Volume 17 Number 21 1989 Nucleic Acids Research

Two highly conserved transcribed regions in the 5S DNA repeats of the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae*

Donald W.Nelson+ and Barry M.Honda*

Department of Biological Sciences, Simon Fraser University, Burnaby, BC V5A 1S6, Canada

Received April 21, 1989; Revised August 17, 1989; Accepted September 26, 1989 EMBL accession nos. X16225, X16226

ABSTRACT

The 5S RNA genes of *Caenorhabditis briggsae* consist of approximately 65 copies of a 1 kb repeat unit and 20 copies of a related 0.7 kb repeat unit, organized in separate tandem clusters. DNA sequence comparisons with the 1kb 5S DNA repeat from the closely related nematode *C. elegans* show that the 5S RNA coding region is perfectly conserved. Both *C. briggsae* 1 kb and 0.7 kb repeats are also efficiently transcribed *in vitro*, suggesting that both represent functional 5S RNA genes. Surprisingly, a second block of 118 bp is also perfectly conserved between the 1 kb repeats, and is less well conserved in the 0.7 kb repeat. In *C. elegans*, this DNA is transcribed to produce an abundant 100 nt transcript (SL RNA) which participates in a trans-splicing process (Krause and Hirsh, Cell *49*:753, 1987). This SL RNA region of the *C. briggsae* 1 kb 5S DNA repeat also appears to be transcribed *in vivo*, while the corresponding region of the 0.7 kb repeat is not.

INTRODUCTION

Eukaryotic 5S RNA genes have provided an important model system to study gene organization, evolution and expression. In general these genes are organized as tandem clusters of homogeneous repeat units (1), although there are exceptional cases of organisms with dispersed or heterogeneous 5S RNA genes (2,3). This tandem organization may facilitate the maintenance of repeat homogeneity and copy number via unequal cross-over and gene conversion (reviewed in 4). Extensive transcription studies have shown that a gene-specific factor interacts with important sequences within the 5S RNA coding region. This internal control region is sufficient for transcription in *Xenopus laevis*, but there are other organisms for which 5′flanking regions are required for efficient transcription *in vitro* (reviewed in 5).

We are interested in transcription in the nematode *C. elegans*, a model organism with numerous advantages for biochemical and genetic analysis (6). Towards this goal, we had previously described the structure and genomic organization of the 5S RNA gene family of *C. elegans*. Its haploid genome contains approximately 110 copies of a single, homogeneous 1kb repeat family encoding 5S RNA, organized as a tandem cluster on the right arm of LG V (7,8). This 1 kb repeat is transcribed in cell-free extracts derived from *C. elegans* embryos (9); we were next interested in which DNA sequences might be required for transcription. In this report we have characterized 5S DNA from a closely related nematode, *C. briggsae*, and have then compared it to *C. elegans* 5S DNA, on the assumption that important, functional sequences will be highly conserved, while non-functional ones will have diverged extensively. Thus, a comparison of 5S DNA repeats between species generally shows that 5S RNA coding sequences are highly conserved, while spacer sequences have diverged (10,11). A similar approach has been used very successfully in

© IRL Press

delineating important coding and regulatory sequences in other genes in Drosophila (12–14).

C. elegans and *C. briggsae* are morphologically indistinguishable but reproductively separate nematode species. Previous studies show that the genomes of these species have diverged extensively, sharing approximately 10% of their genomic sequences (15). Recent work suggests that this sequence identity is confined to genomic sequences which appear in the cellular RNA pool (27, 28, T. Snutch and D.L. Baillie, personal communication).

In this paper we report on the structure and organization of the 5S RNA genes of *C. briggsae*. We show that *C. briggsae* 5S RNA is encoded by two related repeat families which are organized in mutually exclusive tandem clusters. Comparison of *C. elegans* and *C. briggsae* 5S DNA repeat sequences reveals two blocks of highly conserved sequence. As expected, one corresponds to the 5S RNA coding region. However, the other encodes a short primary transcript which is associated with trans-splicing onto the 5′ ends of a variety of cellular mRNAs in *C. elegans* (16,17).

MATERIALS AND METHODS

Restriction enzymes, S1 nuclease and large fragment DNA polymerase (Klenow fragment) were obtained from Pharmacia and used as described (18). Exo III was obtained from New England Biolabs.

Nucleic acid isolations

C. briggsae hermaphrodites were grown on hi-peptone agar plates streaked with wild-type *E. coli* as described previously (19). Nematodes were harvested and washed with 0.04M NaCl, and lysed with proteinase K (200 μg/ml, Sigma) in 0.1M Tris pH 8.5, 0.05M EDTA, 0.2M NaCl, 1% SDS (sodium dodecyl sulfate), at 65°C. Following three phenol and two chloroform-isoamyl alcohol (24:1) extractions, genomic DNA was precipitated at room temperature with 2 volumes 95% ethanol, dissolved in 10mM Tris-HCl pH7, 1mM EDTA and purified on CsCl-ethidium bromide, equilibrium density gradients (20). RNA remaining in the ethanol supernatant was precipitated on further addition of 0.5 volumes 95% ethanol at −20°C, and used without further purification.

