Nucleic Acids Research, 1996, Vol. 24, No. 15 **3017–3022**

Identification and distribution of seven classes of middle-repetitive DNA in the *Arabidopsis thaliana* genome

Helen L. Thompson[+], Renate Schmidt[§] and Caroline Dean*

John Innes Centre, Norwich Research Park, Colney Lane, Norwich NR4 7UH, UK

Received March 19, 1996; Revised and Accepted June 26, 1996

EMBL accession nos[¶]

ABSTRACT

In order to analyse further the genomic distribution of repetitive sequences in the *Arabidopsis* genome, we have identified and characterized seven novel repetitive sequences. Analysis of genomic representation, genomic location and DNA sequence divided the seven repeated sequences into two classes. The first was represented by three cosmid subclones (182A, 74A, 191A) carrying sequences that hybridised to up to 20 genomic fragments and showed sequence homology to the genes, *Arabidopsis CCR2*, *Arabidopsis MYB* and to various ATP-binding transport proteins. These multigene families mapped to various positions within the genome, as judged by hybridization to YAC clones constituting the *Arabidopsis* physical map. The second class was represented by four cosmid subclones (106B, 164A, 163A, 278A) that hybridised to between 20 and 300 genomic fragments. One of these, 106B, is a diverged, partial copy of the LTR of the *Arabidopsis* retrotransposon *Athila*. The other three sequences showed no homology to known genes or proteins. The distribution of these sequences on chromosome 4 was analysed and sequences hybridizing to 106B, 164A and 163A were found exclusively at the centromeric region of this chromosome. Their detailed arrangement at the centromeric region of chromosome 4, relative to other repeated sequence families and single copy sequences, was determined.

INTRODUCTION

Many large plant genomes have a high content of repetitive DNA, both tandem and dispersed. Tandemly repeated sequences are generally found at the centromeres and telomeres (1,2) and are often associated with constitutively condensed heterochromatin (3). Retrotransposon families form a major component of dispersed repetitive DNA (4) and can occur at very high copy numbers in some genomes. For example, the *BIS*-1 retroelement constitutes ~5% of the barley genome (5) and similar examples have been shown in the wheat and lily genomes (6).

In contrast with this picture, the genome of *Arabidopsis thaliana* is relatively small (~100 Mb, reviewed in 7) and has a low repetitive DNA content (~25%). Since *Arabidopsis* is an opportunistic wild plant, it is argued that there has been strong selective pressure for a short generation time resulting in reduced cell cycle time and genome size. If this is the case then any repetitive DNA that has been maintained may be functionally significant.

Of the repetitive DNA in the *Arabidopsis* nuclear genome, ~10% is highly repetitive and a further 10% middle repetitive (7). Three tandemly repeated sequences constituting ~2% of the genome have been characterised, a 180 bp *Hin*dIII repeat, a related 500 bp *Hin*dIII repeat and a 160 bp repeat (8–10). *In situ* hybridisation experiments demonstrated that the 180 bp tandemly repeated sequences co-localised with the heterochromatin surrounding the centromeres on all five chromosome pairs (11,12). Schmidt *et al*. (13) have recently shown that at least six other classes of repeated sequences flank these tandemly repeated sequences around the centromere on chromosome 4. The telomeric repeats have also been defined, each telomere carrying ~350 copies of a 7 bp tandemly repeated sequence (14).

The major component (7–8% of the genome) of the middle repetitive fraction of the *Arabidopsis* nuclear genome is the rDNA, localised within the nucleolus organizing regions on chromosomes 2 and 4. The rest of the middle repetitive DNA is made up of dispersed repeat elements. If these averaged 1–2 kb in length and were distributed randomly, then ~600 dispersed repeats would be distributed, one per 125 kb (15). This long interspersion pattern distinguishes *Arabidopsis* from large plant genomes such as maize where ~ 80% of the genome consists of repetitive sequences (16) which are found interspersed with blocks of middle repetitive and low copy sequences (e.g. the *Adh1* gene; 17,18). There has been little work addressing the identification of these dispersed repeated sequences in the *Arabidopsis* genome. The only two characterised retrotransposon families in *Arabidopsis* are the *Ta* family (19,20) and the *Athila* family (21,22) which occur in 15 and ~150 copies respectively. Most gene families so far characterized in *Arabidopsis* have relatively

