Rte-1, a retrotransposon-like element in *Caenorhabditis elegans*

Sandra Youngman, Henri G.A.M. van Luenen, Ronald H.A. Plasterk*

The Netherlands Cancer Institute, Division of Molecular Biology, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

Received 12 September 1995; revised version received 12 December 1995

Abstract **We have characterized a retrotransposon-like element (Rte-1) in *C. elegans*. It was identified while we were sequencing the *pim related kinase-1* (*prk-1*) gene. The element is 3,298 bp long and flanked by a 200 bp direct repeat. 95 bp of the direct repeat are present in the coding region of *prk-1*. Rte-1 contains an open reading frame, in the opposite orientation of *prk-1*, potentially encoding 625 amino acids, with similarity to reverse transcriptases. The element is most similar to members of the non-LTR group of retrotransposable elements. There is weak homology of the predicted amino acid sequence of Rte-1 to several reverse transcriptase-like genes identified by the *C. elegans* genome sequencing consortium, suggesting that there may be a large family of these elements. Southern blots indicate that there are approximately 10–15 additional Rte-1 elements in the *C. elegans* Bristol N2 genome and a similar number is found in the genomes of two other geographically distinct strains. The insertion pattern of Rte-1 is polymorphic between these strains.**

Key words: Retrotransposon; Non-LTR element; Reverse transcriptase; pim-related kinase; *Caenorhabditis elegans*

1. Introduction

Eukaryotic genomes contain a variety of dispersed repetitive sequences, some of which are transposable elements. It is thought that transposable elements exist in the genome of every eukaryote [1]. The dispersion of transposable elements is due to transposition of these elements within the genome. To date more than six different DNA-transposons have been identified in *Caenorhabditis elegans*, with Tc1 being the best studied [2–8]. Recently, also three *gypsy*/Ty3-like retrotransposons have been described in *C. elegans* [9]. In addition, LTR-like sequences have been described in the nematodes *Ascaris lumbricoides* and *Panagrellus redivivus* [10,11]. However, until this report no non-LTR retrotransposon-like element had been described in nematodes.

Retrotransposons contain an open reading frame (ORF) bearing homology to retroviral reverse transcriptases and they transpose through an RNA intermediate [12]. The total amino acid similarity between these diverse elements is quite low, but they contain seven domains which are well conserved [13]. These highly conserved residues are diagnostic for the identification and alignment of these elements. There are two classes of retrotransposons: the long terminal repeat (LTR) containing elements including *copia* and *gypsy* of *Drosophila melanogaster*, and Ty1 and Ty3 of *Saccharomyces cerevisiae* [14–16], and the non-LTR elements. Examples of this latter class include the long interspersed nucleotide elements (LINEs) of mammals, the I element of *D. melanogaster* and ingi of *Trypanosoma brucei* [17–19]. These elements generate target duplications at their insertion sites [20], and they contain a characteristic A-rich region or poly(A) tail at their 3′-end. In addition these elements have two overlapping ORFs that span most of the element. Amino acid sequence analysis of the longer ORF (ORF2) shows significant homology with reverse transcriptases of retroviruses [21], while the shorter ORF (ORF1) often has homology with the nucleic acid binding domain of the *gag* genes. There are also some elements which contain only one long ORF with homology to reverse transcriptase-like sequences [22,23]. Recently, the non-LTR retrotransposon from *Bombyx mori* (R2Bm) has been shown to encode an endonuclease that also contains reverse transcriptase activity [24,25]. The retrotransposon inserts into a specific sequence of the 28S ribosomal RNA gene. The R2Bm protein binds the 3′-end of R2Bm RNA, associates with its integration site in the target DNA and makes a single stranded break. The R2Bm protein is able to use the 3′ hydroxyl group, generated by the nick in one strand of the DNA, to prime reverse transcription of its RNA template.

While sequencing the *prk-1* gene we found it to contain a large insertion which had none of the characteristics of a *C. elegans* intron, such as a high A/T content and consensus intron border sequences. The insertion, which we named Rte-1, contains a large ORF which potentially encodes a protein of 625 amino acids. Searching the databases with the ORF revealed that the sequence has homology with non-LTR retrotransposons. We have investigated the distribution of Rte-1 in different *C. elegans* strains and two other *Caenorhabditis* species.

2. Materials and methods

2.1. Nematodes

Cultures of *C. elegans* were grown and maintained as previously described by Brenner [26]. The *C. elegans* strains Bristol N2, TR403 and RW4000 in addition to *C. briggsae* and *C. remanei* strains were obtained from the Caenorhabditis Genetics Center in Columbia, MO.

2.2. Sequencing

Subclones containing the *prk-1* gene were obtained from cosmid C06E8. Sequencing was by standard procedures using the Sequenase kit from USB and primers to the sequenced DNA. The sequence of the trans-spliced leader (SL2) primer used to obtain the 5′ end of *prk-1* cDNA is 5′ATCTCGGAGGGTTTAATTACCCAAG. Sequences were analyzed using the blast algorithm by NCBI [27,28].

