J Mol Evol (1987) 25:230–240

Journal of
Molecular Evolution

Structure and Evolution of a Family of Interspersed Repetitive DNA Sequences in *Caenorhabditis elegans*

Kevin M. Felsenstein[1] and Scott W. Emmons

Department of Molecular Biology, Division of Biological Sciences, Albert Einstein College of Medicine, Bronx, New York 10461, USA

Summary. The structure of three members of a repetitive DNA family from the genome of the nematode *Caenorhabditis elegans* has been studied. The three repetitive elements have a similar unitary structure consisting of two 451-bp sequences in inverted orientation separated by 491 bp, 1.5 kb, and 2.5 kb, respectively. The 491-bp sequence separating the inverted 451-bp sequences of the shortest element is found adjacent to one of the repeats in the other two elements as well. The combination of the three sequences we define as the basic repetitive unit. Comparison of the nucleotide sequences of the three elements has allowed the identification of the one most closely resembling the primordial repetitive element. Additionally, a process of co-evolution is evident that results in the introduction of identical sequence changes into both copies of the inverted sequence within a single unit. Possible mechanisms are discussed for the homogenization of these sequences. A direct test of one possible homogenization mechanism, namely homologous recombination between the inverted sequences accompanied by gene conversion, shows that recombination between the inverted repeats does not occur at high frequency.

Key words: Repetitive DNA family — *Caenorhabditis elegans* — Nucleotide sequence — Homologous recombination

Offprint requests to: K.M. Felsenstein
[1] *Current address:* Department of Biochemistry, Columbia University College of Physicians & Surgeons, New York, New York 10032, USA

Introduction

A significant proportion of the DNA of eukaryotic genomes consists of families of short, interspersed repeated sequences a few hundred nucleotides in length (Britten and Kohn 1968; for reviews see Davidson and Britten 1973; Lewin 1974; Jelinek and Schmid 1982; Weiner et al. 1986). In many organisms, a significant proportion of the short-period interspersed repeats occurs as inverted repeats with an arrangement and spacing similar to that of the total repeats (Davidson et al. 1973; Graham et al. 1974; Wilson and Thomas 1974; Cech and Hearst 1975; Deininger and Schmid 1976; Perlman et al. 1976; Jelinek 1977, 1978). In some cases, the inverted repeats are a subset of the total complement of short-period interspersed repeats, arising by fortuitous juxtaposition of two independent elements (Jelinek 1978). In other cases, the inverted repeat itself is the repetitive unit. Examples of this are the transposable inverted repeat units of the *Drosophila* foldback *FB* elements and the sea urchin *TU* elements (Potter et al. 1980; Liebermann et al. 1983).

Despite considerable effort, no function for the majority of the repetitive sequences in eukaryotic genomes has yet been established, and it is possible that most play no essential role in the cellular functions of the genome (Doolittle and Sapienza 1980; Orgel and Crick 1980). Attention therefore focuses on the origin of these sequences as products of the biochemical pathways that maintain and alter the structure and composition of the genome. Britten and Kohn (1968) proposed that families of repeats are produced in proliferative and dispersive events

from a primordial element, and that subsequently the members of the family diverge from one another by mutational drift. Homogenization mechanisms, such as gene conversion and unequal crossing-over, have been suggested to moderate this process. These mechanisms presumably can spread mutations that arise in one family member to other family members, resulting in co-evolution of the entire family or a portion thereof (Brown et al. 1972; Brown and Sugimoto 1973; Kedes 1980; Long and Dawid 1980; Dover 1982; Weiner and Denison 1983). Alternatively, the apparent co-evolution of a repetitive family may be explained by the recent expansion of the family from a single member (Weiner et al. 1986).

We have analyzed the structure of interspersed repetitive elements in the genome of the free-living soil nematode *Caenorhabditis elegans* and interpreted our findings in terms of the model of Britten and Kohn (1968). The genome of *C. elegans,* though small (8×10^7 bp; Sulston and Brenner 1974), is organized in the typical short-period interspersion pattern: repeated elements with a modal length of 300 bp are interspersed at intervals of a few kilobases throughout the genome (Emmons et al. 1979, 1980). Inverted repeats with the same length distribution are randomly arranged, with an average frequency of one in every 33 kb (Emmons et al. 1980). The repeated elements are organized into a large number of families (estimated to be between 100 and 1000), with each family comprising only from 10 to 100 members (Emmons et al. 1979, 1980). This is in marked contrast to larger genomes, where the number of members of single repetitive families is usually much greater and may range up to 100,000 or more (Houck et al. 1979; Anderson et al. 1981). The small size of *C. elegans* repetitive families makes their structure particularly amenable to analysis.

In earlier cloning studies where the constancy of genomic sequences was studied in *C. elegans,* it was shown that most randomly cloned restriction fragments carry a member of a small repetitive family (Emmons et al. 1979). We report here a detailed analysis of the structure of three cloned members of one of these families. Comparison of the structures of the three elements provides evidence for the structure of the primordial repeat unit, and demonstrates facets of the divergence and homogenization processes.

Materials and Methods

Nematodes. Caenorhabditis elegans strain Bristol (designated N2) was obtained from D. Hirsh and is regarded as the wild-type laboratory strain. The Bergerac strains and *C. briggsae* were from the Caenorhabditis Genetics Center. Other *C. elegans* strains were supplied to us by W. Sharrock from the collection of R. Russell. *Caenorhabditis remanei* was obtained from E. Hedgecock, and *Panagrellus redivivus* was supplied by D. Hirsh. Nematodes were grown on agar plates as described by Brenner (1974), or in liquid media as described by Sulston and Brenner (1974), at 16°C for Bergerac or 20°C for the remaining strains.

Nucleic Acids. Whole genomic DNA from populations of worms was isolated by the proteinase K–SDS lysis procedure (Emmons et al. 1979). DNA libraries of Bristol DNA fragments were prepared in the vector λ1059 using 15–20-kb restriction fragments obtained by partial digestion with the restriction endonuclease *Bam*HI, following Karn et al. (1980) and Maniatis et al. (1982). A cosmid library in the vector pTL5 (Lund et al. 1982), consisting of 40–45-kb restriction fragments from the Bristol genome generated by partial digestion with Sau3A, was supplied by S. Roberts. Hybridization probes were labeled by nick-translation (Rigby et al. 1977). Southern hybridizations and isolation of plasmid bacteriophage DNA followed standard methods (Maniatis et al. 1982). Plasmid and bacteriophage screening was carried out according to Hanahan and Meselson (1980) and Benton and Davis (1977), respectively.