*To whom correspondence should be addressed

Present addresses: [+]Molecular Genetics Laboratory, Instituto Giannina Gaslini, Largo G. Gaslini 5, 16148 Genova-Quarto, Genova, Italy and [§]Max-Delbrück-Laboratorium in der MPG, Carl-von-Linne-Weg 10, 50829 Köln, Germany

[¶]X92080, X92081, X93606–X93608, X93610 and X93611

few members, however, there are some exceptions, for example, the β-tubulin gene family which has nine members (23).

To characterise repetitive DNA within the *Arabidopsis* genome further, 300 cosmids were examined for the presence of novel repeated sequences and the results are described here. The sequences identified were analysed to determine their representation in the genome, what proportion were genic, whether they were clustered in the genome and whether they were associated with specific chromosomal areas.

MATERIALS AND METHODS

Cosmid and plant DNA preparation and analysis

A five genome equivalent cosmid library was constructed (C. Lister and C. Dean, unpublished) by cloning size fractionated *Sau*3A partially digested Columbia genomic DNA into the *Bam*HI polylinker site of the pLAFR3 vector (24). The average insert size of the library was 25 kb (I. Bancroft and C. Dean, personal communication). It has been used extensively in the generation of cosmid contigs for use in the EC genome sequencing project and these experiments have shown that the majority of inserts faithfully represent genomic DNA. Cosmid DNA was prepared using a method adapted from (25) and (26).

Plant genomic DNA was prepared from the Columbia ecotype using the protocol from (27).

Hybridisation analysis

DNA was blotted and fixed onto Hybond-N nylon membrane following the recommended protocol (Amersham). Columbia genomic DNA (100 ng) labelled with [α-^{32}P]dCTP by random primer extension, was used to probe Southern blots. Fragments used as probes were gel purified away from vector sequences twice using a Qiaex gel extraction kit (Qiagen). All hybridisations were performed at 65°C and washing was carried out at 65°C in 0.1× SSC, 0.1% SDS.

YAC colony and Southern blot hybridisation protocols were performed as described in (28).

DNA manipulations

Standard protocols were used to subclone the restriction fragments into the Bluescript SK$^+$ (Stratagene) vector as outlined in (29). Ligations were transformed by electroporation into SURE recombination deficient, electrocompetent cells (Stratagene).

DNA sequencing

Double-stranded DNA sequencing was performed on 2 µg of DNA from each subclone using a Pharmacia T7 sequencing kit. Sequencing and PCR primers were made using a Pharmacia LKB Gene Assembler Plus oligo-machine. Sequencing data was compiled and analysed using the UWGCG sequencing packages (30). Database searches were performed using FASTA and BLASTX programmes.

Determination of copy number

Copy number was determined by comparing the hybridisation intensity (on a Southern blot) of dilutions of each subclone insert (equivalent to 1–100 copies assuming a genome size of 100 Mb) with 1 µg of Columbia genomic DNA digested with the relevant enzyme to the subclone.

RESULTS

Identification of cosmids containing repetitive DNA sequences

To identify uncharacterised repetitive DNA sequences, 300 randomly selected cosmids from a library containing DNA from the *Arabidopsis* ecotype Columbia were hybridised with Columbia total genomic DNA. Seventy-two cosmids showed a stronger hybridisation signal than cosmids containing low copy sequences and therefore were considered to contain repetitive DNA. The average cosmid size was ~ 25 kb and so the 72 cosmids represent ~1.8% of the genome.

The 72 cosmids were hybridised with probes containing the total chloroplast genome and coding and intergenic regions of the rDNA and 5S rDNA. The cosmids were also probed with three characterised repetitive DNA sequences; a high copy tandemly repeated 180 bp sequence in pAL1 (8), a 500 bp repeat (9) and the minisatellite arabms1 (31). Of the initial 72 cosmids, 65 hybridised to known repetitive sequences (27 to chloroplast DNA, 22 to the 180/500 bp tandem repeat family, 13 to rDNA and three to arabms1) and were not studied further. They may, however, carry novel repetitive sequences linked to those previously characterized.