2.3. Southern blots

Genomic DNA was isolated as described by Sulston and Hodgkin [29]. Digestion, Southern blotting and hybridization were performed as described by Sambrook et al. [30]. Probes used in the hybridization were obtained by PCR. Probe 1 was generated from a 3.0 kb *Eco*RI–*Xba*I fragment (C06E8RX3.0). This subclone contains no *prk-1* sequence. The probe (2,180 bp) generated from this subclone using the M13 reverse sequencing primer (including the *Eco*RI site) and primer 4113

*Corresponding author. Fax: (31) (20) 669 1383.
E-mail: rplas@ron.nki.nl

0014-5793/96/$12.00 © 1996 Federation of European Biochemical Societies. All rights reserved.
SSDI 0014-5793(95)01525-6

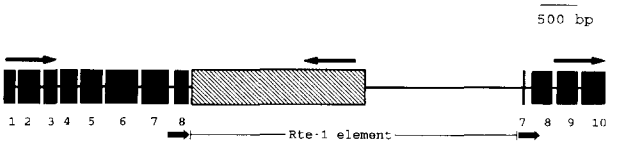


Fig. 1. Schematic diagram showing the location of Rte-1 within *prk-1*. The exons of *prk-1* are indicated by the black boxes. The 200 bp direct repeat (indicated by the arrows flanking the Rte-1 element) consists of the last 3 bp of exon 7, an intron and the first 95 bp of exon 8. The ORF of Rte-1 is shown by the shaded box. The arrows above the exons indicate the direction of transcription of the ORF of *prk-1* and Rte-1. The sequence of cDNA cm01c10 corresponds to nucleotide position 11,495–11,581, 11,629–11,720, 11,769–11,883, 11,931–12,090, 12,141–12,464, 12,519–12,722, 16,339–16,481, 16,536–16,674, and 16,723–17,133 in cosmid C06E8. Exons 1 and 2 correspond to nucleotide positions 11,308–11,377 and 11,424–11,581.

(5′TCTTTCACAAATGGTAGCGAC) represents the ORF of Rte-1. The direct repeat probe C06E8T284 is a 284 bp *Taq*I fragment containing the 200 bp direct repeat. PCR products were isolated from agarose gel prior to radiolabelling by random priming [31]. All Southern blots were washed to a stringency of 0.1 × SSC.

2.4. Accession numbers

The accession numbers for the genes and cosmids discussed in this paper are listed below. Cosmid C06E8, no. CELC06E8. cDNA cm11g11/cm11d12, no. CE11G11/CE11D12. Cosmid C18H2, no. CELC18H2. Cosmid C52A11, no. CEC52A11. Cosmid C07A9, no. CEC07A9. Gene C06E8.4, no. CELCO6E8_2. Gene F56C9.2, no. CELF56C9_8. Gene T07E3.1, no. CELT07E3_6. Gene F58A4.5, no. YMH5_CAEEL. Gene F40F12.2, no. S422831. Gene ZK1236.4, no. Y084_CAEEL. The Rte-1 element corresponds to nucleotide positions 16,216–12,920 in cosmid C06E8 and the ORF to nucleotide positions 14,829–12,956.

3. Results

3.1. Nucleotide sequence

Comparison of the *prk-1* cDNA sequence (cm01c10) and the corresponding genomic region (cosmid C06E8) revealed that there is a large intervening sequence without the consensus intron border sequences. Furthermore, this intervening sequence contains a large ORF in the opposite direction to *prk-1*. Both the cDNA and the cosmid library were generated from the same strain, Bristol N2. Screening the databases with the intervening sequence revealed it to have homology with non-LTR retrotransposons. The element has been named retrotransposon-like element-1 (Rte-1).

The genomic arrangement of *prk-1* and Rte-1 is shown schematically in Fig. 1. The arrangement and numbering of the