DNA Sequence Analysis. The sequence of end-labeled DNA restriction fragments was determined according to the base-specific chemical degradation protocol of Maxam and Gilbert (1980). Cleavage products were displayed on either 6, 8, or 20% polyacrylamide–8 M urea gels and exposed to x-ray film (Kodak XAR5, Eastman Kodak Co.) with an intensifying screen (Lightning Plus, Du Pont Co.) at −70°C. Isolated DNA fragments or mixtures of fragments were either labeled at their 3′ ends using the Klenow fragment of *Escherichia coli* DNA polymerase I (Boehringer/Mannheim Biochemical) with the appropriate [α-^{32}P]dNTP (Amersham Corp.) following the protocol of Drouin (1980), or were labeled at their 5′ ends with [γ-^{32}P]ATP (Amersham Corp.) using T4 polynucleotide kinase (Boehringer/Mannheim Biochemical) as described by Maxam and Gilbert (1980). Ends of labeled fragments were separated by secondary cleavage with an appropriate restriction endonuclease or by strand separation (Maxam and Gilbert 1980). Labeled fragments were purified from agarose or polyacrylamide gels by electroelution into dialysis bags according to McDonnell et al. (1977) and Smith (1980). Concentration and purification were achieved using the Elutip-d Column System (Schleicher and Schuell, Inc.).

Electron Microscopy. DNA was mounted for observation in the electron microscope by the techniques described by Davis et al. (1971) and Ferguson and Davis (1978). Observation of inverted repeats was performed according to the procedures of Emmon et al. (1980). DNA molecules were observed using a JEOL JEM-100S electron microscope.

Isolation of Repeats by S1 Digestion. Inverted repeats were isolated for use as nick-translated probes by treatment of hairpin structures with a single-strand-specific nuclease, as follows. Ten to 250 μg of linearized DNA of a plasmid or phage carrying an inverted repeat was denatured by heating at 100°C for 10 min in TE buffer (10 mM Tris, pH 7.4, 0.1 mM EDTA). The heated sample was cooled quickly on ice and then allowed to reanneal at room temperature for 15 min. The reannealed sample was diluted into S1 reaction buffer (30 mM NaOAc, pH 4.6, 50 mM NaCl, 1 mM $ZnSO_4$, and 5% glycerol), and 100 to 5000 units of S1 exonuclease (Boehringer/Mannheim Biochemical) was added, followed by incubation at 37°C for 30 min. The reaction was stopped with gel loading buffer (Maniatis 1982), and the sample was fractionated by electrophoresis on a 6% acrylamide gel. The gel was stained with ethidium bromide and a band at approximately 450 bp was observed. The band was eluted, labeled by nick-translation, and used as a hybridization probe as described in the text.

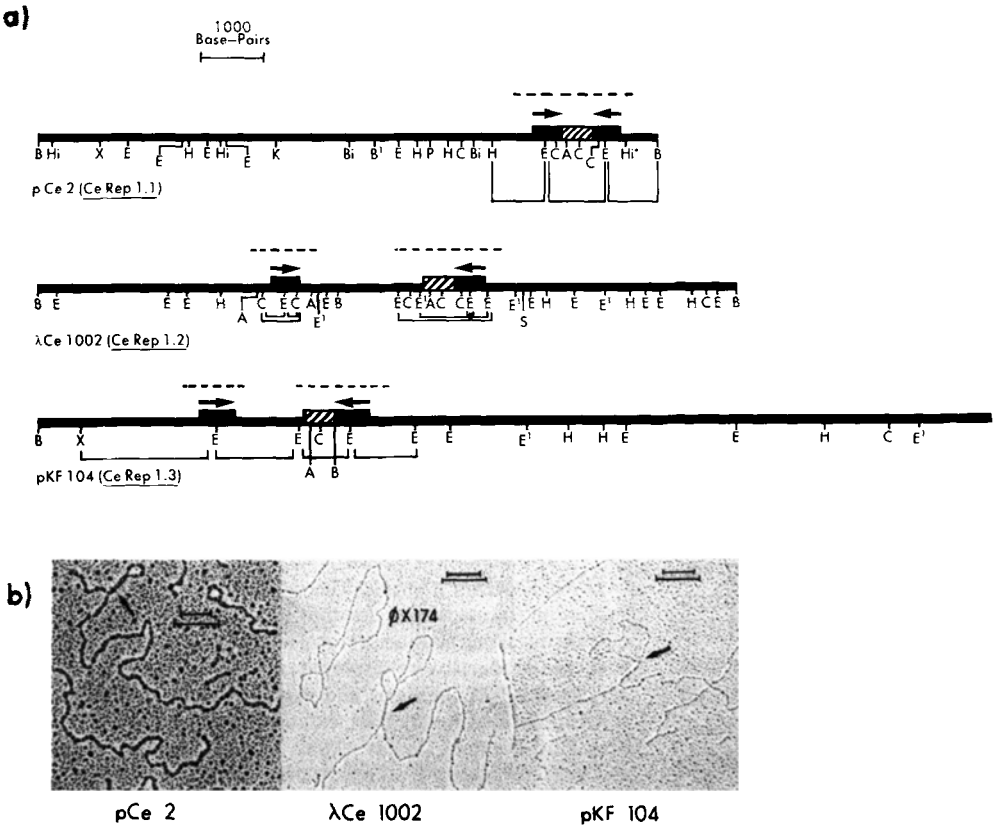

Fig. 1. Location and structure of repetitive elements on three genomic clones. **a** The restriction maps of the three genomic clones were established by means of complete digestion with the enzymes indicated. Symbols used for restriction endonuclease recognition sites are: *AvaI* (A); *Bam*HI (B); *BglI* (B'); *BglII* (Bi); *ClaI* (C); *Eco*RI (E); *Eco*RV (E'); *Hin*cII (Hi); *Hin*dIII (H); *HpaI/Hin*cII (Hi*); *KpnI* (K); *PstI* (P); *SalI* (S); and *XhoI* (X). The bracketed areas below the maps indicate those regions of DNA that were shown to cross-hybridize with the two other genomic clones. The arrows on each map show the position of the 450-bp inverted repeat sequences described in the text. The hatched box designates the conserved internal sequence. The DNA sequences of the regions covered by dashed lines are presented in Fig. 3. **b** Electron micrographs showing the presence of 450-bp inverted repeats on the genomic clones. At least 20 separate DNA molecules with identical structure were measured for each length determination. Length standards used are ØX174 at 5.3 kb for single strands and pBR313 at 9.07 kb for double strands. The upper bar indicates the measurement scale for double-stranded DNA (500 bp) and the lower bar indicates the measurement scale for single-stranded DNA (500 bases)

Results

Isolation and Structure of the CeRep1 Repetitive Family

In our earlier survey of the properties of genomic DNA of *C. elegans*, we found that most cloned restriction fragments carry a representative of a small repetitive family (Emmons et al. 1979). Here we describe the properties of one of these families. Repeats were observed by using cloned fragments as probes in genomic Southern hybridization experiments. The experiments reported here were initiated by screening the same set of cloned genomic fragments for inverted repeats in the electron microscope. Inverted repeats are present in *C. elegans* DNA with an average spacing of 33 kb, as was shown in previous electron microscopic studies of whole genomic DNA (Emmons et al. 1980). One cloned fragment, in plasmid pCe2, was found to have an inverted repeat of about 450 bp separated by about 500 bp. Because of the evidence that some inverted repeats are transposable elements (Potter et al. 1980; Liebermann et al. 1983), and because eukaryotic inverted repeats generally had not been extensively characterized, we decided to analyze the inverted repeat of pCe2 further.

