position of the corresponding 2.5 kb *Dra*I fragment in pAtT12 is noted in Figure 2. An additional

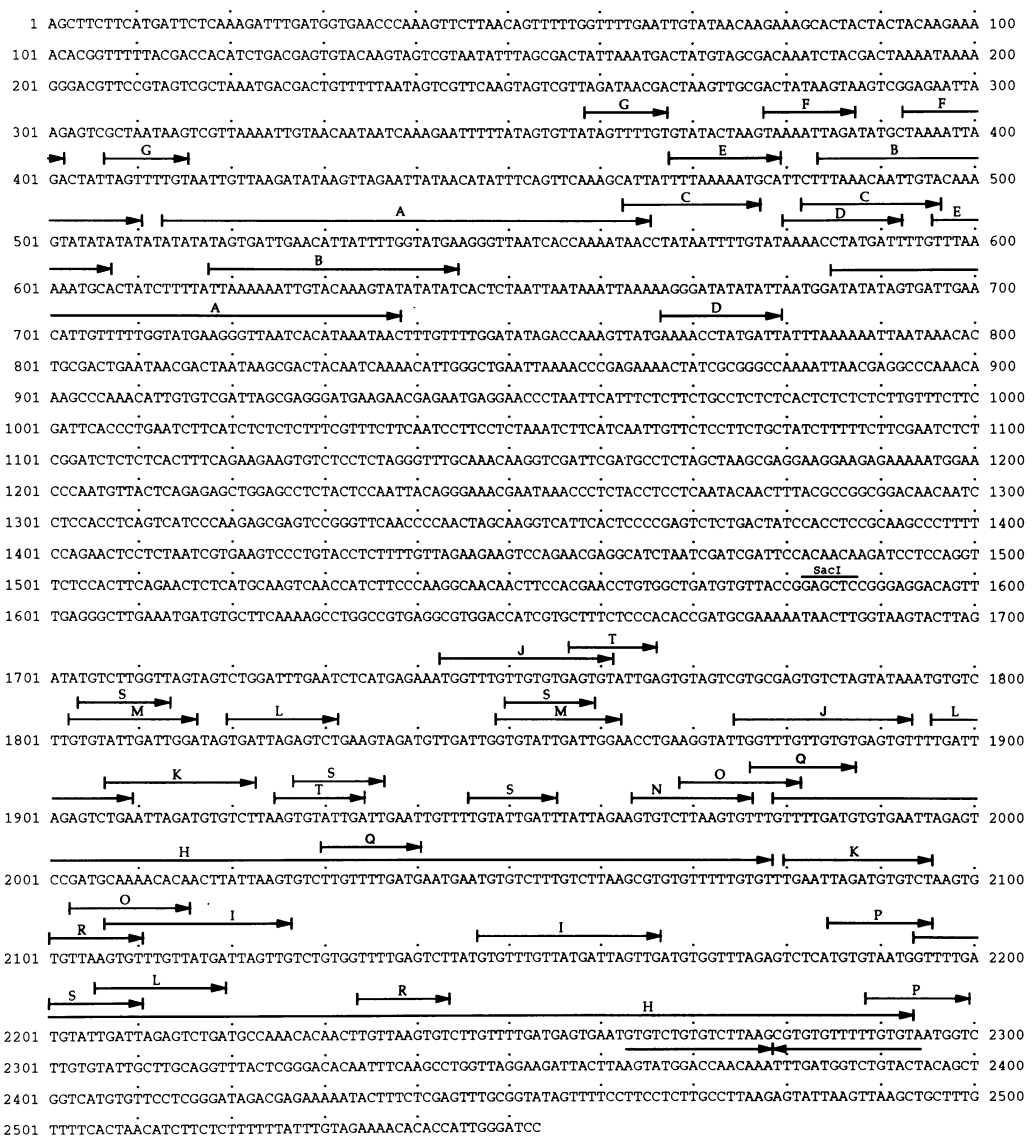

Figure 6. DNA sequence of the pAtT12 flanking region. The DNA sequence of the 2553 bp *Hin*dIII-*Bam*HI fragment corresponding to the flanking region is shown. The boundary between the telomere-similar repeat domain and the flanking region occurs around nucleotide 90, since nucleotides 1−89 share significant similarity with the telomere-similar repetitive elements. The locations of the direct repeats are indicated by the lettered arrows over the sequence. The individual members of the largest direct repeats, A and H, share 94% and 92% nucleotide similarity, respectively, with their cognate. The remaining repeats are identical to their cognate(s), with the exception of B and C which contain 1 bp mismatches. Most motifs are repeated twice, except L (3×) and S (5×). The largest palindromic sequence is shown (residue 2363−2394). The *Sac*I site which defines the left boundary of probe B begins at nucleotide 1582.