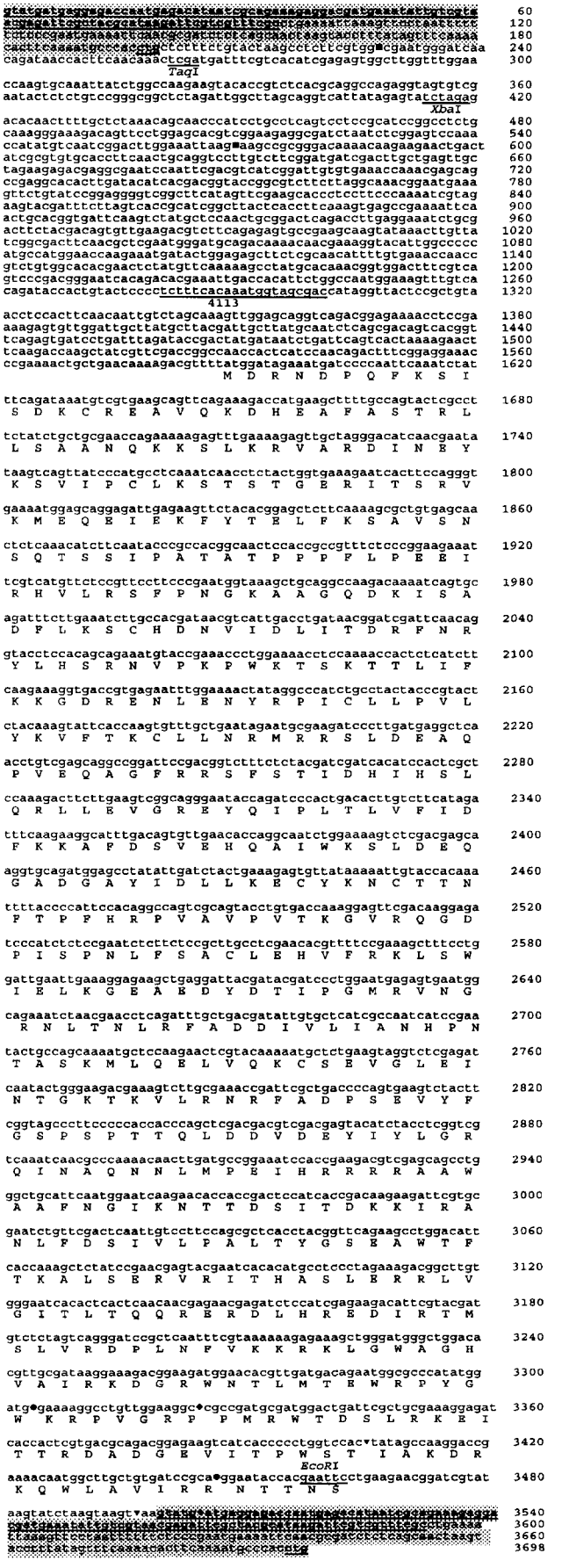


Fig. 2. Nucleotide sequence of Rte-1, including the direct repeats, and the predicted amino acid sequence of the ORF. The 3,698 nucleotides correspond to nucleotide positions 16,416–12,720 in cosmid C06E8. The 200 residues which are highlighted are the direct repeats and the bold, underlined residues are present in cDNA cm01c10. The *Eco*RI, *Xba*I and *Taq*I recognition sites and the primer used to generate the probes by PCR are also indicated. The 341 bp region, nucleotide position 227–568 bracketed with ■ is the region of homology with cDNA cm11d12/cm11g11. The 143 bp region, nucleotide positions 3304–3446, bracketed with ● is the region of homology with C07A9. The 180 bp region, nucleotide position 3324–3503, bracketed with ♦ is the region of homology with C18H2. The 91 bp region, nucleotide position 3405–3496, bracketed with ▼ is the region of homology with C52A11.

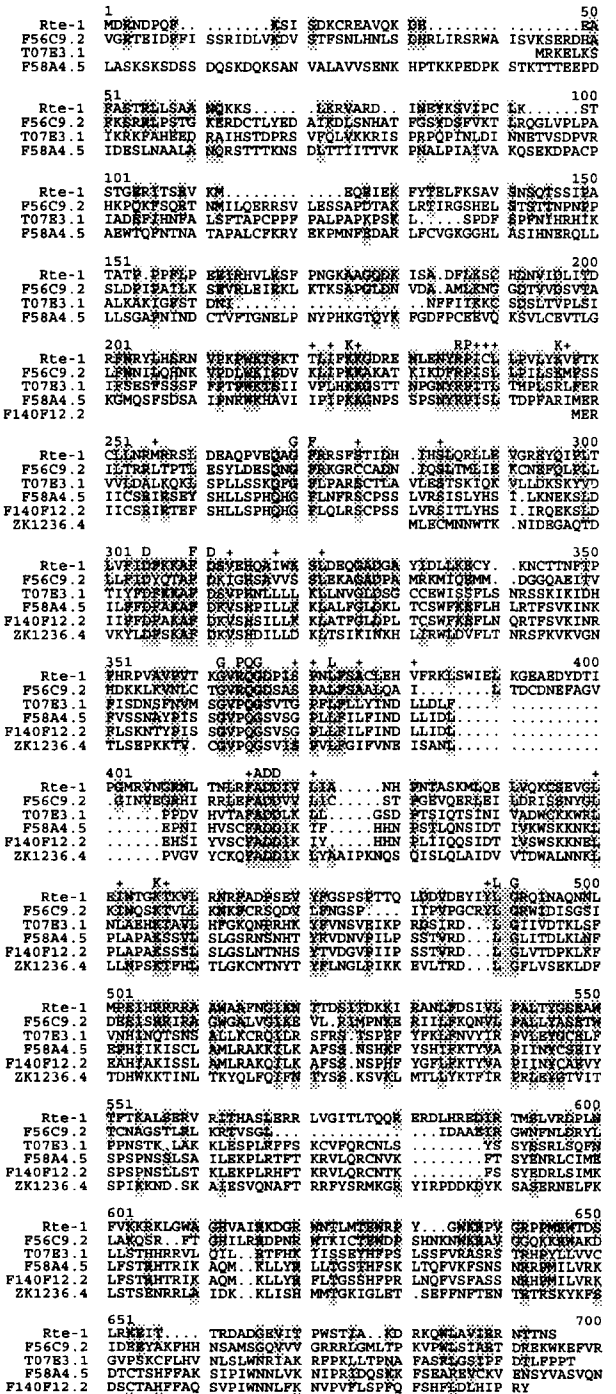
exons and introns of *prk-1* is based on the cDNA sequences and the consensus intron border sequences. Rte-1 is located between *prk-1* exons 7 and 8. The ends of Rte-1 are defined by a 200 bp direct repeat. The duplicated region consists of the last 3 bp of exon 7, an intron and 95 bp of exon 8. The complete nucleotide sequence of Rte-1 including the direct repeats is shown in Fig. 2.