Plasmid pCe2 contains a 10-kb genomic *Bam*HI fragment cloned in the vector pBR313 (Bolivar et al. 1977a). The 10-kb fragment cross-hybridizes with about 10 other genomic restriction fragments besides itself. We found that the repetitive sequence responsible for this cross-hybridization is the inverted repeat identified in the electron microscope. We call the cross-hybridizing family of sequences the *CeRep1* repetitive family. The repetitive element on pCe2 is termed *CeRep1.1*.

Additional members of the *CeRep1* family were isolated by screening clone banks with pCe2. Element *CeRep1.2* is cloned in the recombinant phage

λCe1002, which contains an 11.6-kb genomic segment cloned in the vector λ1059 (Karn et al. 1980). Element *CeRep1.3* is cloned in recombinant plasmid pKF104, containing a 15.4-kb genomic segment subcloned from a cosmid clone into the vector pBR322 (Bolivar et al. 1977b). Restriction maps of the insert in pCe2 and in the two additional cloned genomic segments are shown in Fig. 1a. The location of the repeated sequence on each clone was determined by hybridizing the clones to each other. Cross-hybridizing restriction fragments are designated by brackets below the restriction maps.

Examination in the electron microscope showed that each clone contained an inverted repeat with a separating loop region (Fig. 1b). The inverted repeat was 450 bp long, and the loop region was 500 bp for *CeRep1.1*, 2.5 kb for *CeRep1.2*, and 1.8 kb for *CeRep1.3*. To determine the position of the inverted repeats relative to the cross-hybridizing regions, each clone was digested with one or more restriction enzymes and again examined in the electron microscope. By measuring the distances of the stem and loop structures from the ends of the restriction fragments, the positions of the inverted repeats could be located on the restriction maps. These positions are indicated by arrows in Fig. 1a.

In order to show that the inverted repeat was the same as the cross-hybridizing repeat, the inverted repeat of pCe2 was isolated from flanking DNA, as described under Materials and Methods, and was used as a probe in a genomic Southern blot. The hybridization pattern given by the isolated inverted repeat was the same as that given by the entire pCe2 alone, showing that the inverted repeat is the cross-hybridizing repeat (Fig. 2; compare lanes 1, 2, 3, and 7). Sequences flanking the inverted repeat were also isolated and used as hybridization probes. They only hybridize to the genomic equivalent of the pCe2 insert, and therefore consist entirely of unique DNA (data not shown). A similar analysis was carried out with the inverted repeats of the other two clones, and the same result was obtained (data not shown).

Evidence that the cloned fragments correspond to unrearranged genomic restriction fragments is also presented in Fig. 2 (lanes 3–10). Fragments bearing all or part of a *CeRep1* family repeat in each clone correspond in size to a genomic fragment also bearing a repeat. It can be seen from this experiment that there is a small number of additional family members in the genome that we have not studied. We have no information regarding their structure, except to note that the greater intensity of the 1.0-kb *Eco*RI fragment in Fig. 2 (lane 7) suggests there may be several additional members of the *CeRep1.1* type. The members of the *CeRep1* family are interspersed widely in the genomic DNA, as deduced from the abundance of cross-hybridizing clones in both phage and cosmid clone banks and from the fact that none of the clones we studied contains two members.

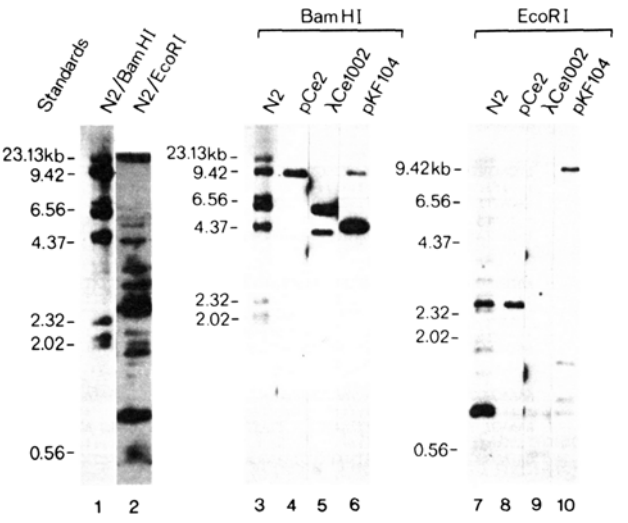

Fig. 2. The *CeRep1* repetitive family is defined by the inverted repeat. Genomic DNA digested with either *Bam*HI or *Eco*RI was hybridized to the inverted repeat of pCe2 [isolated as described under Materials and Methods (lanes 1 and 2)], or to pCe2 (lane 3) or λCe1002 (lane 7). The isolated inverted repeat of λCe1002 gave patterns identical to those shown in lanes 1 and 2. Lanes 4–6 and 8–10 show that fragments cloned correspond to genomic fragments. Probes for these lanes were: lane 4, λCe1002; lane 5, pCe2; lane 6, λCe1002; lane 8, λCe1002; lane 9, pCe2; lane 10, λCe1002.

DNA Sequence of the Repetitive Elements

The DNA sequence of the three cloned repeats is presented in Fig. 3a. The sequenced region of each clone is indicated in Fig. 1a. The sequence reveals the nature of the inverted repeat as well as the presence of a conserved 491-bp sequence in the loop region. The inverted repeat is precisely defined by comparing the two sides of the *CeRep1.1* element. In this instance the two inverted copies of the sequence are identical (Fig. 3b), and the homology of 451 nucleotides ends abruptly on both sides at a run of thymidine residues (boxed in Fig. 3a). The inverted sequences are separated by 491 nucleotides, with which they share no homology.

In the *CeRep1.2* and *CeRep1.3* elements the inverted sequences are imperfect and differ from the sequence found in *CeRep1.1*. In these elements, the intervening material has no homology to other regions of the repetitive elements, except at the right-hand boundary, where a diverged copy of the 491-nucleotide separating sequence of *CeRep1.1* is found.

The inverted repeat sequence and the conserved separating sequence were analyzed for significant internal direct and inverted repeats (Dykes et al. 1975), open reading frames, and transcriptional sig-

