MOLECULAR AND CELLULAR BIOLOGY, June 1990, p. 2765–2773
0270-7306/90/062765-09$02.00/0

Vol. 10, No. 6

Many Transcribed Regions of the *Onchocerca volvulus* Genome Contain the Spliced Leader Sequence of *Caenorhabditis elegans*

WENLIN ZENG, CLARA M. ALARCON, AND JOHN E. DONELSON*

Howard Hughes Medical Institute, Department of Biochemistry, University of Iowa, Iowa City, Iowa 52242

Received 5 December 1989/Accepted 23 February 1990

Genomic DNAs of the related parasitic nematodes *Onchocerca volvulus* and *Dirofilariae immitis*, and a cDNA library of *O. volvulus*, were examined for the presence of the 22-nucleotide spliced leader (SL) found at the 5′ ends of 10 to 15% of the mRNAs in the free-living nematode *Caenorhabditis elegans*. As in *C. elegans*, genes for the SL RNA are linked to the repetitive 5S rRNA genes of *O. volvulus* and *D. immitis*, but unlike *C. elegans*, they are in the same orientation as the 5S rRNA genes within the repeat unit. In *O. volvulus* the SL sequence is also encoded at more than 30 additional genomic locations and occurs at interior sites within many transcripts. Sequence determinations of four different cDNAs of *O. volvulus*, each containing an internal copy of the SL within a conserved 25mer, and one corresponding genomic DNA clone indicate that this sequence is not *trans* spliced onto these RNAs, but is encoded within the genes. The RNAs of two of these cDNAs appear to be developmentally regulated, since they occur in adult *O. volvulus* but were not detected in the infective L3 stage larvae. In contrast, actin mRNAs are present at all developmental stages, and at least one actin mRNA species contains a *trans*-spliced 5′ SL. The internal locations of the SL in various transcripts and its perfect sequence conservation among parasitic and free-living nematodes argues that it serves specific, and perhaps multiple, functions for these organisms.

The presence of a spliced leader (SL) sequence at the 5′ end of an mRNA was first detected in African trypanosomes (4). In these organisms, and in related protozoa of the family Kinetoplastidae, this 39-nucleotide leader sequence is spliced onto the 5′ ends of all mRNAs (33). The trypanosome SL is encoded in a repetitive DNA element of which there are about 200 copies in *Trypanosoma brucei* (8) and from which is transcribed an 85- to 140-nucleotide RNA species with the SL at its own 5′ end. Current evidence suggests that the SL is transferred, in a *trans*-splicing event, from this small initial transcript to precursor RNAs of all protein-coding genes to yield mRNAs possessing the SL at their 5′ ends (6, 22, 27). The raison d'etre of the trypanosome SL is not known, although it has been suggested that it participates in cleavage of polycistronic RNAs from tandem genes to generate monomer RNA units (17) and contributes a 5′ CAP structure to those mRNAs (18).

Recently, it has been observed that 10 to 15% of the mRNAs of several nematodes, including *Caenorhabditis elegans* and *Brugia malayi*, also possess a 5′ SL of 22 nucleotides whose sequence is unrelated to the 39-nucleotide SL of trypanosomes (16, 28). For example, in *C. elegans* three of the four actin mRNA species contain this nematode SL at their 5′ ends, while the fourth actin mRNA does not (16). Other *C. elegans* mRNAs shown to contain the SL include those encoding ubiquitin, glyceraldehyde-3-phosphate dehydrogenase, a heat shock protein, and two ribosomal proteins (2, 13). The *C. elegans* SL sequence of 5′-GGTTTAATTACCCAAGTTTGAG-3′ is encoded within the spacer region of the repeat unit containing the 5S rRNA gene and is transcribed from the complementary strand to an RNA of about 100 nucleotides. It seems likely that, similar to trypanosomes, a *trans*-splicing event transfers the nematode SL from this primary transcript to the 5′ ends of the appropriate RNAs (1, 5, 29, 32). In contrast to trypanosomes, however, this splicing mechanism must be able to identify only precursor RNAs destined to be *trans* spliced and must distinguish *trans*-splicing signals from those for *cis* splicing of introns present in some genes of nematodes (3, 25).