After we had determined the sequence of *prk-1* and the Rte-1 element, it was subsequently determined independently by the genome sequence consortium [32]. We compared the 6 kb of sequence and found four differences. Reassessment of the sequence data showed that our readings were in error. The 5′ end of cDNA clone cm01c10 is located in exon 2. The remaining 5′ sequence of the *prk-1* transcript was determined using DNA obtained from a nested PCR on reverse transcribed total Bristol N2 RNA using primers directed against the trans-spliced leader (SL2) and gene specific sequences. The sequence agrees with the sequence of the consortium.

Screening the databases with the sequence of Rte-1 revealed significant homology to four very small independent loci in *C. elegans*. Rte-1 is homologous (98% identity) to the 337 nucleotides of cDNA cm11d12/cm11g11, a clone partially sequenced by the genome sequencing project. This region of homology is located at Rte-1 nucleotide position 227–568 which is upstream of the Rte-1 ORF (see Fig. 2). The cDNA cm11d12/cm11g11 has been mapped to chromosome 2 whereas Rte-1 is located on chromosome 3. The other three small DNA fragments which are homologous to Rte-1 have also been sequenced by the genome sequencing project. These three loci are located at a similar overlapping position at the 3′ end of the Rte-1 ORF and these regions are indicated in Fig. 2. These regions of homology are summarized as follows; cosmid C07A9 has a 143 bp fragment which is 76% homologous to Rte-1; a 180 bp fragment of cosmid C18H2 is 92% homologous to Rte-1; and cosmid C52A11 has a 91 bp fragment which is 96% homologous. Interestingly these small fragments in cosmids C07A9, C18H2 and C52A11 are very similar to one another. It is surprising that there are these short stretches of homology between Rte-1 and other *C. elegans* loci.

3.2. Amino acid sequence

The longest single ORF of Rte-1 potentially encodes a protein of 625 amino acids and is in the opposite orientation to the *prk-1* ORF. The ORF has the first ATG codon at position 1,588 (Fig. 2) and a second at position 1,805. However, both are a poor match to the optimum initiation sequence for eukaryotic protein translation [33]. Gene Finder takes the second ATG to be the beginning of the protein coding sequence, whereas we consider the first ATG as the beginning of protein translation. The ORF is most homologous to the *C. elegans* predicted gene F56C9.2 (similarity 53%, identity 33%), which the Gene Finder program predicts to be similar to reverse transcriptase. Using the first ATG as the start of protein translation, the two genes Rte-1 and F56C9.2 are of a similar length and there is some homology between the two genes in this first region. Both Rte-1 and F56C9.2 are homologous to the non-LTR group of retrotransposable elements. In addition Rte-1 also shows homology to a number of other predicted reverse transcriptase-like genes in *C. elegans* (T07E3.1, F58A4.5, F40F12.2 and ZK1236.4) (see Fig. 3). The regions of strongest



←

Fig. 3. Alignment of Rte-1, F56C9.2, T07E3.1, F58A4.5, F40F12.2 and ZK1236.4. The homology between Rte-1 and any of the other predicted *C. elegans* reverse transcriptase-like genes is high-lighted. The letters and plus (+) symbols on the top of the alignment are the largely unvaried and chemically similar amino acids identified in a study by Xiong and Eickbush [13]. The numbering refers to the predicted amino acids in Rte-1. Predicted gene F56C9.2 is longer than Rte-1, amino acid 1 of Rte-1 corresponds to amino acid 108 for F56C9.2, and continues for another 5 amino acids not shown on the figure. Predicted gene F58A4.5 is much longer than Rte-1 there are 451 amino acid prior to the beginning of the figure and 113 at the end.