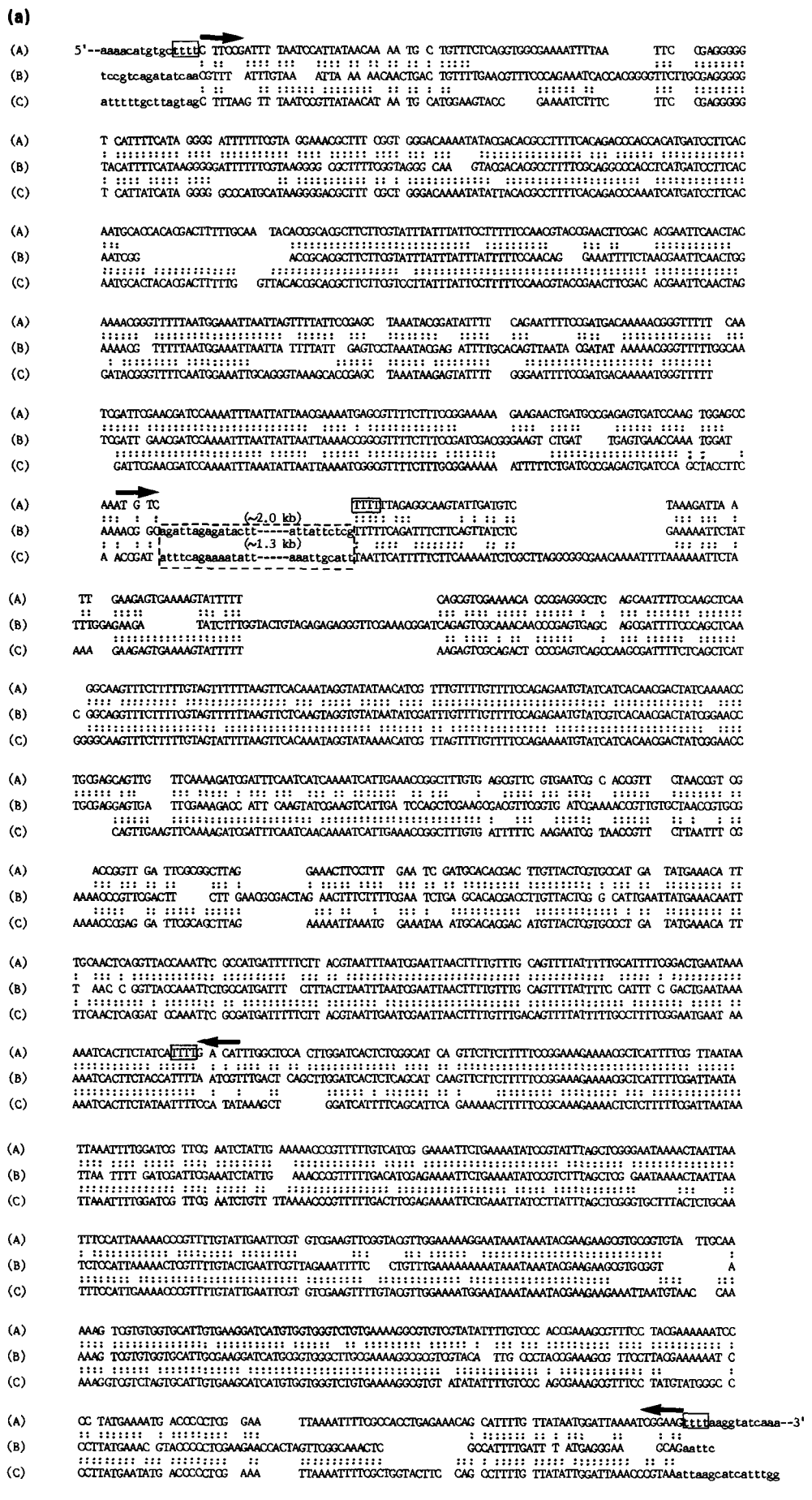

Fig. 3. **a** DNA sequence of the *CeRep1* repeats: A, *CeRep1.1*; B, *CeRep1.2*; C, *CeRep1.3*. The sequences of the three repeats are aligned to give maximum matching without attempting to discriminate between good runs of homology and random correspondence. DNA flanking the repetitive DNA is in lowercase letters. The runs of four Ts bounding the inverted repeats of *CeRep1.1* are boxed. The arrows mark the ends and orientation of the inverted repeats. Double dots show correspondence of *CeRep1.2* or *CeRep1.3* with *CeRep1.1*. Gaps are introduced in the sequence to improve the alignment. The reason for comparing the sequences to *CeRep1.1* is explained in the text. The *CeRep1.1* element is shown in its entirety, while the *CeRep1.2* and *CeRep1.3* sequences both have a large internal gap of nonhomologous, unsequenced DNA, as indicated in the figure

(b)

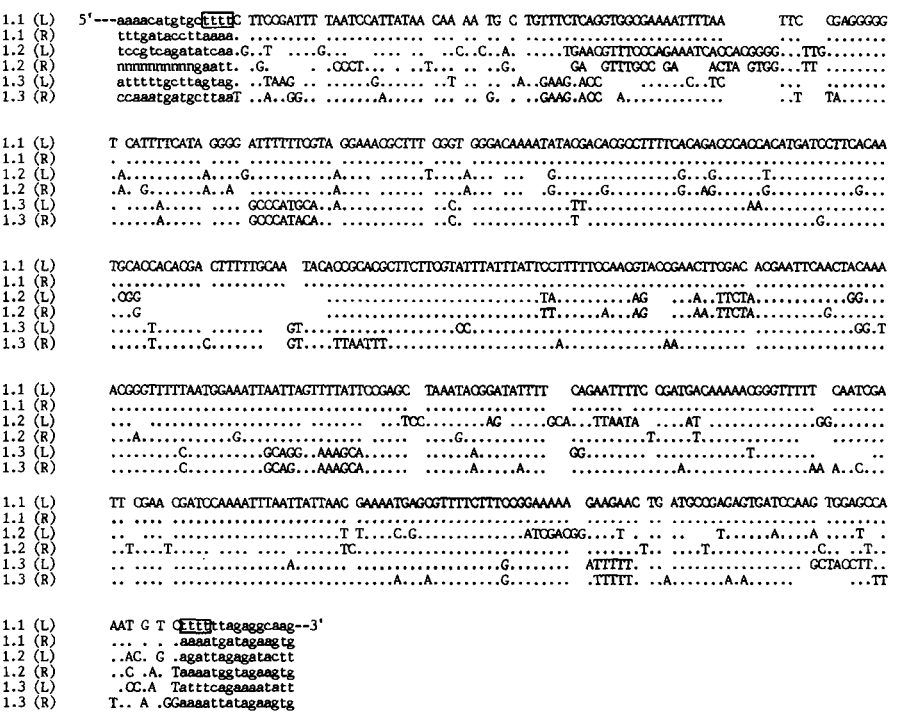

Fig. 3. **b** Sequences of the inverted repeats. The left (L) and right (R) inverted repeat of each element are aligned and compared to the left inverted repeat of *CeRep1.1*. Right repeats are inverted relative to Fig. 3a. Dots indicate a nucleotide identical to *CeRep1.1* (L), and gaps are introduced to improve the alignment. The left and right inverted repeats of *CeRep1.1* are identical. Flanking DNA is in lowercase letters, and flanking direct repeats of four Ts are boxed

nals, such as "CAAT" boxes, "Goldberg-Hogness" boxes, and the prototypic cap site (Proudfoot and Brownlee 1976; Benoist et al. 1980; Corden et al. 1980; Efstratiadis et al. 1980). The only finding is that these sequences contain a large number of short stretches of adenine and thymidine residues, which is considered typical of noncoding regions of eukaryotic genomes. The G+C content of the repeat sequences (34%) is similar to the genome as a whole (35%; Sulston and Brenner 1974).