Onchocerca volvulus is one of several filarial nematodes responsible for different forms of human filariasis, which are major diseases in many tropical countries. Long-term infection by this parasite is one of the leading causes of blindness in Africa. It has been calculated from a comparison of partial sequences of their large-subunit rRNAs that *O. volvulus* and *C. elegans* are phylogenetically related with a K_{nuc} of 0.160 (12). Since drug interference with *trans* splicing could be useful therapeutically (33), we examined the *O. volvulus* genome for the presence of the *C. elegans* SL. It turned out that this sequence is not confined to just the 5S rRNA gene repeat, as it is in *C. elegans*, but is distributed among 30 or more genomic locations. Likewise, in a related parasitic nematode, the dog heartworm *Dirofilariae immitis*, the sequence also occurs both adjacent to the 5S rRNA gene and at other genomic sites. Since a specific 22-mer sequence should occur at random with a frequency of about $(1/4)^{22}$, or once every 1.76×10^{13} base pairs (bp), these sequences occur far more often than expected statistically in the *O. volvulus* haploid genome of 1.5×10^8 bp (10). When examined in more detail, the conserved sequence was found to be 25 nucleotides long in four *O. volvulus* cDNAs and to occur at interior sites, one of which was near the 3′ poly(A) tail and not *trans* spliced onto the RNA.

MATERIALS AND METHODS

Sources of nucleic acids. DNA and total RNA were extracted from adult *O. volvulus* as described previously (10). DNA was isolated from adult *D. immitis* parasites (a gift from John McCall) by the same protocol. DNAs from *C. elegans* and *B. malayi* were gifts from X. Huang and Larry McReynolds, respectively. The previously described adult *O. volvulus* cDNA library was constructed from poly(A)$^+$ RNA of adult parasites collected from patients near

* Corresponding author.

Touboro, West Cameroon, Africa (10). Briefly, the cDNA library was prepared by synthesis of first- and second-strand cDNAs by using avian myeloblastosis virus reverse transcriptase and *Escherichia coli* DNA polymerase I, respectively, digestion with S1 nuclease, addition of *Eco*RI linkers, and ligation of the duplex cDNAs into λgt11. A similarly constructed cDNA library was also prepared from RNA extracted from about 20,000 infective-stage *O. volvulus* L3 larvae. A library of partial *Sau*3A-digested genomic DNAs of *O. volvulus* was constructed in bacteriophage λEMBL3 by using DNA from parasites collected in Mali. Synthetic oligonucleotides were prepared on a Beckman oligonucleotide synthesizer.

Methods. Southern blots (26), Northern (RNA) blots (30), ^{32}P labeling of DNA fragments (11), screening of cDNA and genomic libraries (15, 19), and DNA sequence determinations by the chemical cleavage method (21) or dideoxynucleotide method (14) were conducted as described previously. Subcloning of restriction fragments from recombinant bacteriophage to plasmids and other standard procedures were as described previously (19).

Polymerase chain reaction (PCR) amplification of actin mRNA sequences was conducted by synthesizing first-strand cDNA from adult *O. volvulus* poly(A)$^+$ RNA with reverse transcriptase and oligo(dT) (10). This first-strand cDNA served as the template for the PCR with an SL oligonucleotide and an oligonucleotide containing a consensus complementary actin coding sequence about 200 nucleotides downstream from the ATG start codon (24). PCR amplifications of OV3 and OV9 sequences were conducted with oligonucleotides that flank the internal SL sequences and genomic DNA or recombinant phage DNA that was extracted from the cDNA libraries.

RESULTS

Distribution of the SL sequence in *O. volvulus* and *D. immitis* DNAs. To establish whether the genome of *O. volvulus* encodes the nematode SL sequence, we used a complementary oligonucleotide as a probe in Southern blots. Figure 1A shows the hybridization of this complementary sequence, 5′-dCTCAAACTTGGGTAATTAAACC-3′, to DNAs of *C. elegans* and *O. volvulus* restricted by three different enzymes. The probe hybridizes predominantly to a 930-bp fragment in the *Ava*II, *Dde*I, and *Hin*dIII digests of *C. elegans* DNA, as predicted from the earlier demonstration that this sequence is encoded within the repeat unit containing the *C. elegans* 5S rRNA gene (28). In contrast, the probe hybridizes to more than 30 different fragments in the digests of *O. volvulus* DNA. Although some bands in the *O. volvulus* DNA are more prominent than others, it is not obvious from these digests that the sequence is encoded on a repetitive DNA element in *O. volvulus*, as it is in *C. elegans*.