Recombinant plasmids were prepared by alkaline lysis and CsCl centrifugation as described (18).

Restriction fragments were isolated by fractionation on 1% agarose gels followed by electroelution (18).

Isolation of C. briggsae 5S DNA repeats

25 μg of *C. briggsae* genomic DNA was digested to completion with *Hind*III and size fractionated on 1% agarose. Genomic repeat bands were visualized with ethidium bromide; 1 kb and 0.7 kb repeat bands were excised and recovered by electroelution. Isolated repeats were ligated into *Hind*III linearized pUC13 and transformed into *E. coli* JM83 (21). Recombinants carrying 5S DNA were identified by colony hybridization to a fragment of pCe5S1 encoding *C. elegans* 5S RNA (positions −8 to +120, see ref 7).

Blots and hybridization conditions

C. briggsae genomic DNA was digested to completion with the indicated enzymes and size fractionated on 1% agarose (1 μg/lane). Restriction digests were transferred to nylon filters (Dupont GeneScreen, manufacturer's protcol), and hybridized overnight to isolated restriction fragments nick-translated (22) to a specific activity of 10^8 cpm/μg in 5×SSPE (SSPE = 0.18 M NaCl, 10 mM $NaPO_4$ pH7.0, 1 mM Na_2EDTA) at 65°C. Filters were washed extensively in 0.5×SSPE at 65°C and autoradiographed wet on XAR-5 and XK-1

```
(A)
  1   AAGCTTTTGCTTTTTTTGTATTCAATTTCATAAATATGAATTCAAAGATTTCAATTTTTA
 61   AGATTCATGTTCTTAAAAAATTGCAGAATTTTCATCATTGGTTTGGATTTTCGTGCTTTT
121   TTTCTAATATTTATTTATTTTGTAAATAAATTTGTAAAATGTTTTCCTCACAAATGAGAA
181   GTGTTGTCGAAAAATATAAAATTGTGTCAAATTCAATCAATTTCAATATGAAAAATTTGA
241   ATGTTCAAGATGCATTCGCGTTTTCTTCCCACCACATTGCCTGAGTTTCTGAAATAAATT
301   GTGCATGAATCGACTGAAAATAGATGTGTGTGTATAAAATTTATATTGTAAAAATTTCAG
361   AATCTAAGCCTAAAATTTTAAGAATTCACTAGAATTTAAATGATGACAATTCCGATCGTT
421   TTACAACCCTTTTACGATGCATCGATTTTTCGATTTTCAATGGACCATAACTCCGGCATG
481   CGTTGACCGATTTTCAATTTTTAAAGTTATTTTGTCTCCCCGAGAGGAGACGTTCCAAAA
541   TTTATAGCTAACGCCAAATTTCTTTGGGTCAGTTTCAATGTTTACCTCAAACTTGGGTAA
601   TTAAACCAACTACATCGGCGGGCTTCGCACACTAGTGGAAAACGAGCGGCCGACACCAAT
661   CGAGGCCCGCGTCGGCCACCGCATTTCGACAGCGTGCGCGCGCACACTGCTGGGTGAGTC
721   TTCTTCTACTGTTGGGGGGACTGGGAGAATTCGCTCTTCTGCGTTTCGATTTATTTCACT
781   GATCCCTAGTAGAGTTAAAAGGGGAATGTAGAGGTAGATGTGATGCTTACGACCATATCA
841   CGTTGAATGCACGCCATCCCGTCCGATCTGGCAAGTTAAGCAACGTTGAGTCCAGTTAGT
901   ACTTGGATCGGAGACGGCCTGGGAATCCTGGATGTTGTAAGCTT

(B)
  1   AAGCTTTTTGATTTTTTATATTCATTGTATAAATATGAATTGTGATTCCTGATATTGATT
 61   TTTGATTTTTTTCGGTTTTGAATAAAAAATTTCTatatttttgccaggtcggtttaatta
121   cccaagtttgagtaTTAATATAGGTATTTGTCATCatttttgccaggtcggtttaattac
181   ccaagtttgaggtaGACGGGCGGAGTGTTGTACCCAATTCTCAGATAATACACAGTAGCA
241   TAGTGGCCAACGCGGGCCTCGATCGGTGTCGGGTGCTGGGAAATGTGAAGAGGGGAGACA
301   ACGCTTTCAGCTTTggtttaattacccaagtttgaggtaCTTTGCGTTTTCGACCTAAAG
361   AAATTTAGAGTTGATAACAAATTTTGGAACGTCCCCTTTCGGGGAGACAAAAAATCTAAA
421   GAAATTAAAAATTGGTCAACTATGAACCGAGTTATGATTTTTTGAAATTTTGAAAAATTC
481   CTCTTTTTTTCAATTTCAAATTCTCAATTTTCATCGACGTTATTTTTTTCTTACACAGGA
541   AGAGAAAATTTATATATGAGAAAAGGAGGTGAGAAAAATGCTTACGACCATATCACGTTG
601   AATGCACGCCATCCCGTCCGATCTGGCAAGTTAAGCAACGTTGAGTCCAGTTAGTACTTG
661   GATCGGAGACGGCCTGGGAATCCTGGATGTTGTAAGCTT
```