Table 1. Summary of the size, copy number and genomic location of the subclones

Subclone	Restriction fragment cloned	Size (kb)	Homology	Location	Copy number
164A	*Hin*dIII	0.9		centromeric on chr 4, also mapping to other centromeric regions	~150
106B	*Eco*RI	0.4		centromeric on chr 4, also mapping to other centromeric regions	~300
278A	*Eco*RI/*Bgl*II	1.3		centromeric but not on the present chr 4 contigs	~20
163A	*Hin*dIII	0.45		centromeric on chr 4, also mapping to other centromeric regions	~90
74A	*Eco*RI/*Bgl*II	0.7	*MYB*	dispersed on chr 4 and 5 and unmapped on other chromosomes	~7 (for the MYB homology)
191A	*Eco*RI	0.7	*WHITE*	unknown	~18
182A	*Bgl*II	1.8	*CCR2*	chr 2	~20

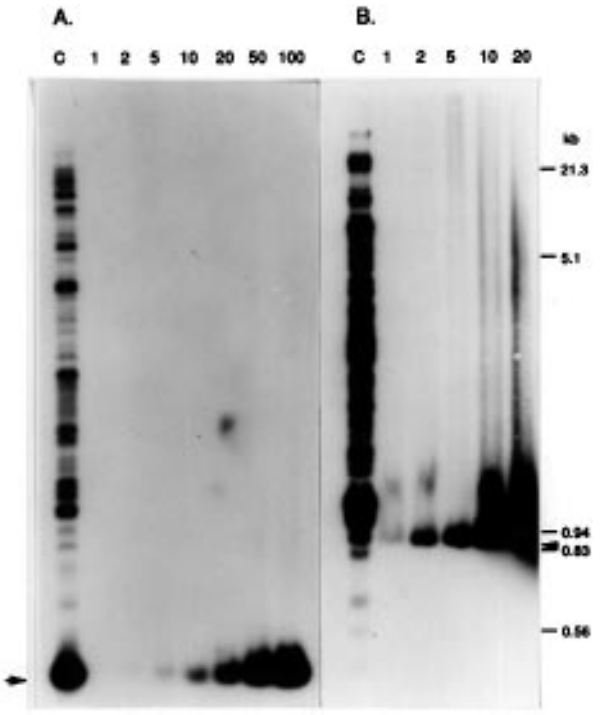

Figure 1. Genomic copy number estimates for subclones 106B and 164A. Each subclone was digested to excise the insert and serial dilutions prepared. 1–100 genomic copies in relation to 1 µg of *Arabidopsis* genomic DNA were determined and loaded on a 0.8% gel. **(A)** One µg Columbia genomic DNA digested with *Eco*RI, (C), alongside 1–100 equivalent copies of 106B. The arrow shows the position of 100 equivalent copies of 106B in the genomic digest. (**B**) One µg of Columbia genomic DNA digested with *Hin*dIII, (C), alongside 1–20 copies of the 164A. The position of five equivalent copies of 164A in the genomic digest is marked by the arrow.

The cloned sequences are present in between 20 and 300 copies in the *Arabidopsis* genome

The smallest restriction fragment within each of the seven remaining cosmids that hybridised strongly to *Arabidopsis* genomic DNA was subcloned (summarized in Table 1). These were called 164A, 106B, 278A, 163A, 74A, 191A and 182A. Only part of a large repetitive element or one of multiple repetitive elements from each cosmid would be analysed using this approach. The copy number of the cross-hybridizing sequences was estimated by determining the number of restriction fragments in the *Arabidopsis* genome hybridising to each subclone. Reconstruction experiments showed that 164A hybridised to ~150 fragments (five of which were the same size as the subclone) and 106B to ~300 (~100 of which were the same size as the subclone) (Fig. 1). 278A hybridised to ~20 fragments, 163A to ~90 fragments and 191A and 182A hybridised to ~20 restriction fragments each (data not shown).