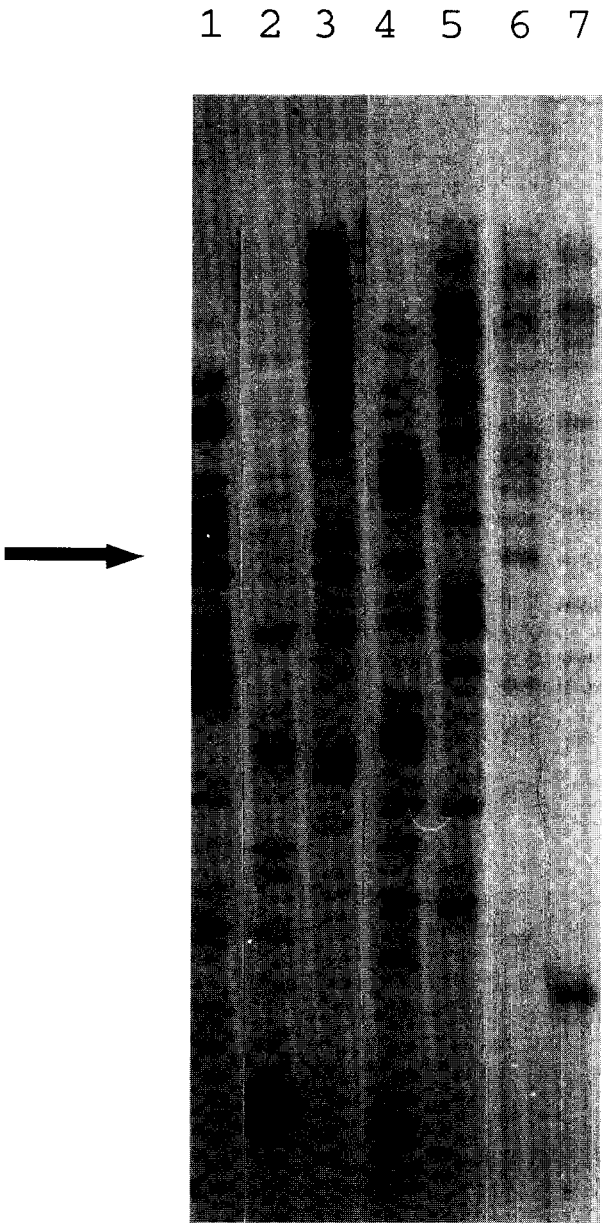


Fig. 4. Southern blot of 5 μg Bristol N2 DNA digested with the following enzymes (1) *EcoRI*, (2) *Hin*dIII, (3) *Bam*HI, (4) *Acc*I, (5) *Sac*I, (6) *Sal*I and (7) *Pst*I and hybridized with probe 1. The 4.5 kb *Eco*RI band that represents probe 1 is indicated with an arrow. This figure shows that Rte-1 is represented about 10–15 ×.

homology are at similar positions previously identified in a study by Xiong and Eickbush [21]. In a study of 13 group II mitochondrial introns and non-LTR retrotransposons they identified 43 unvaried and chemically similar amino acids present in at least 11 of the 13 elements [21]. In a latter study using more elements the results were essential unchanged [13].

3.3. Distribution of Rte-1

To determine the distribution of elements homologous to Rte-1, a probe was generated to be used on Southern blots. Probe 1 is a PCR fragment amplified from a subclone of C06E8 which contains no *prk-1* sequences. This PCR fragment contains the whole ORF plus 312 bp. The sequence of the probe is located on a 4.5 kb *Eco*RI fragment.

Southern blots containing Bristol N2 DNA digested with a number of different restriction enzymes were hybridized to determine whether Rte-1 is represented more than once in the genome. Fig. 4 shows the autoradiograph of the hybridization with probe 1. This shows that the element is repeated 10–15×. The intensities of the bands vary, indicating some variation in the homologies between Rte-1 and the other fragments. The 4.5 kb *Eco*RI fragment is indicated in Fig. 4; it is not the most intense band, the most intense band probably being a doublet. A Southern blot of genomic DNA from the closely related species *C. briggsae* and *C. remanei* shows only one band (Fig. 5). The intensity of the band in *C. briggsae* and *C. remanei* is decreased when compared with the bands in *C. elegans*. The size of this *Eco*RI fragment is approximately 4.5 kb. This is the size of the band expected if Rte-1 were the only copy of the element in *C. elegans* Bristol N2 DNA, located on cosmid C06E8. However, the sizes of the restriction fragments to which

Fig. 5. Southern blot of (1) *C. elegans* (N2), (2) *C. briggsae*, and (3) *C. remanei* DNA (5 μg) digested with *Eco*RI and hybridized with probe 1.