Fig. 1. *Sequence of C. briggsae 5S DNA repeats*
Each sequence is numbered relative to the *Hind*III site separating repeats, with SL RNA and 5S RNA underlined. See also Fig. 2A for schematic drawing of the repeats.
A) Cb5S1. Note that the SL RNA is in the opposite orientation relative to 5S RNA; thus, in this figure the sequence complementary to SL RNA is shown (510 to 607), while the direct sequence of 5S RNA (825 to 943) is underlined.
B) Cb5S0.7. The SL RNA-like sequence (315 to 409), and 5S RNA (580 to 698) are underlined. The 39bp sequence which is non-tandemly repeated in Cb5S0.7 (and partially repeated within the SL RNA-like sequence) is indicated by lower case letters.

(Kodak) film. After autoradiography, filters were stripped by boiling in distilled water, and reprobed as described above.

5S DNA repeat copy number was determined by comparing the genomic hybridization signal to plasmid standard curves. *Hind*III digests of *C. briggsae* genomic DNA (1 μg = 1.1×10^7 haploid genomes), pCb5S1 and pCb5S0.7 (0.5 ng to 8.8 ng = 10 to 175 haploid genome equivalents) were size fractionated on 1% agarose, transfered to nitrocellulose and hybridized to nick-translated Cb5S1 and Cb5S0.7 probes. Hybridizing bands were excised, Cerenkov counts were determined, and bound radioactivity was plotted against number of genome equivalents blotted.

C. briggsae RNA was fractionated on 10% acrylamide 7M urea gels and electroblotted to nylon filters (Dupont GeneScreen, suppliers protocol). Filters were prehybridized for 5 hours in 5×SSPE, 5× Denhart's (0.5% Ficoll, 0.5% PVP, 0.5% BSA), and 0.3% SDS at 45°C. Oligonucleotide probes complementary to SL RNA transcripts, and specific to either the 1 kb (5′L-1) or 0.7 kb repeat (5′L-0.7) (see figure 6 for details) were end-labelled using T4 polynucleotide kinase and [α-^{32}P]ATP (Amersham) as described (18). Probes were hybridized to RNA filters in 5×SSPE, 5× Denhart's, 0.3% SDS overnight at 45°C, washed in 1×SSPE, 0.3% SDS at 45°C and autoradiographed on Kodak XK-1 film at room temperature.

Plasmid sequencing and sequence analysis

Cb5S1 and Cb5S0.7 were cloned in both orientations in the *Hind*III site of pUC13. Overlapping deletions from both ends of each repeat were constructed using the exoIII-S1 protocol of Hennikoff (23). DNA sequences were determined for both strands from

