Repetitive DNA sequences located in the terminal portion of the *Caenorhabditis elegans* chromosomes

Giuseppina Cangiano and Adriana La Volpe*
CNR International Institute of Genetics and Biophysics, 10 Via Marconi, 80125 Naples, Italy

Received December 14, 1992; Revised and Accepted January 28, 1993

ABSTRACT

We describe the distribution along the chromosomes of *Caenorhabditis elegans* of two repetitive DNA families, RcS5 and Cerep3 and interstitial telomeric sequences. Both families show, among other interesting features, a preferential location in the terminal 30% of the chromosomes. It is known that in these regions of the genome the frequency of recombination is much higher than in the central portion, genes are rarer and sequences important for chromosome disjunction may lie.

INTRODUCTION

As a result of the intensive and multifaceted analysis of genomes, it is becoming increasingly manifest that most chromosomal DNA performs functions other than those involved in pure coding activity. Chromosomes need to attach to the scaffold and to the nuclear matrix, to duplicate with fidelity and to segregate efficiently. Homologous and non homologous recombination is active in somatic tissues and germ line cells. Specific sequences along the chromosome guarantee the binding of specific matrix proteins, proper DNA duplication and segregation during the cell cycle, and probably the enhancement or the repression of recombination. *C.elegans* is a well known model system for developmental genetics and genomic studies. The physical map is close to completion and a pilot project to sequence the *C.elegans* third chromosome has been undertaken (1). Comparison between the physical map and the genetic map (2) has confirmed a non uniform distribution of crossovers along the chromosomes. Crossovers are relatively rare in the central part of the chromosomes, in the so called 'gene clusters' (where most mutants map), but their frequency is highly enhanced in the pro-terminal region of the chromosomes, where genetic loci defined by mutations are rarer (fig 1). A survey of the distribution of the coding regions in the genome (by cDNA analysis) has shown that although variation in the frequency of recombination accounts for the clustering of most genes in the centre of the chromosomes on the genetic map, nevertheless coding sequences are physically somewhat rarer in the so called 'chromosomal arms' (3).

Cytological and genetic studies indicate that *C.elegans* centromeres are diffuse along the chromosomes (holocentric chromosomes, 4, 5, 6). However, when the chromosomes segregate in meiosis they align parallel to the spindle axis such that one end or the other points towards the spindle pole (Donna Albertson personal communication).

In an effort to resolve the chromosome structural and functional organization in *C.elegans*, we have recently described the distribution of clusters of satellite-like DNA sequences and short interspersed elements (7). We observed a uniform distribution of five repetitive DNA families along the chromosome. Here we identify and describe a family of repetitive DNA elements, RcS5, which in contrast is preferentially located at the pro-terminal portion of the chromosomes.

We further show that another repetitive DNA family previously described, Cerep3 (8) is similarly distributed. Associated with these two repetitive DNA families we find a high concentration of interstitial telomeric repeats.

MATERIALS AND METHODS

N2-JW-H, W-PA-1H, W-PA-2H, DH424, RW7000, RC301, TR403, CB4507, CB4555, CB4853, CB4854, CB4855, CB4856, CB4857 are wild type, independently isolated *C.elegans* strains, CB1392 is a nuc001 mutant, they were all obtained from the *Caenorhabditis* Genetics Centre. DNA from *C.remanei*, PG-3, *P.redivivus*, *C.briggsae*, PG-2, *C.marpasi*, all closely related nematode species, was a gift from L.W.Nawrocki. All the strains were raised at 18°C. Routine strain maintenance was as described by Sulston and Hodgkin 1988 (9).

*C.elegans* DNA purification from liquid cultures was performed according to Sulston and Hodgkin (9).

The cosmid library, from which the initial representative of the RcS5 family (RcS5 # 1) was obtained, was a gift from Silvana Gargano. It was made by inserting genomic DNA, partially digested with Sau3A, into the BamHI site of tetra-cos vector.

A subset of clones (250), corresponding to about 10% of the *C.elegans* genome, was screened by hybridisation with labelled total genomic DNA as probe. We obtained 12 positive clones, two of which contain ribosomal genes, three contain the transposable element Tc1, two contain elements belonging to the same repetitive DNA family, the others represent individual elements belonging to independent repetitive DNA families. All

* To whom correspondence should be addressed

+X69085, X69082, X68553 and X68498
the other cosmid clones derive from a library also constructed from a partial SauIIIA digest, using the vector Lorist6 (10).