The CeRep1.2 and CeRep1.3 Elements Have Diverged from a Sequence Similar or Identical to that of CeRep1.1

As in other systems, when the nucleotide sequences of the elements are compared, we see single base substitutions, deletions, and insertions of single bases as well as nucleotide blocks. Table 1A lists the frequency of the mutational events in a comparison of the elements. *CeRep1.2* and *CeRep1.3* are more highly diverged from each other than they are from *CeRep1.1*. When the sequences of the three repetitive elements are analyzed in more detail, a pattern emerges suggesting that the *CeRep1.1* element is more closely related to the primordial element for this family than are the *CeRep1.2* and *CeRep1.3* elements. The three sequences can be aligned in such a way that at almost every position at least two are the same, and one of these two is nearly always *CeRep1.1*. Comparing the sequences of the inverted regions and the 491-bp conserved interior sequence, there are only 15 positions where *CeRep1.2* and *CeRep1.3* are the same as each other and different from *CeRep1.1*. At over 200 other positions where any difference occurs, either *CeRep1.2* or *CeRep1.3* is the same as *CeRep1.1*, while the other is different. (These figures are only approximate, as they depend on the precise alignment employed, which is not uniquely defined at some points.) This pattern can only be explained if *CeRep1.2* and *CeRep1.3* are diverging by the independent introduction of mutations from a sequence similar or identical to that of *CeRep1.1*.

Identical Mutations Can Appear in Both Copies of the 450-bp Inverted Sequences of a Single Element

The 450-bp inverted sequences were compared to each other in all possible combinations in Table 1B. In all cases it is seen that the 450-bp sequences within a single element are more homologous to each other than to the 450-bp sequences in the other elements. The divergence between neighboring 450-bp sequences ranges from 0% for the *CeRep1.1* element to about 7% for the *CeRep1.2* and *1.3* elements, whereas the divergence between inverted repeats of different elements is 9–19.5%. Of the total of approximately 50 mutational changes (in com-

Table 1. Comparison of the *CeRep1.1, 1.2,* and *1.3* repetitive sequences

A. Comparison of the entire repeats

Sequences compared	Differences	Frequency[a]
CeRep1.1, CeRep1.2	193	0.14
CeRep1.1, CeRep1.3	141	0.10
CeRep1.2, CeRep1.3	281	0.20

B. Comparison of the 450-bp inverted sequences

Sequences compared	Uncorrected divergence[b] (%)
1.1 (R), 1.1 (L)	0
1.2 (L), 1.1 (L)	>12.5
1.2 (R), 1.1 (L)	>12.7
1.3 (L), 1.1 (L)	>9.0
1.3 (R), 1.1 (L)	>9.1
1.2 (L), 1.2 (R)	7.0
1.3 (L), 1.2 (R)	>17.0
1.3 (R), 1.2 (R)	>19.5
1.3 (L), 1.2 (L)	>16.0
1.3 (R), 1.2 (L)	>18.0
1.3 (L), 1.3 (R)	7.0

[a] Number of differences divided by the total number of nucleotides (1393)

[b] Uncorrected percent divergence is calculated by the total number of substitutions divided by the total number of nucleotides and multiplied by 100. Those divergences with a "greater than" sign preceding them are only approximate, since large deletions, insertions, and substitutions of long runs of nucleotides are defined as single events for simplicity

parison to the *CeRep1.1* sequence) that appear in the four copies of the inverted sequence in the *CeRep1.2* and *CeRep1.3* elements, 12 changes (in the case of *CeRep1.2*) and 11 changes (in the case of *CeRep1.3*) are present in both of the inverted copies within a single element. This is true for single base substitutions and changes affecting blocks of nucleotides (which are counted here as single events). In only four instances are identical changes found in homologous positions of different elements, and these are confined to single base changes. Identical changes in the inverted copies of the 450-bp sequence could imply that a homogenization mechanism is operating to maintain the homogeneity of these sequences, transferring mutations introduced into one copy of the sequence to its inverted copy.

Homologous recombination between the inverted repeats of an element, if accompanied by gene conversion events, could represent such a homogenizing mechanism. Homologous recombination would reverse the orientation of the internal sequence between the inverted sequences. In order to determine whether such recombination was occurring, we analyzed the orientation of the internal sequences of *CeRep1.1, CeRep1.2,* and *CeRep1.3* in several *C. elegans* strains. The strains are derived from independent soil isolates collected at widely separated geographical locations over a period of more than 20 years, and are sufficiently diverged to display great heterogeneity in arrangement of a known transposable element, Tc1 (Emmons et al. 1983; Liao et al. 1983). The orientation of each element was determined using an off-center restriction endonuclease site within the internal sequence as a marker. The results for *CeRep1.1* utilizing the *AvaI* site in the internal region are shown in Fig. 4. Digestion of *C. elegans* genomic DNA with both *AvaI* and *PstI* produces a fragment of 2.4 kb extending from the *AvaI* site within the sequence to a *PstI* site lying in flanking DNA. Inversion of the internal sequence would increase the size of this fragment to 2.7 kb. Using a probe specific for the unique flanking sequence, the size of this fragment was determined in 11 independent wild isolates of *C. elegans.* In all cases, the 2.4-kb fragment was seen, and hence there is no evidence for the occurrence of an inversion event since these strains were separated. Similar results indicating constancy in orientation were also obtained for the *CeRep1.2* and *CeRep1.3* sequences (data not shown). Recombination between the inverted sequences evidently does not occur frequently. If gene conversion events have taken place since these strains separated, such events must have occurred without reciprocal exchange. Alternatively, these strains may not be sufficiently diverged to exhibit evidence for these mechanisms.

CeRep1 Repeats Are Not Found in Related Species and Are Not Highly Polymorphic within C. elegans

Two lines of evidence suggest that the *CeRep1* family lacks an essential functional role within the cell. The first is that this family of repeats appears to have arisen recently within *Caenorhabditis.* Figure 5 shows that the *CeRep1* family is not present in two other species of the genus, *C. briggsae* and *C. remanei,* nor in the more distantly related nematode *P. redivivus.*

The second line of evidence is that the *CeRep1* sequence is not appreciably represented in cellular RNA. No homologous material was detected in Northern hybridization experiments employing whole cell RNA, or in a cDNA clone bank of polyA$^+$ (provided by B. Meyers; data not shown).

One class of dispensable repetitive sequences in eukaryotic genomes consists of transposable elements. Among the known transposable elements in other organisms are several with an inverted repeat structure (Potter et al. 1980; Liebermann et al. 1983; Zuker et al. 1984). Accordingly, we asked whether the *CeRep1* family is polymorphic within *C. elegans.* The arrangement of *CeRep1* elements in var-