Chromosomal location of the seven repeated sequences

The repeated sequences were mapped onto the *Arabidopsis* genome by taking advantage of the physical map of chromosome 4 generated in YAC clones (13). All the available subclones were hybridized to colony filters carrying YAC clones from the CIC library (32) and positive hybridization was confirmed by Southern blot analysis. The subclones, 164A and 106B, hybridized to a proportion of the YAC clones to which the respective whole cosmids had hybridized. The cosmids, CC164 and CC106, had been found to hybridize to YAC clones that exclusively mapped to the centromeric region of chromosome 4 and cross-hybridising sequences were not detected elsewhere on the chromosome (13). The hybridization profile of each subclone was analysed with respect to the relative overlap of the YAC clones between markers mi233 and mi87 and around HY4/nga8. These overlaps created intervals varying between 10 and 300 kb to which the repetitive sequences could be mapped (Fig. 2). 164A hybridised to multiple intervals located across the whole centromeric region including the interval mapping in between the 180 bp repeat loci that contained the single-copy marker mi87. 163A was found associated with 164A but was detected only on one side of the 180 bp repeat loci, 106B related sequences were more clustered around the 180 bp repeat loci, being closely associated with sequences cross-hybridizing to 164A in this area. There was one 106B locus located someway within the short arm of chromosome 4, close to the marker mi233.

The association of sequences carried in 164A and 106B with YAC clones, containing the 180 bp repeat array and not mapping to chromosome 4, suggests that 164A and 106B sequences are localised around centromeres of at least some of the other chromosomes. 278A hybridised to a subset of the 164A/106B YAC clones. Thus sequences cross-hybridizing to 278A localise at one or more centromeric regions. 278A cross-hybridizing sequences were not found on the YAC contigs available for the centromeric region of chromosome 4.

The remaining subclones 182A, 74A and 191A did not localise to the centromere on chromosome 4 and were found at dispersed locations throughout the genome. 182A hybridised to two CIC YAC clones, CIC5B4 and CIC7E2. Both co-ordinates have been previously mapped to chromosome 2 (33). 191A hybridised to one YAC clone CIC11G5 whose position is unknown at present. For 74A ~70 corresponding YAC clones were identified in YAC colony hybridisations. Some of these co-ordinates map to chromosomes 4 and 5. Up to four loci for sequences related to this cosmid have been identified on chromosome 4 and up to 12 potential locations have been found dispersed along the length of chromosome 5 (R. Schmidt and C. Dean, unpublished data).

Sequence analysis of the repeated sequences in the subclones

The DNA sequences were analysed for distinct features such as tandem arrays, microsatellite sequences and direct and inverted repeat motifs. Four clones, 163A, 106B, 278A and 164A, contained direct and inverted repeat motifs of between 9 and 24 bp in one or two copies.

Homology searches to all available DNA and peptide sequence databases were performed. 182A (EMBL accession no. X93610) was found to have significant DNA homology to the *CCR2* gene (93% in the 276 bp overlap). This gene encodes a glycine-rich protein, contains an RNA-binding motif containing and is a member of a small gene family with approximately six members (34). 182A hybridised to two CIC YAC clones positioned on the physical map of chromosome 2.

The sequence analysis of 74A (EMBL accession no. X93607) revealed two distinctive features, a $(GA)_{38}$ microsatellite repeat