To determine whether the 5S rRNA gene and the 22-mer sequence are linked in the DNAs of *O. volvulus* and other parasitic nematodes, we used the *C. elegans* 5S rRNA gene as a hybridization probe in Southern blots (Fig. 1B). To obtain this probe, the 930-bp region of an agarose gel containing a *Hin*dIII digest of *C. elegans* DNA (Fig. 1A) was eluted, these fragments were subcloned into plasmid IBI30, and recombinant plasmids containing the 22-mer sequence were identified. About 80% of the sequence of one of these *C. elegans* 930-bp inserts was determined and found to contain the 5S rRNA gene and the 22-nucleotide SL sequence on complementary strands as reported previously

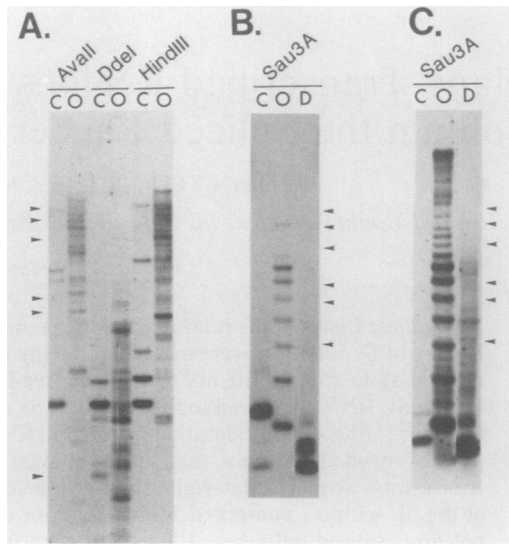

FIG. 1. (A) Autoradiogram of a Southern blot in which 5 μg of genomic DNA from *C. elegans* (lanes C) or *O. volvulus* (lanes O) was digested with *Ava*II, *Dde*I, or *Hin*dIII and probed with an oligonucleotide complementary to the 22-nucleotide SL sequence of *C. elegans*. (B and C) Autoradiograms of Southern blots in which *Sau*3A digests of genomic DNAs from *C. elegans* (lanes C), *O. volvulus* (lanes O), or *D. immitis* (lanes D) were probed with a restriction fragment containing the *C. elegans* 5S rRNA gene (panel B) or the 22-mer oligonucleotide (panel C). Arrowheads indicate the positions of DNA *Hin*dIII fragments (panel A) or ladders of 1-kb markers (panels B and C).

(16). The spacer region of this cloned repeat possesses a few point changes from the published *C. elegans* 5S rRNA gene repeat, as might be expected for a different clone of the same repetitive element. Since the sequence of 5S rRNA is highly conserved across species, a 224-bp *Pvu*I fragment from this cloned repeat containing the 5S rRNA gene was used to probe *Sau*3A digests of DNA from three different nematodes (Fig. 1B). Hybridization occurs to a major fragment in *C. elegans* DNA (lane C), a ladder of fragments in *O. volvulus* DNA (lane O), and three small fragments in DNA from the dog heartworm, *D. immitis* (lane D).

The Southern blot of Fig. 1B was washed and reprobed with the complementary SL oligonucleotide (Fig. 1C). The oligonucleotide hybridized to the same ladder of *O. volvulus* fragments as the 5S rRNA probe did, but it also hybridized to many additional fragments, suggesting that the SL sequence is both linked to the 5S rRNA genes and dispersed throughout the genome. The hybridization pattern of *D. immitis* DNA likewise indicates that the SL sequence may be both linked to the 5S rRNA genes and scattered elsewhere in the genome. In contrast, the blot indicates that in *C. elegans* the SL is encoded predominantly, and perhaps only, on a repetitive DNA element.

The 5S rRNA genes and the SL sequence are linked in *O. volvulus* and *D. immitis*. To verify linkage of the 5S rRNA genes and the SL sequence in *O. volvulus* and *D. immitis*, the 5S rRNA gene repeats were cloned from the DNAs of these organisms and their sequences were determined. The 5S rRNA gene repeat of *O. volvulus* was obtained by eluting the 400-bp region of an agarose gel containing a *Sau*3A digest and cloning the fragments into the *Bam*HI site of plasmid IBI30. The equivalent DNA repeat of *D. immitis* was similarly cloned after elution of the appropriate regions of both a