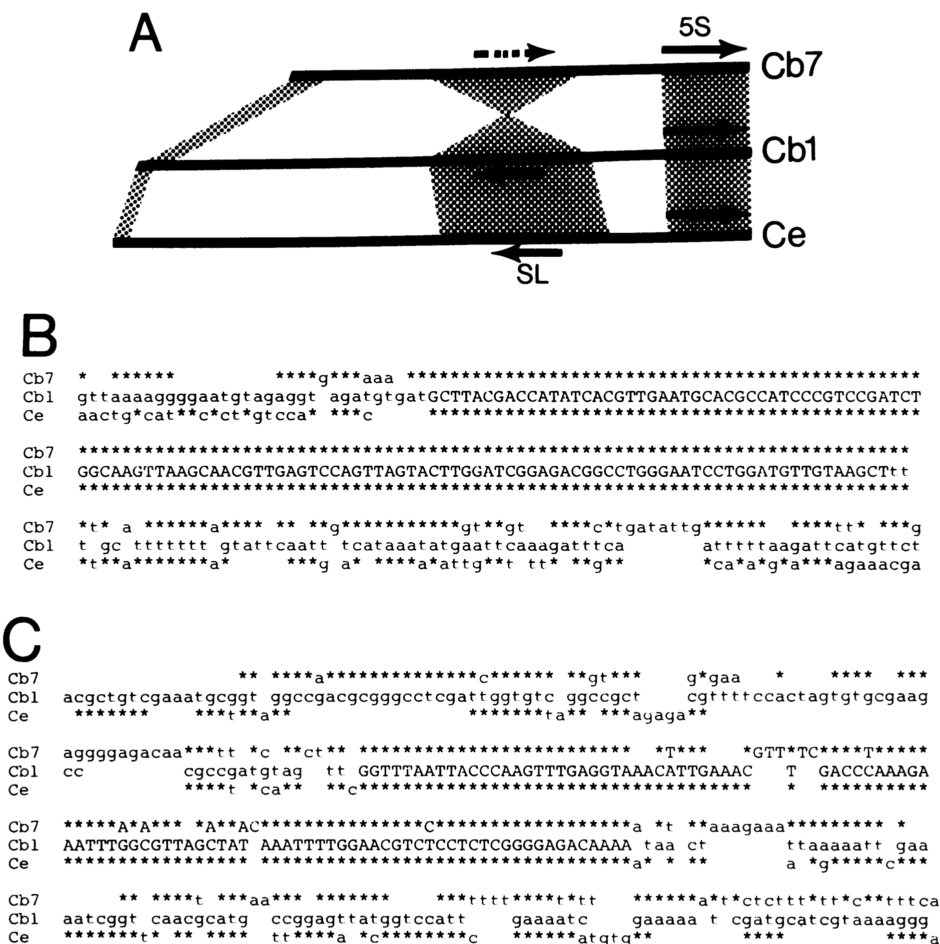

Fig. 2. *Comparison of C. elegans and C. briggsae 5S DNA repeat sequences.* Pairwise sequence comparisons were done using the SEQNCE program of Delaney (Delaney Software, Vancouver, Canada), using the 'Homology' function and varying initial match length to get optimal sequence alignment.

A) Line drawing showing the gross organization of the repeats and their conserved regions. Each repeat (Cb7=Cb5S0.7, Cb1=Cb5S1, and Ce=Ce5S1) is drawn to scale and the locations of the transcription units (SL=SL RNA, 5S=5S RNA) indicated by the arrows. Conserved regions are indicated by shading. Note that the conserved region encompassing the SL RNA-like region of Cb5S0.7 is inverted relative to SL RNA of Cb5S1 and Ce5S1.

B) Sequence alignment of 5S RNA coding region including 5′ and 3′ flanking sequences. The repeats shown are: Cb7=Cb5S0.7 (positions 560 to 74); Cb1=Cb5S1 (794 to 73); and Ce=Ce5S1 (positions −24 to +186, ref. 7). Only the Cb5S1 sequence is shown in its entirety. The 5S RNA coding region is indicated in upper case letters; flanking sequences are given in lower case. Conserved positions are indicated by (*), non-conserved positions are indicated by the appropriate nucleotide, and gaps are indicated by a blank space ( ). NOTE: The published sequence of the 119 nucleotide 5S RNA (26), and the published sequence of *C. elegans* 5S DNA were determined to be identical (7). However, there is a typographical error in Fig. 1C of ref. 7, which shows an extra G residue at position +114. This G has been removed for the comparisons shown in the Figure (also corrected, EMBL accession no. X16224).

denatured plasmid templates as described (24,25). Sequence analysis was done using the SEQNCE program of Delaney (Delaney Software, Vancouver, Canada).

Transcription reactions

Embryonic extracts were prepared from *C. elegans* or *C. briggsae* and transcription reactions performed essentially as previously reported (9). Two templates were present in each reaction: test *C. briggsae* templates pCbe5S1 or pCbe5S0.7 (0.1 nM) and pCe5SMAXI (a reference template). pCe5SMAXI is a derivative of pCe5S1 (*C. elegans* 5S DNA) lacking 185 bp of 3′ flanking sequence including the wild-type transcription termination signals. It yields a 165 nt transcript (internal standard) in the transcription reactions. All templates were used as covalently-closed, circular plasmids at 0.1 nM. Final reaction volumes were 25 μl and incubations were for 1 hour at 25°C. Reactions were terminated by the addition of SDS to 1% and phenol-chloroform extracted. Nucleic acids were ethanol precipitated, resuspended in 100% formamide and fractionated on 10% polyacrylamide, 7M urea sequencing gels (24).

RESULTS

Isolation of C. briggsae 5S DNA repeats

The 5S RNAs of *C. elegans* and *C. briggsae* have previously been shown to be identical in nucleotide sequence (26). We took advantage of this sequence identity by using a *C. elegans* probe to visualize the *C. briggsae* genomic sequences encoding 5S RNA. Southern blots of *C. briggsae* genomic restriction digests were probed with a 128 bp fragment of pCe5S1 corresponding to the *C. elegans* 5S RNA coding region. For this, a 128bp *Hind*III fragment (positions −8 to +120, unpublished results) from a 5′ deletion/substitution derivative of pCe5S1 (7) was obtained using the exo III-SI protocol of Henikoff (23). This 5S RNA coding probe identified two *C. briggsae* genomic repeat families (1 kb and 0.7 kb *Hind*III fragments) which potentially encoded 5S RNA (data not shown).

In order to isolate representative members of these two repeat families, 1 kb and 0.7 kb size fractions of a *C. briggsae* genomic *Hind*III digest were isolated, ligated into pUC13 and screened for hybridization to the 5S coding probe. Two recombinant plasmids, pCb5S1 and pCb5S0.7, were chosen as representative clones and used for all subsequent analyses. The putative 5S RNA coding sequences of both repeats (identified by hybridization to the 5S coding probe) are located immediately 5′ to the *Hind*III site which separates single repeats (see the sequence data in Fig. 1). Apart from this 5S RNA coding region, Cb5S1 and Cb5S0.7 share no common restriction sites and cross-hybridize poorly.