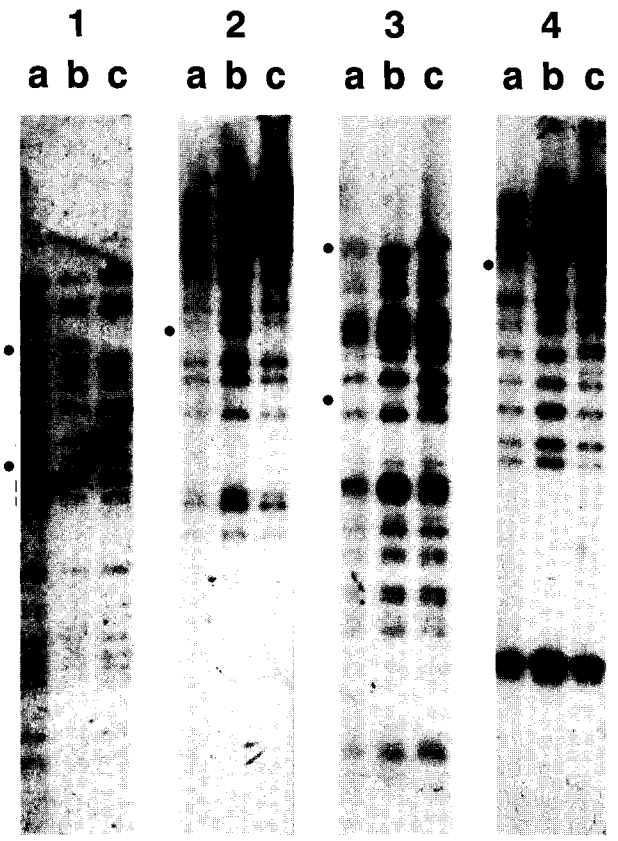


Fig. 6. Southern blot showing the banding patterns of (1) N2, (2) TR403 and (3) RW4000 DNA digested with the restriction enzymes (a) *Eco*RI, (b) *Bam*HI, (c) *Acc*I and (d) *Pst*I and hybridized with probe 1. The most prominent RFLPs have been indicated by dots.

probe 1 hybridizes, when the DNA is digested with additional enzymes is not consistent with the expected fragment sizes in cosmid C06E8 (data not shown). This suggests that the size of the *Eco*RI fragment detected by probe 1 in *C. briggsae* and *C. remanei* is a coincidence. We also compared the banding patterns of different *C. elegans* strains, N2 with RW7000 and TR403 (Fig. 6). These three strains were isolated in different parts of the world. Although all three strains have similar banding patterns, they contain differences which could indicate that the insertion pattern is polymorphic but could also be explained by other RFLPs unrelated to Rte-1.

3.4. Is the 200 bp direct repeat an integral part of Rte-1?

To determine whether the 200 bp repeat is a terminal repeat present at the ends of all the Rte-1-like sequences, we generated a probe containing the direct repeat from subclone C06E8T284. This subclone is a 284 bp *Taq*I fragment (Rte-1 nucleotide position −15 to 265). The fragment was hybridized to an *Eco*RI digest of Bristol N2 DNA. It hybridized to two *Eco*RI bands of 4.5 kb and 5 kb as expected and in addition to seven fainter bands shown in Fig. 7. This shows that the other 10–15 Rte-1 like elements have a different terminal sequence.

4. Discussion

4.1. Is Rte-1 a non-LTR element?

We have characterized Rte-1 located within the *prk-1* gene. This element shows strongest homology to the non-LTR group of retrotransposable elements. The ORF of Rte-1 contains the seven reverse transcriptase-like domains [13,21]. These seven reverse transcriptase domains are found in four classes of elements: non-LTR retrotransposable elements, group II mitochondrial introns, retroviruses and LTR retrotransposable elements and share approximately one-half of the conserved residues. Particularly diagnostic of non-LTR retrotransposable elements is the highly conserved Y/FXDD box, flanked by several hydrophobic residues (residues 415–418, Fig. 3). All non-LTR retrotransposable elements (including Rte-1) have an alanine at point X, while all members of the LTR branch have a hydrophobic residue at this position. Rte-1 is a member of the non-LTR retrotransposable elements, although it contains not all characteristic features of the non-LTR group of sequences. Non-LTR elements are usually organized with two overlapping ORFs whereas Rte-1 has only one ORF. Other non-LTR elements with a single ORF besides Rte-1 have been reported, for example R2Bm [22]. There is no A-rich region at the 3′ end of Rte-1, as would be expected for a non-LTR retrotransposon. However, R2Bm has an invariable tail of only four adenine residues [25]. Non-LTR containing retrotransposons have similar cysteine-motifs associated with the 3′ portions of their ORF1 and ORF2. These motifs are generally conserved as $C\text{-}X_2\text{-}C\text{-}X_4\text{-}H\text{-}X_4\text{-}C$ and $C\text{-}X_{1/3}\text{-}C\text{-}X_{7/8}\text{-}H\text{-}X_4\text{-}C$, respectively [13,34]. The latter motif corresponds to a CCHC box known

Fig. 7. Southern blot of Bristol N2 DNA digested with *Eco*RI and hybridized with C06E8T284, the repeat probe.

to be associated with the binding of single stranded nucleic acids [35]. Rte-1 has neither of these motifs. However not all non-LTR retrotransposon have these cysteine motifs.