hybridization probe was derived from the extreme right end of pAtT12, designated probe B (see Figure 2). As shown in Figure 4, probe B hybridizes to approximately five bands on genomic Southerns, some of which are polymorphic between the two ecotype parents. A panel of Southern blots containing *Dra*I digested genomic DNA corresponding to the F2 segregants of the inter-ecotype cross were hybridized with probes A and B to determine a genotype for each plant at the locus corresponding to the pAtT12 insert. A genetic map position for the locus was then assigned relative to other RFLP and genetic markers.

As shown in Figure 5, the pAtT12 insert was derived from genomic sequences which reside in the middle of chromosome 1, in the vicinity of the centromere. The pAtT12 RFLP is closely linked to several other RFLP markers and the morphological marker *ch1*; because of this tight linkage the order of the markers in this region can not be unambiguously defined (See Figure 5A). To construct the simplified region of the RFLP map shown in Figure 5B, we placed a constraint on the order of the *ch1* and NIA2 markers by correlation with the genetic map. The NIA2 RFLP is defined by a nitrate reductase structural gene which corresponds to the *chl3* (reduced nitrate reductase activity leading to chlorate resistance) genetic marker (23, Jack Wilkinson and Nigel Crawford, pers. com.). Given this information, the most likely map position for the pAtT12 insert DNA is within 1 cM of *ch1* in the interval between *ch1* and NIA2.

Koornneef and co-workers have determined that centromere 1 is closely linked to *ch1*, and resides in the genetic interval between the *tt1* and *ch1* markers (24, 25). Based on these data, it is likely that the telomere-similar repetitive array cloned in pAtT12 is associated with the centromeric heterochromatin of chromosome 1.

Structure of the dispersed repetitive sequences in pAtT12 flanking region

As evidenced by the number of bands seen in Figure 4, the flanking region of pAtT12 contains repetitive DNA sequences. Subsequent attempts to find probes in the flanking region which hybridize to only one locus have failed. Consequently, the entire flanking region is comprised of repeated DNA sequences. Moreover, the repeats map to at least two different chromosomal locations since the Landsberg-specific RFLP allele L1 recognized by probe B maps to the central region of chromosome 5 (data not shown).

We determined the DNA sequence of the 2.6 kb flanking region of pAtT12 in order to study the structure of the dispersed repetitive element (See Figure 6). The junction between the telomere-similar repeats and the flanking region lies at approximately nucleotide 90 on the sequence shown in Figure 6. The sequence of the flanking region has several notable features, most prominent are two AT-rich domains which contain numerous direct repeats. The first domain (nucleotides 319 through 798) is 80% AT and the second (AT content = 66%) domain lies between nucleotides 1742 and 2300. The complex organization of the larger direct repeats ($\geq$9 bp) is shown. Most of the repeat motifs are reiterated only twice, and are frequently superimposed. Other notable features of the flanking region sequence are a CT-rich region (82% CT, nucleotide 952–1092) and a large palindrome ((32 nucleotides, nucleotide 2363 to 2394). Only one small open reading frame was found (149 amino acids, residues 1162–1608). Searches of the databases with the peptide sequence predicted from the small open reading frame did not reveal any significant matches.

DISCUSSION

We have described the structure of an *A. thaliana* genomic clone, pAtT12, derived from the centromere region of chromosome 1. The majority of the pAtT12 insert is comprised of highly reiterated telomere-similar repetitive DNA. The telomere-similar repeats are approximately 500 bp in length and contain a simple-sequence region of variable size composed of degenerate telomere repeats. Simoens et al. (26) independently reported the sequence of a 500 bp *A. thaliana* repeat family which corresponds to the telomere-similar repetitive elements described here. Analysis of the genomic clone pAtT12 indicated that the telomere-similar repeats are arranged in tandem arrays which are frequently interspersed with related repeats which lack the telomere-similar simple-sequence domain.