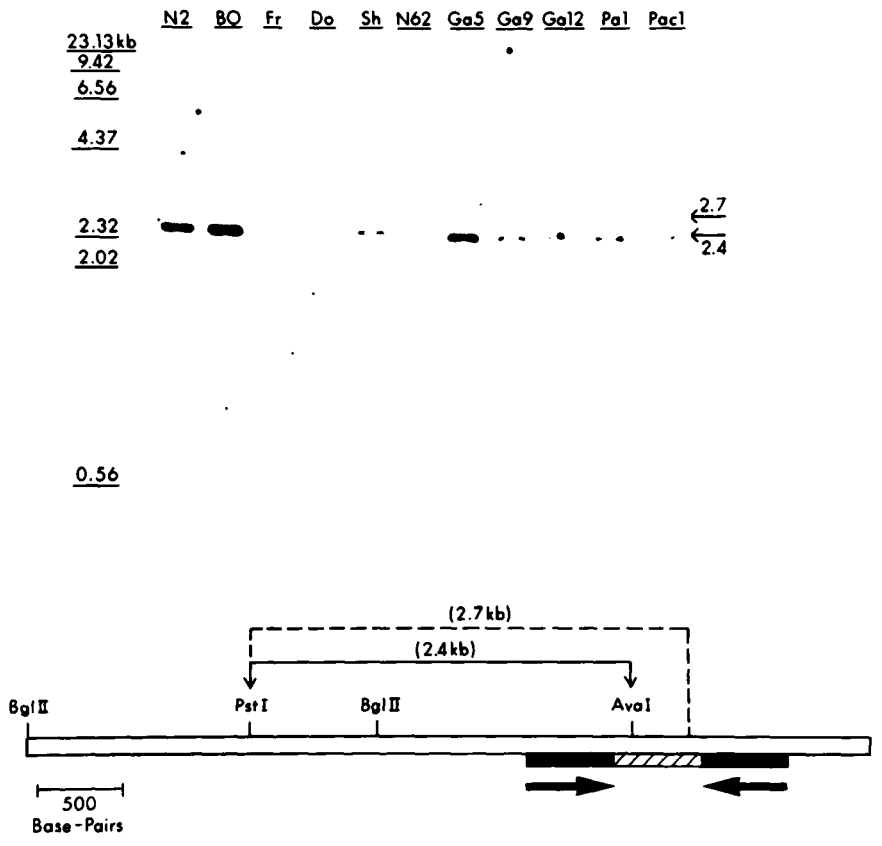

Fig. 4. The *CeRep1.1* unit is in a fixed orientation in various *C. elegans* strains. Genomic DNA from various geographically and temporally isolated *C. elegans* strains was digested with the restriction enzymes *AvaI* and *PstI* and electrophoresed on a 0.7% agarose gel. Following transfer of the gel to nitrocellulose, the filter was hybridized with the 2-kb *BglII* fragment shown on the map. In all cases a 2.4-kb fragment was detected. The relative positions of a 2.4-kb and a 2.7-kb fragment are indicated by the arrows

ious strains is shown in Fig. 5. It does not show the degree of variability that would be expected if *CeRep1* elements were actively transposing at high frequency. Additionally, the insertion of such sequences usually results in the duplication of the sequence at the target site, resulting in a direct repeat (4–11 bp) at the ends of the inserted sequence (Farabaugh and Fink 1980; Gafner and Philippsen 1980; Levis et al. 1982). In only the *CeRep1.1* unit does such a sequence exist, with four thymidine residues found at the 5′ and 3′ ends of the sequence. The *CeRep1.2* unit shows two adenine residues at its ends, and the *CeRep1.3* unit shows no duplication. The significance of the sequence in the *CeRep1.1* unit is not known and the other data do not support the hypothesis that the element is transposable.

Discussion

Origin of the CeRep1 Repetitive Family

Although short inverted repeats are present throughout the DNA of many eukaryotic organisms, the structure of this class of repetitive sequences has not been extensively studied. In mammalian genomes, many of the short inverted repeats consist of Alu sequences, which have been well characterized (Jelinek and Schmid 1982). Alu sequences are present as both isolated repeats and in inverted repeat pairs. Because of the similarity in the length distribution of interspersed and inverted repeats in the *C. elegans* genome (Emmons et al. 1980), we speculated that, as with Alu repeats, these two classes of sequences would overlap. The results reported here indicate that this may not be the case. It will be of interest to determine whether the remaining members of the *CeRep1* family are all inverted repeats with similar structures.

We argue that the *CeRep1* family lacks an essential function in *C. elegans,* primarily because this family is not present in the very similar species *C. briggsae* and *C. remanei.* This conclusion is supported by the observation that two of the family members are diverging from the primordial repeat by the accumulation of mutations. The presence of these repeats in the genome must therefore be explained in terms of the mechanism that has repeatedly introduced them.

In some respects the structure of *CeRep1* repeats resembles the structure of known eukaryotic transposable elements, DIRS-1 elements of *Dictyostelium* (Zuker et al. 1984) and TU elements of sea urchins (Liebermann et al. 1983) are flanked by inverted repeats of several hundred nucleotides. FB elements of *Drosophila* have an overall inverted repeat structure (Potter et al. 1980). Both TU and FB elements have variable sequences between the in-