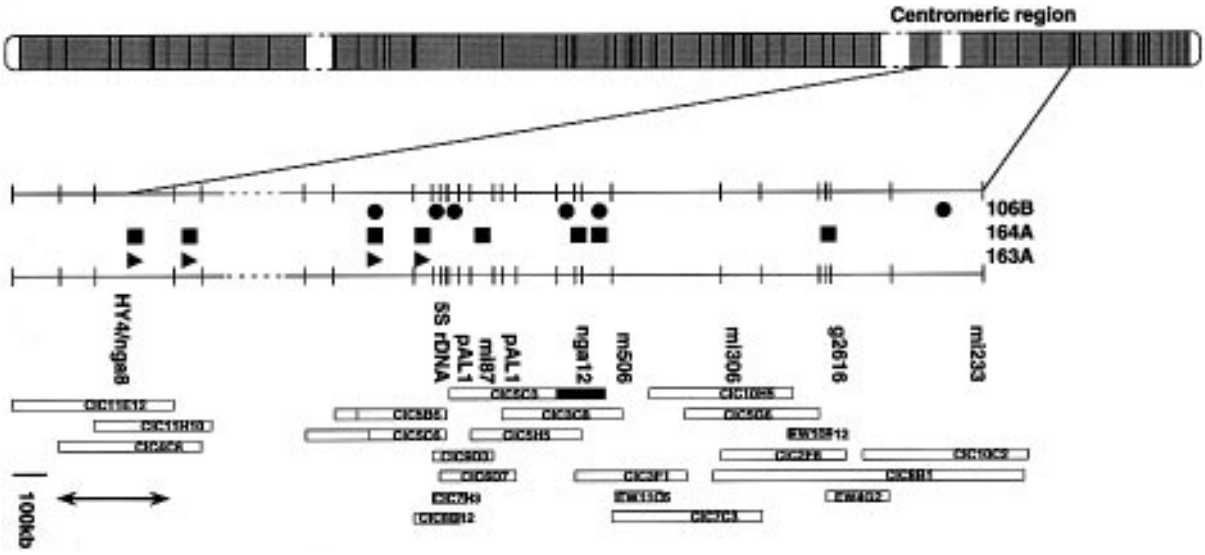

Figure 2. Distribution of repeated sequences in the centromeric region of chromosome 4. The YAC contig map of chromosome 4 is shown at the left of the figure. The four YAC contigs are represented as grey boxes. The YAC contigs, constituting the physical map in the centromeric region of chromosome 4, are shown to the right. The regions shown as broken lines are gaps between the contigs where no YAC clones have as yet been identified. All YAC clones which have been used for the distribution analysis of the repeated DNA sequences are represented as vertical rectangles. The YAC clones are drawn to scale and any chimaeric regions are shaded black. The locations of the RFLP markers, the 180 bp *Hin*dIII and 5S rDNA repeat arrays are indicated relative to the YAC clones. The overlap of the YAC clones defines the various intervals which can be assessed for the presence of sequences related to the repetitive subclones. The boundaries of the intervals are indicated by the horizontal lines. Hybridization of a subclone to a particular interval is shown by a black symbol. 106B related sequences are indicated as filled circles, while 164A and 163A related sequences are represented by squares and triangles respectively.

in the middle of the clone and a region showing homology to the MYB class of transcriptional regulators. The *MYB* genes in *Arabidopsis* and *Zea mays* contain two ~60 amino acid imperfect repeats at the N-terminus, thought to be necessary for DNA binding. The predicted peptide sequence of 74A shows strong homology to these conserved repeats in the *Arabidopsis MYB* gene *GL1* (55% in the first repeat and 75% in the overlap region of the second repeat), a gene required for trichome differentiation in *Arabidopsis* (35).

$(GA)_n$ microsatellites, like the dinucleotide repeat found in 74A, are one of the least abundant microsatellites in *Arabidopsis* (36). They are highly polymorphic and estimated to occur once every 244 kb. To determine whether the GA microsatellite or the *MYB* homologous sequences contributed to the highly repetitive pattern of this clone, PCR primers were designed to separately amplify the microsatellite and the *MYB* DNA binding region. The resulting PCR products were used to probe Columbia genomic Southern blots. The *MYB* PCR probe hybridised to seven *Eco*RI fragments whereas the microsatellite hybridised to ~100 fragments (data not shown) demonstrating that the $(GA)_{38}$ microsatellite accounted for the hybridisation pattern of 74A. This was further confirmed by analyzing the YAC clones corresponding to 74A using Southern blot analysis. The *MYB* homologous sequences hybridised to only two of the 16 putative loci for 74A, which had been anchored on chromosomes 4 and 5.