The genomic DNA library in yeast artificial chromosomes was that made for the C.elegans mapping project and has been previously described (11, 12, 13, 7). The vector was pYAC4 and the Saccharomyces cerevisiae host strain was AB1380 (MATa, y+· ara-3, trp-1, ade2-1, can-1 100, lys 2-1, his-5).

Southern transfer was performed as described by Southern (14) with the following modifications: the DNA (1 µg of total genomic DNA per lane) was transferred to Hybond-N membrane and immobilized by UV treatment. Hybridisations of libraries and Southern nylon filters were performed in 6xSPC (20xSPC = 2M NaCl, 0.6M Na2HPO4, 0.02M EDTA pH 6.2), 0.9% Sarkosyl, 10% dextran sulphate and 100µg/ml single strand DNA carrier at 65°C for 6-16 hours. We used approximately 107 cpm of oligolabelled DNA probes prepared as described by Feinberg and Vogelstein (15, 16) with a specific activity of greater than 109 cpm/µg per hybridisation bag. The telomeric probe was a single strand 48mer containing eight copies the TTAGGC repeat identified as the Ascaris telomeric repeat (17). It was purchased from Eurogentech (Belgium) and end-labelled with T4 polynucleotides kinase and [γ32P]-ATP. Hybridisation was performed in the mixture described above at 60°C.

Following the hybridisation, filters were washed four times either in 1xSPC, 1%SDS at 50°C or in 2xSPC, 1%SDS at 42°C (relaxed conditions) and air dried.

Removal of probe: In order to reuse the filters, after each independent hybridisation, probes were removed from filters by repeating the following steps twice: 10 min room temp. in 0.2 M NaOH, 0.5 M NaCl, 10 min room temp. in 0.5 M Tris pH 7.4, 0.5M NaCl and 10 min room temp. in 1xSSPE (20xSSPE = 3M NaCl, 0.2M NaH2PO4, H2O, 0.02M EDTA, pH 7.4). Efficiency of removal was monitored by re-exposing used filters before reuse.

Cosmids and plasmids used as probes were maintained in E.coli rec A strain HB101 (F−, hsd S20, (r plugs), recA13, ara-14, proA2, lacY1, galK2, rpsL20 (Smr), xyL5, mil-1, supE44, h-). Selecting for kanamycin or ampicillin resistance. The strain was checked periodically for the recA phenotype.

All RcS5 subclones were constructed in pGEM3/4 or pUC18/19. Cerep3 plasmid was a gift from Scott Emmons.

Sequencing was performed according to Sanger and Coulson (18). The complete sequences obtained during this work have been sent to the EMBL Data Library, and only relevant portions are shown. Accession numbers are indicated in the legend of fig. 5.

Computer programs used in the physical mapping analysis are described in Sulston et al. (19). Sequence data were analysed with the GCG (20) software package.

RESULTS

Physical distribution of RcS5 and Cerep3 elements in the genome

The first element belonging to the RcS5 family, called RcS5 # 1, was isolated from a genomic DNA library using as probe total genomic C.elegans DNA; clones containing DNA from the genomic repetitive fraction give much stronger hybridisation signals (7). The repetitive portion of the original cosmid was then subcloned and used as probe on two types of indexed library grids on nylon membrane (12, 7). The first kind of grid represents a subset of cosmids already positioned on the map, the 1934 cosmids cover about 80% of the genome in a non-overlapping manner. The second kind of grid is a set of overlapping YAC clones representing almost 95% of the genome. We demonstrated elsewhere that all the cloned members of a repetitive family can be positioned in this way on the already mapped portion of the genome (7). The use of two indexed libraries carried by a prokaryotic and an eukaryotic host may overcome problems of instability (in one host or the other) of some DNA sequences.

In figure 2 we report the physical distribution of RcS5 elements on the physical map of the five autosomes and on the X chromosome. RcS5 elements appear to be preferentially located outside the 'gene clusters'. Another family with a similar distribution is the Cerep3 family. This latter family has been identified by Felsestein and Emmons in 1988 and comprises elements about one kb long, some of which have been shown to affect replication and segregation of plasmids in Saccharomyces cerevisiae. The distribution of Cerep3 elements along the chromosomes is also shown in figure 2.