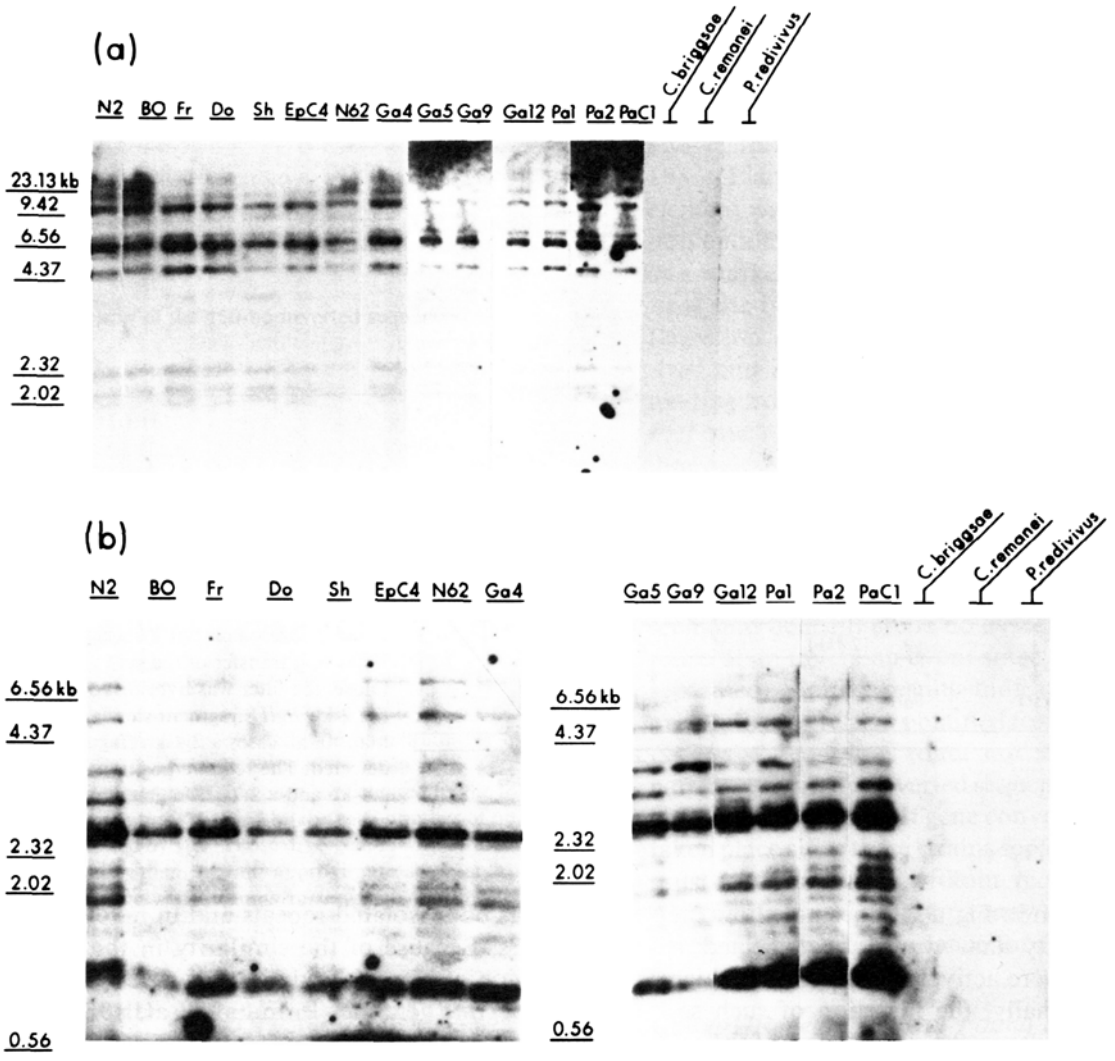

Fig. 5. The *CeRep1* family is confined to *C. elegans* but is not highly polymorphic within this species. Genomic DNA from various strains of *C. elegans*, in addition to three other nematodes, *C. briggsae*, *C. remanei*, and *P. redivivus*, was treated with either the endonuclease *Bam*HI (**a**) or *Eco*RI (**b**), fractionated by electrophoresis on a 0.7% agarose gel, transferred to nitrocellulose, and hybridized to pCe2. Hybridization to the *C. elegans* strains was under standard conditions (Maniatis 1982), whereas hybridization to non-*C. elegans* strains was performed under low stringency conditions. Size markers are indicated along the vertical axes in kilobase pairs (kb).

verted repeats; both differ from *CeRep1* elements, however, in being made up in part from multiple direct repeats of a short sequence. DIRS-1 elements differ from *CeRep1* repeats in having a constant central region (Cappello et al. 1985). The runs of Ts flanking the *CeRep1.1* repeat are reminiscent of target site duplications. However, these are found at the inside boundaries of the inverted repeats as well as flanking the entire element. Thus, *CeRep1* elements might resemble composite transposons in bacteria and *Drosophila*, which consist of variable DNA flanked by active transposons (Kleckner 1981; Paro et al. 1983). However, the inverted repeats themselves have no other resemblance to transposable elements, such as short inverted terminal repeats and internal open reading frames. Nevertheless it remains possible that *CeRep1* repeats are derived from or related to an active transposon, and are inserted into the genome by a transpositional mechanism. Alternatively, they may be created by the action of cellular mechanisms involved with alteration or maintenance of the genomic DNA, for example, the process of gene conversion (see below).

Divergence and Co-evolution within the CeRep1 Repetitive Family

There are two significant findings in the analysis of the *CeRep1* repetitive units. The first is that the *CeRep1.1* repetitive unit is similar to the primordial repeat from which the *CeRep1.2* and *CeRep1.3* units have independently diverged. The second finding is that the inverted 450-bp sequence within a particular unit tends to be maintained. This latter finding

suggests that horizontal or concerted evolution of the sequences is occurring.

Horizontal or concerted evolution of inverted repeats, such as found here for *CeRep1* repeats, has been observed in the case of *Drosophila* heat-shock genes (Brown and Ish-Horowicz 1981) and *C. elegans* heat-shock genes (Russnak and Candido 1985). In the case of the *CeRep1* repeats, the rate of gene conversion must be lower than the rate of spontaneous mutation, because the inverted repeats of *CeRep1.2* and *CeRep1.3* are not identical. We attempted to gain additional evidence for an interaction between the inverted repeats of *CeRep1* sequences by determining whether reciprocal recombination occurs between them. As the strains we have compared are all quite similar at the DNA level (Fig. 5), it is probable that they are not sufficiently diverged to exhibit changes due to either spontaneous mutation or gene conversion. This could explain the lack of evidence for reciprocal recombination.

Alternatively, gene conversion may take place without reciprocal recombination, as has been directly demonstrated in yeast (Klein 1984). On theoretical grounds, it is expected that reciprocal recombination between short homologous regions does not occur in eukaryotes, since such recombination would lead to a scrambling of the genome due to the prevalence of interspersed repeated sequences. It may be that the process of gene conversion is in part responsible for the creation, structure, and maintenance of the *CeRep1* family. The conservation of the 491-bp separating sequence might be explained if it is a *cis*-acting element involved in catalyzing such a reaction.

Acknowledgments. We thank D. Albertson for her comments on the manuscript and J. Rogers for helpful discussions. We are grateful to M. Sanicola, S. Roberts, and B. Meyers for providing DNA, RNA, and clone banks. We thank M. Malone and S. Smith for careful preparation of media. This work was supported by grants from the NIH (GM 29178) and the NSF (DCB 8316338) to S.W.E. K.M.F. was supported by NIH Training Grant GM 07491 and S.W.E. was supported by American Cancer Society grant FRA-227. Some nematode strains used in this work were provided by the Caenorhabditis Genetics Center, which is funded by a grant from the NIH to the University of Missouri.

References

Anderson DM, Scheller RH, Posakony JW, McAllister LB, Trabert SG, Beall C, Britten RJ, Davidson EH (1981) Repetitive sequences of the sea urchin genome: distribution of members of specific repetitive families. J Mol Biol 145:5–28

Benoist C, O'Hare K, Brealtnach R, Chambon R (1980) The ovalbumin gene-sequence of putative control regions. Nucleic Acids Res 8:127–142

Benton WD, Davis RW (1977) Screening λgt recombinant clones by hybridization to single plaques *in situ*. Science 196:180–182

Bolivar F, Rodrigues RL, Betlach MC, Boyer HW (1977a) Construction and characterization of new cloning vehicles I. Ampicillin-resistant derivatives of the plasmid PMB9. Gene 2:75–93

Bolivar F, Rodrigues RL, Greene PJ, Betlach MC, Hegneckar HL, Boyer HW, Crosa JH, Falkow S (1977b) Construction and characterization of new cloning vehicles II. A multipurpose cloning system. Gene 2:95–113

Brenner S (1974) The genetics of *Caenorhabditis elegans*. Genetics 77:71–94

Britten RJ, Kohn DE (1968) Repeated sequences in DNA. Science 161:529–540

Brown AJL, Ish-Horowicz D (1981) Evolution of the 87A and 87C heat-shock loci in *Drosophila*. Nature 290:677–682

Brown DD, Sugimoto K (1973) The structure and evolution of ribosomal and 5S DNAs in *Xenopus laevis* and *Xenopus mulleri*. Cold Spring Harbor Symp Quant Biol 38:501–505

Brown DD, Wensink PC, Jordan E (1972) Comparison of the ribosomal DNAs of *Xenopus laevis* and *Xenopus mulleri*: the evolution of tandem genes. J Mol Biol 63:57–73

Cappello J, Hardelsman K, Lodish HF (1985) Sequence of *Dictyostelium* DIRS-1: an apparent retrotransposon with inverted terminal repeats and an internal circle junction sequence. Cell 43:105–115

Cech TR, Hearst JE (1975) An electron microscopic study of mouse foldback DNA. Cell 5:429–446

Corden J, Wasylyk B, Buchwalder A, Sassoni-Corsi P, Kedinger C, Chambon P (1980) Promoter sequences of eukaryotic protein coding genes. Science 209:1406–1414