4.2. Is Rte-1 a full-length element?

Full-length non-LTR elements are usually large. The LINEs in mammals are usually 6–7 kb, the ingi element in *Trypanosoma brucei* is 5.2 kb and R2Bm of *Bombyx mori* is 4.2 kb [17]. Hence the 3,298 bp long Rte-1 element is on the small side. A common feature of LINE sequences is that they are often truncated with a common 3′ end. This is thought to be due to an interruption of the reverse transcriptase process. There are examples of truncated elements in *C. elegans*; for example the predicted genes F40F12.2 and ZK1236.4 do not contain all seven reverse transcriptase domains. However as Rte-1 has the seven reverse transcriptase domains, and the DNA sequence of the ORF of both F56C9.2 and Rte-1 are of a similar length (2,316 and 2,064 bp, respectively), this may be the actual size of this group of elements in *C. elegans*.

4.3. Does Rte-1 have terminal repeats?

Rte-1 is flanked by a 200 bp direct repeat, it is not easy to determine whether this is host DNA or part of the element. Target duplications at the ends of transposable elements are usually small (5–10 bp; for reviews see Berg and Howe [1]). The duplicated region does contain an intron with conserved border sequences which is suggestive that the DNA was originally part of prk-1. Sequencing the homologous *prk-1* gene from another species would help to determine the sequence of the original *prk-1* gene and would indicate whether the duplicated region is originally part of the *prk-1* gene or part of the Rte-1 element. The mouse homolog (pim-1) [36] of *prk-1* is shorter and does not extent until the region of *prk-1* which contains the Rte-1 element.

The direct repeat has no homology to long terminal repeats found at the ends of LTR retrotransposons. The LTRs of LTR retrotransposons are characterized by three regions, U3, R and U5, which have a role in the replication of the virus. Screening the data bases with the 200 bp nucleotide sequence revealed no sequence similar to this repeat. Southern blot analysis with the direct repeat indicates that there are possibly seven other weakly hybridizing copies. One of these copies corresponds to the locus represented by cDNA cm11d12/cm11g11, because the probe used in the hybridization includes some of the region (38 bp) which is also homologous to the partially sequenced cDNA. The 10–15 elements which are similar to Rte-1 will eventually be sequenced as an integral part of the genome sequencing project and it will be interesting to determine whether they have terminal repeats. Based on the Southern blot data we would expect that these terminal repeats, if present, will not be similar in sequence to the direct repeats of Rte-1.

4.4. Is Rte-1 active?

There are 10–15 elements which are similar to Rte-1 which suggests that this element may be or has been transposing. The RFLPs detected between geographically distinct isolates of *C. elegans* also suggests that the element may be active. However, the RFLPs may also be due to point mutations or other genomic rearrangements. Rte-1 appears to be present in only a single copy in the closely related species *C. briggsae*, and *C. remanei*. The low intensity of the bands in *C. briggsae* and *C. remanei* compared to *C. elegans* indicates that the element is apparently not well conserved. Interestingly, the recently identified mariner-like repetitive sequence of *C. elegans* also hybridizes weakly to only one fragment in *C. briggsae* and *C. remanei* [3]. No retrotransposon-induced mutations have been reported thusfar in *C. elegans*, which may suggest that none of these elements are currently active. But it is also possible that the preferred integration sites are predominately located within introns or that they are efficiently spliced out when they are inserted in coding sequences. In the case of the *prk-1* gene it would seem that the gene is transcribed, because cDNA clones of *prk-1* were found. The cDNA contains one copy of the direct repeat; if the direct repeat was originally part of the element, it now contributes to the coding potential of the gene while the Rte-1 element is spliced out.

4.5. Family of Rte-1 elements in C. elegans

Rte-1 is most homologous to F56C9.2, but F56C9.2 is organized differently from Rte-1, with two non-overlapping exons. To date there have been three possible full-length elements (Rte-1, F56C9.2 and F58A4.5), two possible pseudo genes (C07A9 and T07E3.1) and two truncated elements (F40F12.2 and ZK1236.4) identified. All of these sequences are predicted by the Gene Finder program to contain reverse transcriptase-like genes. Predicted genes F58A4.5, F40F12.2, ZK1236.4 and T07E3.1 are more similar to each other than to either Rte-1 or F56C9.2, this could suggest that there are at least two different families of non-LTR retrotransposon in *C. elegans*. The seven predicted reverse-transcriptase-like non-LTR retrotransposons sequenced by the genome sequencing consortium is an amazingly high number of elements in 10 Mb of sequenced DNA.

Acknowledgements: We thank the Caenorhabditis Genetics Center for the cosmids, the cDNA clones and the *Caenorhabditis* species. We also thank our colleagues Jan Chris Vos and Piet Borst for critical comments and advice on the manuscript. This work was supported by a Wellcome Travelling Research Fellowship to S.Y. and a Pionier grant from the Netherlands Organization for Scientific Research (NWO) to R.H.A.P.