Strains and species comparison

We have compared the patterns obtained by digesting the genomic DNA of several independently isolated C.elegans wild type strains with the restriction endonuclease Eco RI and hybridising the Southern transfer with RcS5 or Cerep3 probes (figure 3). The RcS5 probe is an Eco RI fragment from the RcS5(ZK1091) element containing only RcS5 sequences (see figure 5 for details), the Cerep3 probe is an EcoRI-BamHI fragment from pCe17(Cerep3.1) containing a Cerep3 entire element (Felsestein and Emmons 1988). Some of the C.elegans strains used in this experiment have a high copy number of the transposable element Tc1 and among these some contain a mutator activity that results in an increase in the rate of transposition ( 'hopper' genetic background, 21). We observe that the RcS5 probe generally recognises the same number of bands of similar size, suggesting
Figure 2. Physical distribution of RsS5 and Cerep3 elements along the physical map of the five autosomes and of the chromosome X. The distribution has been obtained by hybridising two types of indexed library grids on nylon membranes: a subset of 1934 cosmids already positioned on the map covering about 80% of the genome in a non-overlapping manner and a set of overlapping YAC clones representing almost 95% of the genome (12, 7). The probes used in the screening are: a HindIII fragment containing the entire RsS5 #1 element and a 728bp flanking DNA region (on the left side of the element illustrated in figure 5, ac. no.: X68553), an EcoRI-BamHI fragment from pCel7(Cerep3.1) containing a Cerep3 entire element and about 200bp of flanking sequences on both sides (8), and a synthetic oligonucleotide 48bp in length containing 8 repeats of the putative telomeric hexamer (TTAGGC). Hybridisations of libraries with RsS5 and Cerep3 probes were performed in 6×SPC, 0.9% Sarkosyl, 10% dextran sulphate, 100μg/ml single strand DNA carrier at 65°C, hybridization with the single stranded telomeric probe was performed in the same mix at 60°C. In all cases, filters were washed four times in 1×SPC, 1%SDS at 50°C. The RsS5 elements and the Cerep3 elements are reported as small vertical rods in the second and third row respectively. Interstitial telomeric repeats are represented by round bold dots under the physical map.

an average constant copy number, although RFLPs are detectable especially in CB4854 and CB4856 strains. These are not ‘hopper’ strains, but show a low Tc1 copy number and mobility. Those RFLPs are therefore not likely to be due to accidental transposition of an active Tc1 next to a RsS5 element. The patterns obtained on the same strains with the Cerep3 probe appear more variable, though not so well resolved due to the high number of Cerep3 loci in genome.

We have also tested for possible cross-hybridisations between our repetitive probes and DNAs from other free living nematode species (figure 4, upper panels). We also have on the same gel three C. elegans N2 derived lines. No detectable RFLPs were observed among N2 derived lines. Using an internal fragment of the RsS5(ZK1091) element, common to all the members of the family, we did not detect any cross-species hybridisation under relaxed stringency conditions (hybridisation at 65°C in 6×SPC,
Figure 3. Southern transfer and hybridisation with RsS5 and Cerep3 specific probes of independently isolated C. elegans wild type strains. DNA was digested with EcoRI and run on a 1% agarose gel in Tris-Acetate buffer, following the hybridisation in 6xSPC, 0.9% Sarkosyl, 10% dextran sulphate at 65°C, filters were washed four times in 1xSPC, 1% SDS at 50°C. From left to right W-PA-1H(Ca,USA), W-PA-2H(Ca,USA), DH424 (Ca,USA), RW7000* (France), RC301* (Germany), TR403* (WI, USA), CB4507 (Ca,USA), CB4555* (Ca,USA), CB4853 (Ca,USA), CB4854 (Ca,USA), CB4855 (Ca,USA), CB4856 (Haway), CB4857 (Ca,USA). Strains with an asterisk have a high copy number of Tel elements. Origin of the strains is indicated in parenthesis. Upper panel: the filter was hybridised with an RsS5 specific probe. Lower panel: a second copy of the same gel transferred on filter was hybridised with a Cerep3 specific probe.

Figure 4. Southern transfer and hybridisation with RsS5 and Cerep3 specific probes of different closely related nematode species. DNA in all cases was digested with EcoRI and run on a 1% agarose gel in Tris-Acetate buffer, following the hybridisation in 6xSPC, 0.9% Sarkosyl, 10% dextran sulphate at 65°C, filters were washed four times in 2xSPC, 1% SDS at 42°C. The gel in the upper panel contains, left to right, C. remanet, PG-3, C. briggsae, PG-2, C. noma, followed by three N2 derived strains A, B, C, an empty lane and P. redivivus DNA. The gel was hybridised with an RsS5 specific probe, (Eco RI-Sac I fragment from the RsS5(ZK1091) element containing only RsS5 sequences, see figure 5 for details) left, and a Cerep3 specific probe (EcoRI-BamHI from pCe17(Cerep3.1) containing a Cerep3 entire element, 8) right. In the lower panel, we have a different gel hybridised with two subclones of an RsS5 element. The gel contains, left to right, C. remanet, PG-3, P. redivivus, C. briggsae, PG-2, C. noma. Saccharomyces cerevisiae DNA as negative control, two empty lanes and CB1392. The filter on the left was hybridised with the RsS5 internal fragment described above, the one on the right with a SacI—HindIII fragment spanning the junction between the RsS5(ZK1091) element and its righ flanking sequence (accession number: X69082).