As discussed in the introduction, there is precedence for telomere-like sequences located at the centromeres in certain animal cells. It has been proposed that centromeric telomere-similar sequences are remnants of telomeric repeat motifs left behind after Robertsonian fusion events (i.e., end-to-end fusion between two telocentric or acrocentric chromosomes creating a single metacentric chromosome) (8, 27). This mechanism can not easily account for the presence of the centromeric telomere-similar arrays in *A. thaliana* which contain non-telomeric sequences in addition to degenerate telomeric motifs. It is possible, however, that telomere-similar repeats are normally found in subtelomeric regions, and became internalized by a 'telomere fusion' event that left behind telomere-flanking DNAs if not the true telomeres themselves. An alternative hypothesis for the origin of the *A. thaliana* centromeric telomere-similar repetitive arrays stems from the observations of Simoens et al. (26). These authors have shown that the 500 bp telomere-similar repeat is closely related to the abundant *A. thaliana* 180 bp *Hin*dIII repeat described by Martinez-Zapater et al. (28), and suggested that the telomere-similar element arose by insertion of telomeric sequences into the 180 bp element. The telomere-similar repeats may then have spread to the centromere regions by recombination with related repeats already present in the centromeric heterochromatin. At present, we do not know if the *A. thaliana* telomere-similar repeats are found at non-centromeric locations.

In pAtT12, the telomere-similar repeats reside next to an unusual repetitive element which is reiterated approximately five times in the *A. thaliana* genome. We have demonstrated that one copy of this element is located in the central region of chromosome 5, in addition to the copy linked to centromere 1. The general organization of the moderately repetitive element is reminiscent of repeats found in the centromere region of *Schizosaccharomyces pombe* chromosomes, designated dg and dh (or K) (29, 30, 31, 32). Like the *A. thaliana* centromere-linked repeat, these *S. pombe* repeats contain AT-rich regions and domains with numerous short direct repeats. Short regions of nucleotide similarity were found in comparisons between the pAtT12 flanking region and the *S. pombe* repeats, but the significance of these similarities is difficult to assess. Although the *S. pombe* repeats have been shown to be important for centromere function, the DNA sequence elements within these repeats which are required for *S. pombe* centromere have not been established (34). We are currently constructing a fine structure physical and genetic map of *A. thaliana* centromere 1 in order to characterize plant centromere DNA sequences, and we hope to determine if the sequence characterized here plays any role in centromere structure or function.

The description of the structure and localization of repeated DNAs should aid current efforts to develop an integrated physical/genetic/cytological *A. thaliana* map as a resource and model for study of higher plants. It also demonstrates the utility of RFLP mapping techniques for chromosomal localization of repeated DNA sequences in organisms, such as *A. thaliana*, where cytology and *in situ* hybridization techniques are poorly developed.

ACKNOWLEDGEMENTS

We are especially grateful to Hong-gil Nam and Bart den Boer for their efforts in establishing the *A. thaliana* RFLP mapping project. We also express our thanks to: Dan Voytas for help with sequencing protocols and critical reading of the manuscript; Brian Seed for preparing oligonucleotide sequencing primers; Brian Hauge for communicating information; Jérôme Giraudat for help with RFLP mapping; and Bill D.B. Loos for computer mapping analysis. This work was supported by a grant from Hoechst AG to Massachusetts General Hospital.

Note added in proof

Maluszynska and Heslop-Harrison have recently demonstrated by *in situ* hybridization techniques that the abundant *A.thaliana* 180 bp *Hind*III tandem hybridizes to all five centromere regions. This result complements our localization of related tandem repeats to the centromere region of chromosome 1. [Maluszynska,J. and Heslop-Harrison,J.S. (1991) Localization of tandemly repeated DNA sequences in Arabidopsis thaliana. *Plant Journal*, in press].