References

[1] Berg, D.E. and Howe, M.M. (1989) Mobile DNA, American Society for Microbiology, Washington, D.C.
[2] Liao, L.W., Rosenzweig, B. and Hirsh, D. (1983) Proc. Natl. Acad. Sci. USA 80, 3585–3589.
[3] Sedensky, M.M., Hudson, S.J., Everson, B. and Morgan, P.G. (1994) Nucleic Acids Res. 22, 1719–1723.
[4] Levitt, A. and Emmons, S.W. (1989) Proc. Natl. Acad. Sci. USA 86, 3232–3236.
[5] Collins, J., Forbes, E. and Anderson, P. (1989) Genetics 121, 47–55.
[6] Yuan, J., Finney, M., Tsung, N. and Horvitz, H.R. (1991) Proc. Natl. Acad. Sci. USA 88, 3334–3338.
[7] Dreyfus, D.H. and Emmons, S.W. (1991) Nucleic Acids Res. 19, 1871–1877.
[8] Moerman, D.G. and Waterston, R.H. (1989) in: Mobile DNA (Berg, D.E. and Howe, M.M. eds.) pp. 537–556, American Society for Microbiology, Washington, D.C.
[9] Britten, R.J. (1995) Proc. Natl. Acad. Sci. USA 92, 599–601.
[10] Aeby, P., Spicher, A., Chastonay, Y. de, Müller, F. and Tobler, H. (1986) EMBO J. 5, 3353–3360.
[11] Link, C.D., Graf-Whitsel, J. and Wood, W.B. (1987) Proc. Natl. Acad. Sci. USA 84, 5325–5329.
[12] Boeke, J.D. and Corces, V.G. (1989) Annu. Rev. Microbiol. 43, 403–434.
[13] Xiong, Y. and Eickbush, T.H. (1988) Mol. Biol. Evol. 5, 675–690.

[14] Mount, S.M. and Rubin, G.M. (1985) Mol. Cel. Biol. 5, 1630–1638.
[15] Marlor, R.L., Parkhurst, S.M. and Corces, V.G. (1986) Mol. Cell. Biol. 6, 1129–1134.
[16] Hansen, L.J., Chalker, D.L. and Sandmeyer, S.B. (1988) Mol. Cell. Biol. 8, 5245–5256.
[17] Hutchison III, C.A., Hardies, S.C., Loeb, D.D., Shehee, W.R. and Edgell, M.H. (1989) in: Mobile DNA (Berg, D.E. and Howe, M.M. eds.) pp. 593–617, American Society for Microbiology, Washington, D.C.
[18] Fawcett, D.H., Lister, C.K., Kellett, E. and Finnegan, D.J. (1986) Cell 47, 1007–1015.
[19] Kimmel, B.E., Moiyoi, O.K. and Young, J.R. (1987) Mol. Cell. Biol. 7, 1465–1475.
[20] Weiner, A.M., Deininger, P.L. and Efstratiadis, A. (1986) Annu. Rev. Biochem. 55, 631–661.
[21] Xiong, Y. and Eickbush, T.H. (1990) EMBO J. 9, 3353–3362.
[22] Burke, W.D., Calalang, C.C. and Eickbush, T.H. (1987) Mol. Cell. Biol. 7, 2221–2230.
[23] Murphy, N.B., Pays, A., Tebabi, P., Coquelet, H., Guyaux, M., Steinert, M. and Pays, E. (1987) J. Mol. Biol. 195, 855–871.
[24] Xiong, Y. and Eickbush, T.H. (1988) Cell 55, 235–246.
[25] Luan, D.D., Korman, M.H., Jakubczak, J.L. and Eickbush, T.H. (1993) Cell 72, 595–605.
[26] Brenner, S. (1974) Genetics 77, 71–94.
[27] Gish, W. and States, D.J. (1993) Nature Genet. 3, 266–272.
[28] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) J. Mol. Biol. 215, 403–410.
[29] Sulston, J. and Hodgkin, J. (1987) in: The Nematode *Caenorhabditis elegans* (Wood, W.B., ed.) pp. 587–606, Cold Spring Harbor Laboratory, Cold Spring Harbor.
[30] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
[31] Feinberger, A.P. and Vogelstein, B. (1983) Anal. Biochem. 132, 6–13.
[32] Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fraser, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laisster, N., Latreille, P., Lightning, J., Lloyd, C., Mortimore , B., O'Callaghan, M., Parson, J., Percy, C., Rifken, L., Roopra, A., Saunders, D., Shownkeen, R., Sims, M., Smaldon, N., Smith, A., Smith, M., Sonnhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkinson-Sproat, J. and Wohldman, P. (1994) Nature (London) 368, 32–38.
[33] Kozak, M. (1987) J. Mol. Biol. 196, 947–950.
[34] Jakubczak, J.L., Xiong, Y. and Eickbush, T.H. (1990) J. Mol. Biol. 212, 37–52.
[35] Berg, J.M. (1990) J. Biol. Chem. 265, 6513–6516.
[36] Selten, G., Cuypers, H.T., Boelens, W., Robanus-Maandag, E., Verbeek, J., Domen, J., van Beveren, C. and Berns, A. (1986) Cell 46, 603–611.