Genomic organization of C. briggsae 5S RNA genes

Our preliminary genomic Southern blots indicated that the 1 kb and 0.7 kb 5S DNA repeats of *C. briggsae* are organized in tandem clusters, but do not distinguish between the mixing of these two repeat families within a single tandem cluster and their separation into independent, repeat-specific clusters. In order to resolve this ambiguity, we used spacer

C) Sequence alignment of conserved spacer sequences. Cb7=Cb5S0.7 (positions 243 to 498). Because the SL RNA sequence is in an orientation opposite to that of 5S RNA, Cb1 shows the *complementary* strand of the Cb5S1 sequence (positions 427 to 695) given in Figure 1. Similarly, Ce gives the complementary strand of the Ce5S1 sequence (positions −126 to −348) in ref. 7. The putative SL RNA sequence is indicated in upper case letters; flanking sequences are given in lower case. NOTE: There is a sequencing error in the original published sequence of the *C. elegans* 5S DNA; it should have an extra C residue at position −229 in Fig. 1C of ref. 7 (unpublished results; see also the gel data in Fig. 5A, ref. 16 for independent sequence data). This correction (an extra G in the complement) has been added in the present Figure.

sequences specific to each repeat class to probe a battery of *C. briggsae* genomic restriction digests. The results show that each repeat family is organized independently of the other in the *C. briggsae* genome. Enzymes which cut in Cb5S1 but not in Cb5S0.7 all leave the 0.7 kb repeat family as a high mol wt array while reducing the 1 kb repeat family to its unit length (data not shown). The Cb5S1 and Cb5S0.7 repeat families are therefore not interspersed within a single tandem cluster, but are organized into mutually exclusive tandem clusters of homogeneous repeat units.

The relative hybridization signals from the *C. briggsae* genomic Southerns suggest that the 1 kb 5S DNA repeat is present in higher copy number than the 0.7 kb repeat. We have compared the intensity of the genomic hybridization signals to plasmid standard curves and estimate approximately 65 copies of Cb5S1 and 20 copies of Cb5S0.7 per *C. briggsae* haploid genome (data not shown).

In addition to 5S DNA tandem clusters, some eukaryotes also possess single copies of 5S DNA dispersed throughout their genomes (2). The lack of additional hybridizing bands observed in Southern blots suggests that the *C. briggsae* genome contains few, if any dispersed copies of either repeat family.

Structural features of the C. briggsae 5S DNA repeats

In order to examine the relationship between the two *C. briggsae* 5S DNA repeat families, we determined the complete nucleotide sequences of Cb5S1 and Cb5S0.7 (Figure 1). Both *C. briggsae* 5S DNA repeats share general sequence features previously described for the *C. elegans* 5S DNA repeat (7). In each case, the 5S RNA coding sequence is separated from the RNA pol III transcription termination signal by a *Hind*III site. The 3′ flanking sequence is AT-rich with numerous runs of consecutive A's and T's. A search for internal sequence repetition in these repeats revealed a 39 bp sequence which is perfectly repeated twice and partially a third time in Cb5S0.7 (lower case nucleotides in Figure 1b). No internal repetition was observed in Cb5S1.

Comparison of Cb5S1 and Cb5S0.7 sequences confirms the lack of extensive sequence conservation suggested by the previous hybridization results. Only the 5S RNA coding region and 3′ flanking sequences are directly conserved between the two repeats (see figure 2). The 5S RNA coding sequence itself is perfectly conserved, while the 3′ flanking 40 bp is slightly divergent. Further downstream, and upstream of the 5S RNA coding region, the two repeats are not significantly similar.

A second region of partially conserved sequence (approximately 250 bp in length) is revealed when Cb5S1 is compared to the complement of Cb5S0.7 (see figure 2). The conservation is not as complete as noted above for the 5S RNA coding region, and is punctuated by blocks of divergent sequence.

Comparison of C. elegans and C. briggsae 5S DNA repeat sequences

Comparison of *C. elegans* and *C. briggsae* 1 kb 5S DNA repeat sequences reveals two blocks of extended sequence identity (Figure 2). One of these corresponds and is limited to the 5S RNA coding region and 15 bp of 3′ flanking sequence. The other is located upstream in the spacer region (−273 to −177 in Ce5S1, ref 7; and 510 to 607 in Cb5S1, see Figure 2). In this spacer sequence, a core of 118 bp is perfectly conserved while flanking sequences diverge rapidly on one side (within 20 bp) and slowly on the other (within 60 bp). The remainder of these repeats share little sequence identity (Please note the corrections of the original published pCe5S1 sequence; see Fig 2b, 2c).