REFERENCES

1. Zakian, V.A. (1989) *Annu. Rev. Genet.* **23**, 579–604.
2. Walmsley, R.W., Chan, C.S.M., Tye, B.-K., and Petes, T.D. (1984) *Nature* **310**, 157–160.
3. Pace, T., Ponzi, M., Dore, E., and Frontali, C. (1987) *Mol. Biochem. Parasitol.* **24**, 193–202.
4. Dore, E., Pace, T., Ponzi, M., Picci, L., and Frontali, C. (1990) *Mol. Cell. Biol.* **10**, 2423–2427.
5. De Lange, T., Kooter, J.M., Michels, P.A.M., and Borst, P. (1983) *Nucl. Acids Res.* **11**, 8149–8163.
6. Fry, K., and Salser, W. (1977) *Cell* **12**, 1069–1084.
7. Southern, E.M. (1970) *Nature* **227**, 794–798.
8. Allshire, R.C., Gosden, J.R., Cross, S.H., Cranston, G., Rout, D., Suguwara, N., Szostak, J.W., Fantes, P.A., and Hastie, N.D. (1988) *Nature* **332**, 656–659.
9. Meyne, J., Baker, R.J., Hobart, H.H., Hsu, T.C., Ryder, O.A., Ward, O.G., Wiley, J.E., Wurster-Hill, D.H., Yates, T.L., Moyzis, R.K. (1990) *Chromosoma* **99**, 3–10.
10. Levinson, A., Silver, D., and Seed, B. (1984) *J. Mol. Appl. Genet.* **2**, 507–517.
11. Viera, J., and Messing, J. (1982) *Gene* **19**, 259–268.
12. Yanisch-Perron, C., Viera, J., and Messing, J. (1985) *Gene* **33**, 103–109.
13. Richards, E.J., and Ausubel, F.M. (1988) *Cell* **53**,127–136.
14. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (1987) Current Protocols In Molecular Biology. John Wiley & Sons, New York.
15. Olszewski, N.E., Martin, F.B., and Ausubel, F.M. (1988) *Nucl. Acids Res.* **16**, 10765–10782.
16. Church, G.M., and Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1991–1995.
17. Grunstein, M., and Hogness, D.S. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 3961–3965.
18. Boseley, P.G., Moss, T., and Birnsteil, M.L. (1980) *Meth. Enzymol.* **65**, 478–494.
19. Nam, H.-G., Giraudat, J., den Boer, B., Moonan, F., Loos, W.D.B., Hauge, B.M., and Goodman, H.M. (1989) *The Plant Cell* **1**, 699–705.
20. Lander, E.S., Green, P., Abrahamson, J., Barlow, A., Daly, M.J., Lincoln, S.E., and Newberg, L. (1987) *Genomics* **1**, 174–181.
21. John, B. (1988) Heterochromatin: Molecular and Structural Aspects. Cambridge University Press, Cambridge, pp.1–147.
22. Schweizer, D., Ambros, P., Grundler, P., and Varga, F. (1987) *Arabid. Inf. Serv.* **25**, 27–34.
23. Cheng, C.-l., Dewdney, J., Nam, H.-g., den Boer, B.G.W., and Goodman, H.M. (1988) *EMBO J.* **7**, 3309–3314.
24. Koornneef, M. (1983) *Genetica* **62**, 33–40.
25. Koornneef, M., and Van der Veen, J.H. (1983) *Genetica* **61**, 41–46.
26. Simoens, C.R., Gielen, J., Van Montagu, M, and Inze, D. (1988) *Nucl. Acids Res.* **16**, 6753–6766.
27. Holmquist, G.P., and Dancis, B. (1979) *Proc. Natl. Acad. Sci USA* **76**, 4566–4570.
28. Martinez-Zapater, J.M., Estelle, M.A., and Somerville, C.R. (1986) *Mol. Gen. Genet.* **204**, 417–423.
29. Nakaseko, Y. Adachi, Y., Funahashi, S., Niwa, O., and Yanagida, M. (1986) *EMBO J.* **5**, 1011–1021.
30. Nakaseko, Y., Kinoshita, N., and Yanagida, M. (1987) *Nucl. Acids Res.* **15**, 4705–4715.
31. Clarke, L., Amstutz, H., Fishel, B., and Carbon, J. (1986) *Proc. Natl. Acad. Sci USA* **83**, 8253–8257.
32. Fishel, B., Amstutz, H., Baum, M., Carbon, J., and Clarke, L. (1988) *Mol. Cell. Biol.* **8**,754–763.
33. Koornneef, M. (1987) Genetic Maps. Cold Spring Harbor Laboratory, Cold Spring Harbor.
34. Clarke, L., and Baum, M.P. (1990) *Mol. Cell. Biol.* **10**, 1863–1872.