191A (EMBL accession no. X93606) showed significant nucleotide sequence homology to the *Arabidopsis* EST at 8149 DNA, 85% in the 69 bp overlap. It also showed predicted peptide homology to various ATP-binding transport proteins, including the *Drosophila* white protein (sp:P10090). Comparison with other ATP-binding transport proteins identified a putative ATP binding site. 191A hybridized to ~18 fragments in Columbia genomic DNA but to only one CIC YAC clone, the position of which on the physical map is unknown. This would suggest that all the associated sequences to 191A are clustered within one region.

The subclone 106B (EMBL accession no. X93611) was found to have significant DNA sequence homology to the long terminal repeats (LTR) of the recently characterised *Athila* retrotransposon (EMBL accession no. X81801) (21,22). The clone showed 70% DNA sequence homology to the left LTR and 69% homology to the right LTR. The level of homology suggests that the sequence in clone 106B represents a diverged copy of an LTR from this group of retroelements.

The remaining three subclones, 163A (EMBL accession no. X93608), 164A (EMBL accession no. X92080) and 278A (EMBL accession no. X92081) had no significant homology to any database entries or to each other.

DISCUSSION

We describe here the identification of novel middle repetitive DNA sequences in the *A.thaliana* ecotype Columbia genome. This fraction of the *Arabidopsis* genome is relatively uncharacterised. Analysis of genomic representation, chromosomal location and DNA and predicted peptide sequence divided the repeated sequences into two classes summarised in Table 1. The first class represented gene sequences, with up to 20 associated genomic fragments. From the proportion of the repeats that fell into this class, it is likely that a high proportion of the dispersed repeated sequences in the *Arabidopsis* genome are made up of large gene families.

The second class of sequences had no homology to known genes. They were present in between ~20 and 300 copies and were clustered in the genome. No highly repetitive DNA

sequences other than those previously characterised by (8) and (9) were detected in the initial cosmid screening experiments.

One of the subclones, 106B, was found to be a diverged, partial sequence of the LTR of the *Athila* retroelement. It is possible that the cosmid CC106 contains an intact copy of a related element. The *Athila* retrotransposon is found in up to 150 copies in the *Arabidopsis* ecotype C24 (21,22). It has been estimated that up to 200 retroelements may be present in the Arabidopsis genome (37), thus *Athila* is likely to constitute a significant number of these. Since the estimated copy number of 106B is ~300, this may indicate that all the LTRs are associated with intact elements (one at each end). Diverged copies of the *Athila* element have been found and are likely to represent inactive remnants of old integration events (22). One *Athila* element was found integrated within the 180 bp repeat tandem array and from a distribution analysis of *Athila* within YAC and λ libraries Pelissier *et al.* (22) concluded that the elements were concentrated at the centromeric regions. Our analysis of the distribution of repeated sequences, including 106B, across the centromeric region of the physical map of chromosome 4 shows that sequences related to 106B and hence the LTRs of *Athila* are clustered in this region particularly around the 180 bp repeat arrays. We have also detected sequences related to 106B up to 1.5 Mb away from the 180 bp repeat arrays.

Only a relatively small fraction of the genome, ~7.5% (300 cosmids of average size 25 kb), was analysed in this study and it is clear from other studies that a large number of other middle repetitive elements, not detected in this study, also co-localize with the paracentromeric heterochromatin in chromosome 4. These include sequences present on two RFLP markers, m456 and mi167, and two YAC end-probes, CIC5C6LE and CIC6D7RE (13). Preliminary analysis of cosmid CC106 revealed that it hybridized to five additional intervals, as compared with 106B, across the centromeric region of chromosome 4. This suggests the presence of middle repetitive sequences, other than 106B, in CC106. Thus a lot more analysis is required before we will be close to fully characterizing the repetitive DNA of the *Arabidopsis* genome.