Comparison within RsS5 elements and within their flanking sequences

We have determined the DNA sequence of four independent members of the RsS5 family. Two, RsS5(T25H8) on chromosome I and RsS5(ZK1091) on chromosome V, represent complete elements, 1.3kb in length and composed of two terminal inverted repeats of about 250bp separated by an intervening sequence around 800bp in length. Within the intervening sequence of the two complete elements we find a duplication of over 150 bp. RsS5(ZK1091) has also a duplication of 69bp within the inverted repeat at the 3' end. The other two, RsS5 #1, unmapped, and RsS5(T17H5) on chromosome II, are truncated elements missing the 3' inverted repeat and part of the intervening sequence (see figure 5). Sequence identity between elements ranges between 86% and 93%. We did not find any open reading frame inside the elements making it unlikely that any of these could be an active member of an unhindered transposon family such as Tc1.

An interesting feature of the flanking regions of the RsS5
elements is the presence of telomeric repeats. The inverted repeats of each element ends with the sequence AGGGTT (in opposite orientations), and adjacent to each element either at the 5’ end (in RsS5(T17H5)) or at the 3’ end (in RsS5(T25H8), RsS5(ZK1091) and RsS5 #1) we find a number of repetitions of the putative telomeric hexamer of *C. elegans* TTAGGC (as defined by C. Wicky by means of Bal31 experiments, personal communication). This finding accounts for the results obtained in the Southern experiments with the panel of the free living nematode species: in fact, the probe that only recognises *C. elegans* DNA was an Eco RI—Sac I fragment from the RsS5(ZK1091) element containing only RsS5 sequence, while the other probe, detecting sequences in all the nematode species, was a SacI—HindIII fragment encompassing the 3’ end of RsS5(ZK1091) and entering the flanking region for 98 bp. We also hybridised the filter shown in figure 4, bottom panel, with a oligomer containing 8 tandem repetitions of the TTAGGC hexamer, and we detected the same pattern obtained with the SacI—HindIII probe including the flanking sequence of RsS5(ZK1091) (data not shown). Evidence for the occurrence of intrachromosomal telomeric repeats in *C. elegans* has been obtained in other laboratories (Donna Albertson, Yuko Kozono, and Chantal Wicky, personal communication). Sequence comparison between this 98bp fragment flanking RsS5(ZK1091) and the sequence of the cloned *Ascaris* telomere (17) is shown in fig. 6 (top).