Both of these conserved sequences are also found in the *C. briggsae* 0.7 kb 5S DNA repeat (Figure 2). The 5S RNA coding sequence is perfectly conserved while the spacer

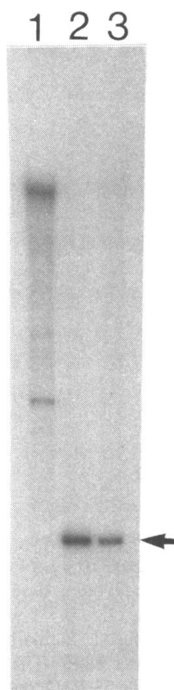

Fig. 3. *Transcription of C. briggsae 5S RNA in a C. elegans cell-free transcription system.*
The transcription of: lane 1 = pCb5S1; lane 2 = pCbe5S1 and lane 3 = pCbe5S0.7 templates. pCbe5S1 and pCbe5S0.7 were constructed by ligating, into the *Hind*III site 3′ to the 5S RNA coding region of pCb5S1 or pCb5S0.7, the 450bp *Hind*III-*Bam*HI fragment of pCe5S1, which contains sequences immediately 3′ to the *C. elegans* 5S RNA (including the TTTTT pol III terminator, see ref. 7). These templates were transcribed, and the resulting labelled RNAs were processed as described in the text. The arrow indicates the position of 5S RNA.

region is partially conserved in the complementary orientation, as noted in the previous section.

Are both C. briggsae 5S DNA repeat families functional?

The identification of two related but distinct 5S DNA repeat families in the *C. briggsae* genome raises the possibility that one of these families may be partially or completely inactive *in vivo*. The perfect identity of the 5S RNA coding sequences of both repeat families makes it impossible to examine their individual expression *in vivo*. We have therefore examined their ability to program 5S RNA transcription in a *C. elegans* cell-free extract.

Both Cb5S1 and Cb5S0.7 were cloned as *Hind*III genomic fragments, which resulted in the separation of 5S RNA coding sequences from the putative RNA pol III termination signals. The transcription of pCb5S1 and pCb5S0.7 results in the appearance of several large transcripts *in vitro* (Figure 3), presumably due to termination at alternative sites in flanking vector sequences. Rather than re-isolate the *C. briggsae* genomic repeats, Cb5S1 and Cb5S0.7 were recloned adjacent to the RNA pol III termination signals carried by Ce5S1 to produce the templates pCbe5S1 and pCbe5S0.7 (see Figure 3, legend). The 3′ flanking signals are relatively conserved between the two species (see Fig. 2b), and are thus unlikely to influence transcription efficiency under our reaction conditions. 5S RNA

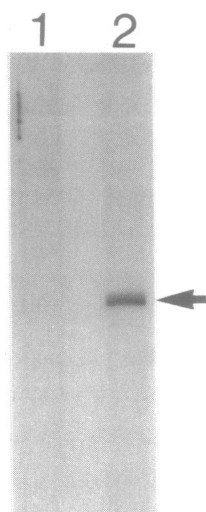

Fig. 4. *In vivo expression of SL RNA coding regions*
Northern blots probed with oligonucleotide probes. 10 μg. of total *C. briggsae* RNA was fractionated, blotted and hybridized to 'repeat-specific' oligonucleotide probes as described in Materials and Methods. To detect putative SL RNA transcripts without cross-hybridization, oligo 0.7-1 was synthesized as the complement to nucleotides 339 to 357 of the Cb5S0.7 sequence given in Fig. 1. Similarly, oligo 1-1 was synthesized as nucleotides 565 to 583 of the Cb5S1 sequence in Figure 1. The observed RNA has the same gel mobility as the *C. elegans* transcript described by Krause and Hirsh (16), between our 5S RNA and tRNA markers on gels (data not shown). Lane 1, oligo 0.7-1; lane 2, oligo 1-1; the arrow indicates the putative SL RNA transcript.

transcription is also relatively insensitive to changes in 3′ flanking sequences (Nelson and Honda, unpublished results). Both hybrid templates are efficiently transcribed *in vitro* (Figure 3) to produce 5S RNA, suggesting that both repeats can function as 5S RNA templates *in vivo*. Both templates are also efficiently transcribed in *C. briggsae* extracts (not shown).

While this work was in progress, we learned of the results of Krause and Hirsh (16) describing an abundant *C. elegans* transcript (Spliced Leader or SL RNA) of about 100 nucleotides (nt), which appears to donate its 5′ 22 nt (SL) to the 5′ termini of a variety of cellular mRNAs via a trans-splicing mechanism. Surprisingly, this 100 nt SL RNA coding sequence appears in the *C. elegans* 5S DNA repeat, 176 bp upstream of, and in the opposite orientation to, the 5S RNA coding sequence.

The SL RNA coding region corresponds precisely to the spacer sequences which are perfectly conserved betwen the *C. elegans* and *C. briggsae* 1 kb 5S DNA repeats. This region is also conserved, albeit imperfectly and in the opposite orientation, in the *C. briggsae* 0.7 kb 5S DNA repeat. Significantly, the 22 nt which appears to be trans-spliced, and the adjacent splice donor site are perfectly conserved in Cb5S0.7, suggesting that, if transcribed, the divergent SL RNA transcript could function in the trans-splicing process.

In order to determine if the *C. briggsae* 5S DNA repeats are transcribed to produce an SL RNA transcript *in vivo*, we chose oligonucleotides complementary to SL RNA transcript sequences which are different in the 1.0 and 0.7 kb sequences (see Figure 4, legend). These 'repeat-specific' oligonucleotides were used to probe Northern blots of total *C. briggsae* RNA fractionated on denaturing polyacrylamide gels. The results (Figure 4)

show that while the 1 kb repeat-specific SL RNA probe detects an abundant 100 nt RNA analogous to that described in *C. elegans*, the 0.7 kb repeat-specific probe does not reveal a detectable transcript. Somehow then, changes in SL RNA coding sequences and/or flanking sequences in the 0.7 kb repeat have resulted in the inactivation of the SL RNA transcription unit.