Davidson EH, Britten RJ (1973) Organization, transcription, and regulation in the animal genome. Quart Rev Biol 48:565–613

Davidson EH, Hough BR, Anderson CS, Britten RJ (1973) General interspersion of repetitive with non-repetitive sequence elements in the DNA of *Xenopus*. J Mol Biol 77:1–23

Davis RW, Simon MN, Davidson N (1971) Electron microscope heteroduplex methods for mapping regions of base sequence homology in nucleic acids. Methods Enzymol 21:413–248

Deininger PL, Schmid CW (1976) An electron microscope study of the DNA sequence organization of the human genome. J Mol Biol 106:773–790

Doolittle WF, Sapienza C (1980) Selfish genes, the phenotype paradigm and genomic evolution. Nature 284:601–603

Dover G (1982) Molecular drive: a cohesive mode of species evolution. Nature 299:111–117

Drouin J (1980) Cloning of human mitochondrial DNA. J Mol Biol 140:15–34

Dykes G, Bambara R, Marians K, Wu R (1975) On the statistical significance of primary structural features found in DNA–protein interaction sites. Nucleic Acids Res 2:327–345

Efstratiadis A, Maniatis T, Lawn RM, O'Connor C, Spritz RA, DeRiel JK, Forget BG, Weissmann SM, Slightom JL, Blechl AE, Smithies O, Baralle FE, Shoulders CC, Proudfoot MJ (1980) The structure and evolution of the human β-globin gene family. Cell 21:653–668

Emmons SW, Klass MR, Hirsh D (1979) Analysis of the constancy of DNA sequences during development and evolution of the nematode *Caenorhabditis elegans*. Proc Natl Acad Sci USA 76:1333–1337

Emmons SW, Rosenzweig B, Hirsh D (1980) Arrangement of repeated sequences in the DNA of the nematode *Caenorhabditis elegans*. J Mol Biol 144:481–500

Emmons SW, Yesner L, Ruan K, Katzenberg D (1983) Evidence for a transposon in *Caenorhabditis elegans*. Cell 32: 55–65

Farabaugh PM, Fink GR (1980) Insertion of the eukaryotic transposable element Ty1 creates a 5-base pair duplication. Nature 286:352–356

Ferguson J, Davis RW (1978) Advanced techniques in biological electron microscopy II, vol 2. Koejler JR (ed) Springer Verlag, Berlin, p 123

Gafner L, Philippson P (1980) The yeast transposon Ty1 generates duplications of target DNA on insertion. Nature 286: 414–418

Graham DE, Neufeld BR, Davidson EH, Britten RJ (1974) Interspersion of repetitive and non-repetitive DNA sequences in the sea urchin genome. Cell 1:127–137

Hanahan D, Meselson M (1980) Plasmid screening at high colony density. Gene 10:63–67

Houck CM, Rinehart FP, Schmid CW (1979) A ubiquitous family of repeated DNA sequences in the human genome. J Mol Biol 132:289–306

Jelinek WR (1977) Specific nucleotide sequences in HeLa cell inverted repeated DNA: enrichment for sequences found in double-stranded regions of heterogeneous nuclear RNA. J Mol Biol 115:591–601

Jelinek WR (1978) Inverted repeated DNA from Chinese hamster ovary cells studied with cloned DNA fragments. Proc Natl Acad Sci USA 75:2679–2683

Jelinek WR, Schmid CW (1982) Repetitive sequences in eukaryotic DNA and their expression. Annu Rev Biochem 51: 813–833

Karn J, Brenner S, Barnett L, Cesareni G (1980) Novel bacteriophage cloning vector. Proc Natl Acad Sci USA 77:5162–5176

Kedes LH (1980) Histone genes. Annu Rev Biochem 48:837–870

Kleckner N (1981) Transposable elements in prokaryotes. Annu Rev Gen 15:341–404

Klein HL (1984) Lack of association between intrachromosomal gene conversion and reciprocal exchange. Nature 310: 748–753

Levis R, Dunsmuir P, Rubin GM (1982) Terminal repeats of the drosophila transposable element copia: nucleotide sequence and genomic organization. Cell 21:581–588

Lewin B (1974) Gene expression, vol 2. Eukaryotic chromosomes. John Wiley & Sons, New York, New York

Liao LW, Rosenzweig B, Hirsh D (1983) Analysis of a transposable element in *Caenorhabditis elegans*. Proc Natl Acad Sci USA 80:3585–3589

Liebermann D, Liebermann BH, Weinthal J, Childs G, Maxson R, Mauron A, Cohen SN, Kedes L (1983) An unusual transposon with long terminal repeats in the sea urchin *Strongylocentrotus purpuratus*. Nature 306:342–347

Long ED, Dawid IB (1980) Repeated genes in eukaryotes. Annu Rev Biochem 49:727–766

Lund T, Grosveld FG, Flavell RA (1982) Isolation of transforming DNA by cosmid rescue. Proc Natl Acad Sci USA 79: 520–524

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning—a laboratory manual. Cold Spring Harbor Publications, Cold Spring Harbor, New York

Maxam AM, Gilbert W (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol 65: 499–560

McDonnell MW, Simon MN, Studier FW (1977) Restriction fragments of T7 DNA. J Mol Biol 110:119–146

Orgel LE, Crick FHC (1980) Selfish DNA: the ultimate parasite. Nature 284:604–606

Paro R, Goldberg MC, Gehring WJ (1983) Molecular analysis of large transposable elements carrying the white locus of *Drosophila melanogaster*. EMBO J 2:853–860

Perlman S, Phillips C, Bishop J (1976) A study of foldback DNA. Cell 8:33–42

Potter S, Truett MA, Phillips M, Maher A (1980) Eukaryotic transposable genetic elements with inverted terminal repeats. Cell 20:639–647

Proudfoot NJ, Brownlee GG (1976) 3′ Non-coding region sequences in eukaryotic messenger RNA. Nature 263:211–214

Rigby PWJ, Dieckman M, Rhodes C, Berg P (1977) Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick-translation with DNA polymerase I. J Mol Biol 113:237–251

Russnak RH, Candido EP (1985) Locus encoding a family of small heat-shock genes in *C. elegans*. Two genes duplicated to form a 3.8 kilobase inverted repeat. Mol Cell Biol 5:1268–1278

Smith HO (1980) Recovery of DNA from gels. Methods Enzymol 65:371–380

Sulston JE, Brenner S (1974) The DNA of *Caenorhabditis elegans*. Genetics 77:71–104

Weiner AM, Denison RA (1983) Either gene amplification or gene conversion may maintain the homogeneity of the multigene family encoding human U1 small nuclear RNA. Cold Spring Harbor Symp Quant Biol 48:1141–1149

Weiner AM, Deininger PL, Efstratiadis A (1986) Non-viral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu Rev Biochem 55:631–661

Wilson DA, Thomas CA (1974) Palindromes in chromosomes. J Mol Biol 84:115–144

Zuker C, Capello J, Lodish HF, George P, Chung S (1984) *Dictyostelium* transposable element *DIRS-1* has 350 base pair inverted terminal repeats that contain a heat-shock promoter. Proc Natl Acad Sci USA 81:2660–2664

Received August 12, 1986/Revised and accepted January 9, 1987