**D**istribution of interstitial telomeric repeats in the *C. elegans* genome

Transposable elements flanked by telomere-like repeats have been isolated in several protozoa (22, 23). Could the occurrence of intrachromosomal telomeric hexamers be closely linked to RsS5 elements? We mapped the interstitial telomeric repeats making use of the same indexed filters already used for the physical mapping of the RsS5 and Cerep3 families. The filters were hybridised with a terminal labelled 48mer containing eight tandem repeats of the TTAGGC hexamer. Interstitial telomeric repeats once again turned out to be clustered in the terminal 30% of the chromosomes (see bold dots in figure 2). They appear to be more abundant than the repetitive elements belonging to RsS5 and Cerep3 families, although they are often found associated either with Cerep3 or with RsS5 or with both, in fact cosmids containing RsS5 and/or Cerep3 often hybridise also with the telomeric probe. There is a positive correlation between enhancement of crossing overs in this part of the genome and frequency of interstitial telomeric repeats (figure 7).
account for the increase in the recombination frequency in the personal communication). This observation suggests that the as well as in be necessary for proper disjunction in C.elegans have recombination abnormalities suggesting that exchange may portions. C.elegans chromosomes are holocentric (also found in Ascaris, might invading the last 30% of the chromosomes in C. elegans for example, telomeric repeats moved recombinogenic in other organisms (26, 27, 28, 29). In 7). Telomeric repeats have been shown to be highly enhancement of recombination along the chromosomes (see figure between abundance of interstitial telomeric repeats, and other organisms (24, 25). We observe a positive correlation towards the spindle pole while the other guarantees the maintenance of the bivalent (probably by terminalization of the chiasmata). Notably, it is not always the same end to be used for each function (Albertson personal communication). Some mutations, that affect meiotic disjunction, have been shown to have recombination abnormalities suggesting that exchange may be necessary for proper disjunction in C.elegans as well as in other nematodes, in some insects and plants. Kinetocores are not visible during meiosis I, the two main functions of the centromeres (attachment to the spindle and holding the bivalent together) being adopted by the chromosome termini (Albertson personal communication). One end of the chromosome is oriented towards the spindle pole while the other guarantees the maintenance of the bivalent (probably by terminalization of the chiasmata). Notably, it is not always the same end to be used for each function (Albertson personal communication). Some mutations, that affect meiotic disjunction, have been shown to have recombination abnormalities suggesting that exchange may be necessary for proper disjunction in C.elegans as well as in other organisms (24, 25). We observe a positive correlation between abundance of interstitial telomeric repeats, and enhancement of recombination along the chromosomes (see figure 7). Telomeric repeats have been shown to be highly recombinogenic in other organisms (26, 27, 28, 29). In Saccharomyces cerevisiae, for example, telomeric repeats moved inside the chromosome create a hot-spot of recombination (Petes, personal communication). This observation suggests that the occurrence of stretches of degenerate telomeric hexamers, invading the last 30% of the chromosomes in C.elegans, might account for the increase in the recombination frequency in the ‘chromosome arms’. In this perspective these DNA elements might play an important functional role in chromosome disjunction favouring the formation of sub-terminal chiasmata during meiosis. Middle repetitive DNA families, containing several TTAGGGC repeating units, have been detected in Ascaris lumbricoides, where these DNA elements are completely eliminated from the somatic genome during the process of chromatin elimination (17), and in the filarial species Loa Loa (30). Klion et al. (30) claim that such repeated DNA is a species-specific diagnostic probe. Under our stringency conditions as little as 8 tandem duplications of the TTAGGC repeat cross-hybridise with all nematode species analysed. Sequence similarity between C.elegans interstitial telomeric repeats and Loa Loa repetitive DNA sequence is shown in figure 6 (bottom).

We also observed that interstitial telomeric repeats are associated in C.elegans with bona fide repetitive DNA elements. The RsS5 and Cerep3 families are also preferentially located in the proterminal last 30% of the chromosomes. Elements belonging to the Cerep3 family have been shown to direct replication and correct segregation in Saccharomyces cerevisiae, stabilising yeast plasmids during mitosis and meiosis and also lowering the plasmid copy number (8). Taking into account the non-uniform distribution of Cerep3 elements in C.elegans, it is suggestive to imagine that they might play a role in segregation in this organism.

Internally located telomeric sequences in the germ-line chromosomes of Tetrahymena mark the ends of transposon-like elements (31). The authors suggest that transposition of these elements involves a linear form and that telomeric repeats are added to the ends of these free linear elements. Transposons bounded by C4A4 telomeric repeats were also found in Oxytricha fallax (22) and shown to excise precisely during macronuclear development (23). RsS5 might be a vestigial mobile element of the kind observed in protozoa and the cause of interspersing of telomeric repeats in the C.elegans ‘chromosome arms’. Once moved inside the chromosome, telomeric repeats might have been furtherly amplified by gene conversion, unequal-crossing over or DNA polymerase slippage (32, 33, 34). Alternatively, the amplification of telomeric repeats in the chromosome arms has independently occurred and the putative transposon has acquired the hexamer TTAGGC as preferential target site of integration.
ACKNOWLEDGEMENTS

We are grateful to Clara Frontali, Alan Coulson and John Pulitzer for stimulating discussions and helpful suggestions concerning the manuscript. We thank Donna Albertson, and Chantal Wicky for suggestions and for sharing informations prior publication. We thank Scott Emmons for the Cerep3 plasmid, and L.W.Nawrocki for non-C.elegans nematode DNAs. Some nematode strains used in this work were provided by the Caenorhabditis Genetics Center, which is funded by the NIH National Center for Research Resources. This work was supported in part by Commission of European Communities Contract SC1*-CT91-0633, CNR Progetto Speciale Bioinformatica and CNR Target Project Biotechnology and Bioinstrumentation.

REFERENCES