DISCUSSION

In an effort to use an evolutionary approach (see also refs 26, 27 and refs. therein) to identify the functionally important features of *C. elegans* 5S RNA gene structure and organization, we have characterized the homologous genes of *C. briggsae*, a closely related nematode species. In both species, the haploid genome contains approximately 100 copies of 5S RNA coding sequence. In *C. elegans*, these genes comprise a single homogeneous 1 kb genomic repeat family which is tandemly clustered on the right arm of linkage group V (7,8). In contrast, *C. briggsae* 5S RNA genes are split into two discrete repeat families; approximately 65 copies of a 1 kb repeat and 20 copies of a 0.7 kb repeat. Each repeat family is homogeneous in size and structure, and is organized independently of the other. However, we do not know whether each repeat family exists within one (as in *C. elegans*) or a few separate tandem clusters. Members of each repeat family also appear to be homogeneous in size and sequence, as judged from the lack of detectable variants following digestion and Southern blotting, using a large number of different restriction enzymes. However, we cannot rule out minor differences in DNA sequence between individual repeats.

Comparison of *C. elegans* and *C. briggsae* 5S DNA repeat sequences suggests that they diverged from a common ancestral repeat; all three share two extended blocks of highly conserved sequence. The relative orientations of these conserved sequences is the same for the *C. elegans* and *C. briggsae* 1 kb repeats, suggesting that they most closely resemble the ancestral repeat. The complementary orientation of the conserved spacer sequences relative to the 5S RNA coding region in the *C. briggsae* 0.7 kb repeat indicates that it may have diverged by a complex series of events including the inversion of spacer sequences. The organization of the two *C. briggsae* 5S DNA repeat families into mutually exclusive tandem clusters may serve to maintain homogeneity within each repeat family while allowing the two families to diverge from one another (4).

This work was undertaken on the assumption that the conservation of 5S DNA repeat sequence reflects its functional importance. As anticipated, all three repeats contain perfectly conserved 5S RNA coding sequences. While we cannot distinguish between the 5S RNA transcripts derived from the *C. briggsae* 1 kb and 0.7 kb 5S DNA repeat *in vivo*, the observation that both repeats program the efficient transcription of 5S RNA in a *C. elegans* or homologous *C. briggsae* cell-free extract suggests that both could represent functional 5S RNA genes.

Surprisingly, the conserved spacer sequence also appears in the *C. elegans* cellular RNA pool, as an abundant SL RNA transcript which is thought to participate in a trans-splicing process (16,17). Our results indicate that an analogous transcript is also found in *C. briggsae*. Northern blot analysis, using repeat-specific oligonucleotide probes complementary to putative SL RNA transcripts, suggests that the 1 kb repeat is transcribed *in vivo*, while the 0.7 kb repeat appears to be transcriptionally inactive. This is also reflected in the sequence conservation: the SL RNA coding regions of the *C. elegans* and *C. briggsae* 1 kb repeats are identical. This perfect conservation presumably reflects the functional importance of this sequence in the two nematode species. Other recent work has

demonstrated the presence of sequences implicated in trans-splicing; these sequences are highly conserved across the phylum nematoda (17, 29).

The corresponding, SL RNA-like region of the *C. briggsae* 0.7 kb repeat differs extensively in sequence, and is apparently not expressed. The SL transcript of *C. elegans* may be transcribed from either the 5S DNA tandem cluster, or from dispersed genomic sequences which hybridize to the 5S DNA repeat (7). However, the apparent lack of a significant number of such dispersed sequences in the *C. briggsae* genome, as well the observed complete sequence conservation relative to the *C. elegans* sequence, suggest that the 1 kb 5S DNA repeat itself encodes both the 5S and SL RNAs *in vivo*. We have not observed SL RNA transcription from any of these 5S DNA templates in the *C. elegans* cell-free extract; either SL RNA is transcribed by RNA polymerase II (which is inhibited by the levels of alpha-amanitin present in our transcription reactions), or essential components are missing from the extract. It is also formally possible that very short transcripts were made but not detected in these experiments.

Are there other conserved blocks of sequence which might reflect functionally important DNA e.g. promoter type sequences for the the two conserved transcripts? We have identified four short (10−15 bp) blocks of sequence which are conserved in the region between the 5′ ends of the two transcripts in the *C. elegans* and *C. briggsae* 1 kb repeats. While these conserved blocks are colinear in each repeat, their positions relative to the two transcription units and to one another vary considerably. At least one of these sequence elements appears to be involved in modulating 5S RNA transcription in the *C. elegans* cell-free extract (Nelson and Honda, unpublished results).