If the YAC clones hybridizing to the 180 bp tandem repeat loci do represent at least the core of the centromere (13), then it is interesting to compare the arrangement of the repetitive sequences relative with other characterized centromeric regions. A recent study characterising the centromere of *Drosophila* minichromosome *Dp1187* found that a central core of ~220 kb containing complex DNA, of single copy and middle repetitive sequences, was essential for centromere function (38). A region of ~200 kb, either side of the essential core, containing highly repetitive DNA sequences was also required for completely normal inheritance. This region was believed to be involved in sister chromatid adhesion and indirectly assist in kinetochore formation. The centromeric region of the fission yeast *Schizosaccharomyces pombe* has also been characterised and displays some of the features of the *Drosophila* centromere. A central core of 6.8 kb of single copy DNA is directly flanked by moderately repetitive sequences one of which, the K-type repeat contains a region 2.1 kb long which is critical for centromere formation. The central region is flanked by ~100 kb of satellite DNA (39). The examples of the *Drosophila* and *S.pombe* centromeres allow us to draw a comparison with the centromeric region of chromosome 4. 164A cross-hybridizing sequences are found in the ~100 kb interval between the 180 bp satellite repeat loci. The RFLP marker mi87, that maps to this interval, hybridises to a single genomic fragment so at least some of this DNA is not repetitive. This region might equate to the essential centromere core. Families of other repetitive sequences are also present up to 1.5 Mb away from the 180 bp tandem repeat loci on each chromosome arm and these may represent the non-essential functional flanking DNA. It will be interesting to see if the asymmetry of the 163A hybridizing sequences is functionally significant.

The repeated sequences comprising the heterochromatin around centromeres are generally species-specific (40). However, some centromeric repeats have been shown to be chromosome-specific, for example the pBcKB4 and pBoKB1 repeats in *Brassica* chromosomes (41). The repetitive element in 278A would appear to be an example of this class of repeat. It is associated with 164A, 106B and 180 bp repeat loci sequences but is not present on the YAC contigs currently available for the centromeric region of chromosome 4. The availability of the physical maps for the other chromosomes, in combination with *in situ* hybridisation analysis, will allow the distribution of repeats such as these to be analysed and to further examine the organization of *Arabidopsis* centromeres.

ACKNOWLEDGEMENTS

The authors would like to thank Joanne West, Karina Love and Zoë Lenehan for their contributions to the physical map, and Dr Clare Lister for the cosmid library and her advice. We should also like to thank Dr Keith Mitchelson for the arabms1 probe and Dr Richard Macknight for proofreading. This work was funded by a BBSRC grant (PG208/PG0608) to C.D. and a BBSRC studentship to H.L.T.

REFERENCES

1 Bedbrook,J.R., Jones,J., O'Dell,M., Thompson,R.D. and Flavell,R.B. (1980) *Cell*, **19**, 545–560.
2 Hutchinson,J. and Lonsdale,D. (1982) *Heredity*, **48**, 371–376.
3 John,B. and Miklos,G.L.G. (1979) *Int. Rev. Cytol.*, **58**, 1–114.
4 Hirochika,H. and Hirochika,R. (1993) *Jpn. J. Genet.*, **68**, 35–46.
5 Moore,G., Cheung,W., Schwarzacher,T. and Flavell,R.B. (1991) *Genomics*, **10**, 469–476.
6 Smyth,D.R. (1991) *Chromos.*, **100**, 355–359.
7 Meyerowitz,E.M. (1992) In Koncz,C., Chua,N-H. and Schell,J. (eds), *Methods in Arabidopsis Research*. World Scientific, Singapore, pp. 100–118.
8 Martinez-Zapater,J.M., Estelle,M.A. and Somerville,C.R. (1986) *Mol. Gen. Genet.*, **204**, 417–423.
9 Simoens,C.R., Gielen,J., Van Montagu,M. and Inze,D. (1988) *Nucleic Acids Res.*, **16**, 6753–6766.
10 Richards,E.J., Goodman,H.M. and Ausubel,F.M. (1991) *Nucleic Acids Res.*, **19**, 3351–3357.
11 Maluszynska,J. and Heslop-Harrison,J.S. (1991) *Plant J.*, **1**, 159–166.
12 Bauwens,S., Van Oostveldt,P., Engler,G. and Van Montagu,M. (1991) *Chromos.*, **101**, 41–48.
13 Schmidt,R., West,J., Love,K., Lenehan,Z., Lister,C., Thompson,H., Bouchez,D. and Dean,C. (1995) *Science*, **270**, 480–483.
14 Richards,E.J. and Ausubel,F.M. (1988) *Cell*, **53**, 127–136.
15 Pruitt,R.E. and Meyerowitz,E.M. (1986) *J. Mol. Biol*, **187**, 169–183.
16 Hake,S. and Walbot,V. (1980) *Chromos.*, **79**, 251–270.
17 Springer,P.S., Edwards,K.J. and Bennetzen,J.L. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 863–867.
18 Avramova,Z., SanMiguel,P., Georgieva,E. and Bennetzen,J.L. (1995) *Plant Cell*, **7**, 1667–1680.
19 Voytas,D.F., Konieczny,A., Cummings,M.P. and Ausubel,F.M. (1990) *Genetics*, **126**, 713–721.
20 Konieczny,A., Voytas,D.F., Cummings,M.P. and Ausubel,F.M. (1991) *Genetics*, **127**, 801–809.