The remainder of the 5S DNA repeat sequences share little sequence identity, although some general features are conserved. Downstream of the 5S RNA coding region, all three repeats are AT-rich, with numerous runs of consecutive A and T residues. The region separating the SL and 5S RNA coding sequences of the *C. elegans* and *C. briggsae* 1 kb repeats is, on the other hand, GC-rich. The inversion of spacer sequences noted in the *C. briggsae* 0.7 kb repeat maintains a GC-rich sequence upstream of the SL coding region, while the region upstream of the 5S RNA coding region is replaced by an AT-rich sequence.

The efficient transcription of 5S RNA from all three 5S DNA repeats indicates that the replacement of GC-rich 5′ flanking sequences with AT-rich sequences in the *C. briggsae* 0.7 kb repeat does not dramatically affect the expression of this transcription unit. In contrast, the inversion of homologous spacer sequences in the 0.7 kb repeat appears to have inactivated SL RNA transcription, despite the maintenance of the GC-rich 5′ flanking sequence. Further studies will be required to determine if these and other sequences are important in transcription, the maintenance of repeat homogeneity, or other functions.

ACKNOWLEDGMENTS

We thank Ian Scott for able technical assistance, Mike Krause for communicating his results to us prior to publication, and many colleagues for their support, and invaluable comments and discussion.

This work was supported by a grant from NSERC Canada to B.M.H. D.W.N. was the recipient of an NSERC Canada studentship.

*To whom correspondence should be addressed

+Present address: Department of Cellular and Developmental Biology, Harvard University, Cambridge, MA 02138, USA

REFERENCES

1. Long,E.O. and Dawid,I.B. (1980) Ann. Rev. Biochem. *49*, 727–764.
2. Rosenthal,D.S. and Doering,J.L. (1983) J. Biol. Chem. *258*, 7402–7410.
3. Selker,E.V., Yanofsky,C., Driftmier,K., Metzenberg,R.L., Alzner-DeWeerd,B. and RajBhandary,U.L. (1981) Cell *24*, 819–828.
4. Arnheim,N. (1983) In Nei,M. and Koehn,R.K. (eds), Evolution of Genes and Proteins. Sinauer Associates, Inc., Sunderland, MA., pp.38–61.
5. Geiduschek,E.P. and Tocchini-Valentini,G.P. (1988) Ann. Rev. Biochem. *57*, 873–914.
6. Wood,W.B. (1988) The nematode *Caenorhabditis elegans*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
7. Nelson,D.W. and Honda,B.M. (1985) Gene *38*, 245–251.
8. Nelson,D.W. and Honda,B.M. (1986) Can. J. Genet. Cytol. *28*, 545–553.
9. Honda,B.M., Devlin,R.H., Nelson,D.W., and Khosla,M. (1986) Nucl. Acids Res. *14*, 869–881.
10. Fedoroff,N.V. (1979) Cell *16*, 697–710.
11. Samson,M-L. and Wegnez,M. (1984) Nuc. Acid. Res. *12*, 1003–1014.
12. Bodmer,M. and Ashburner,M. (1984) Nature *309*, 425–430.
13. Blackman,R.K. and Meselson,M. (1985) J. Mol. Biol. *188*, 499–515.
14. Henikoff,S. and Eghtedarzadeh,M.K. (1987) Genetics *117*, 711–725.
15. Hirsh,D., Emmons,S.W., Files,J.G. and Klass,M.R. (1979) In Axel,R., Maniatis,T. and Fox,C.F. (eds), Eukaryotic Gene Expression. Academic Press, NY. pp.205–218.
16. Krause,M. and Hirsh,D. (1987) Cell *49*, 753–761.
17. Bektesh,S.L., Van Doren,K. and Hirsh,D.I. (1988) Genes and Devel. *2*, 1277–1283.
18. Maniatis,T., Fritsch,E.R. and Sambrook,J. (1982) Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
19. Brenner,S. (1974) Genetics *77*, 71–94.
20. Emmons,S.W., Klass,M.R. and Hirsh,D. (1979) Proc. Natl. Acad. Sci. (USA) *76*, 1333–1337.
21. Vieira,J. and Messing,J. (1982) Gene *19*, 259–268.
22. Rigby,P.W., Dieckmann,M., Rhodes,C. and Berg,P. (1977) J. Mol. Biol. *113*, 237–251.
23. Henikoff,S. (1984) Gene *28*, 351–359.
24. Sanger,F., Nicklen,S. and Coulsen,A.R. (1977) Proc. Natl. Acad. Sci USA *74*, 5463–5467.
25. Hattori,M. and Sakaki,Y. (1986) Anal. Biochem. *152*, 323–328.
26. Butler,M.H., Wall,S.M., Luehrsen,K.R., Fox,G.E. and Hecht,R.M. (1981) J. Mol. Evol. *18*, 18–23.
27. Prasad,S. and Baillie,D.L. (1989) Genomics, in press.
28. Zucker-Aprison,E. and Blumenthal,T. (1989) J. Mol. Evol. *28*, 487–496.
29. Takacs,A.M., Denker,J.A., Perrine,K.G., Maroney,P.A. and Nilsen,T.W. (1988) Proc. Natl.Acad.Sci.(USA) *85*, 7932–7936.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.