21 Pelissier,T., Tutois,S., Deragon,J.M., Tourmente,S., Genestier,S. and Picard,G. (1995) *Plant Mol. Biol.*, **29**, 441–452.
22 Pelissier,T., Tutois,S., Tourmente,S., Deragon,J.M. and Picard,G. (1996) *Genetica*, **97**, 141–151.
23 Snustad,D.P., Haas,N.A., Kopczak,S.D. and Silflow,C.D. (1992) *Plant Cell*, **4**, 549–556.
24 Staskawicz,B., Dahlbeck,D., Keen,N. and Napoli,C. (1987) *J. Bacteriol.*, **169**, 5789–5794.
25 Birnboim,H.C. and Doly,J. (1979) *Nucleic Acids Res.*, **7**, 1513–1523.
26 Ish-Horowiscz,D. and Burke,J.F. (1981) *Nucleic Acids Res.*, **9**, 2989.
27 Tai,T. and Tanksley,S. (1991) *Plant Mol. Biol. Rep.*, **8**, 297–303.
28 Schmidt,R., Cnops,G., Bancroft,I. and Dean,C. (1992) *Aust. J. Plant Physiol.*, **19**, 341–351.
29 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
30 Devereux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387–395.
31 Mitchelson,K.R., Luczak,B. and Porter,A.J.R. (1994) *J. Exp. Bot.*, **45**, 1655–1659.
32 Creusot,F., Fouilloux,E., Dron,M., Lafleuriel,J., Picard,G., Billault,A., Le Paslier,D., Cohen,D., Chaboute,M.E., Durr,A., Fleck,J., Gigot,C., Camilleri,C., Bellini,C., Caboche,M. and Bouchez,D. (1995) *Plant J.*, **8**, 763–770.
33 Zachgo,E.A., Wang,M.L., Dewdney,J., Bouchez,D., Camilleri,C., Belmonte,S., Huang,L., Dolan,M. and Goodman,H.M. (1996) *Genome Res.*, **6**, 19–25.
34 Carpenter,C.D., Kreps,J.A. and Simon,A.E. (1994) *Plant Physiol.*, **104**, 1015–1025.
35 Oppenheimer,D.G., Herman,P.L., Sivakumaran,S., Esch,J. and Marks,M.D. (1991) *Cell*, **67**, 483–493.
36 Bell,C.J. and Ecker,J.R. (1994) *Genomics*, **19**, 137–144.
37 Hirochika,H., Fukuchi,A. and Kikuchi,F. (1992) *Mol. Gen. Genet.*, **233**, 209–216.
38 Murphy,T.D. and Karpen,G.H. (1995) *Cell*, **82**, 599–609.
39 Baum,M., Ngan,V.K. and Clarke,L. (1994) *Mol. Biol. Cell*, **5**, 747–761.
40 Shaw,D.D. (1994) *Trends Ecol. Evol.*, **9**, 170–175.
41 Harrison,G.E. and Heslop-Harrison,J.S. (1995) *Theor. Appl. Genet.*, **90**, 157–165.