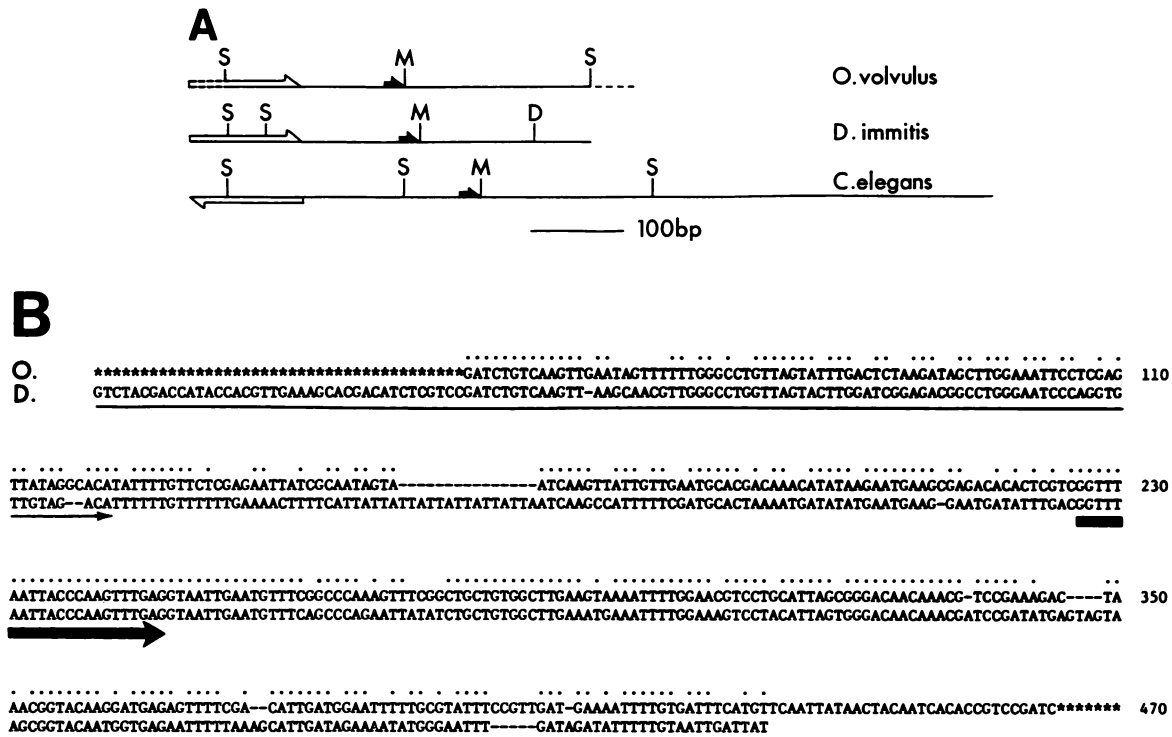

FIG. 2. (A) Genomic organization of DNA repeat units containing the genes for 5S rRNA and the SL sequence of *O. volvulus*, *D. immitis*, and *C. elegans*. Open arrows represent the 5S rRNA genes, and black arrows show the 22-nucleotide SL sequence. The arrows point in the direction of transcription. Dashed regions at both ends of the *O. volvulus* repeat indicate the segment of this repeat that was not cloned. Restriction sites for *Sau*3A (S), *Mnl*I (M), and *Dde*I (D) are shown. (B) Sequence comparison of the DNA repeat units of *O. volvulus* (O) and *D. immitis* (D) that were cloned as indicated by the solid lines in panel A. The 5S rRNA genes are indicated by the narrow underline arrow, and the SL sequences are indicated by the bold underline arrow. The two sequences are aligned at the start of the 5S rRNA genes. Dots indicate conserved nucleotides, and dashes indicate deletions or insertions to maximize sequence homology. The asterisks at both ends of the *O. volvulus* sequence indicate that the repeat unit continues in both directions from the segment whose sequence was determined.

*Taq*I digest and a *Dra*I digest. Recombinant clones of these fragments were screened with the *Pvu*I fragment containing the *C. elegans* 5S rRNA gene to identify specific clones with the desired inserts. The sequences of these inserts were then determined (Fig. 2).

Similar to *B. malayi* and *Ascaris lumbricoides* (23, 28), and in contrast to *C. elegans* (16), the 5S rRNA gene and the SL sequence are linked on the same strand in both *O. volvulus* and *D. immitis*. In the *O. volvulus* DNA repeat, a spacer region of 104 nucleotides separates the 5S rRNA gene and the SL RNA gene, whereas in *D. immitis* the same region is 15 nucleotides longer. The *D. immitis* sequence (Fig. 2B) is a composite of the sequences of overlapping *Taq*I and *Dra*I fragments and represents the complete sequence of this DNA repeat. The *O. volvulus* sequence, however, is derived from a single *Sau*3A fragment and is only a portion of the complete DNA repeat unit. No attempt was made to clone the entire repeat, since the sequence of this *Sau*3A fragment contains both the 5S rRNA gene and the SL sequence. Comparison of the two repeat sequences reveals that the SL RNA genes are more highly conserved than are the sequences of the 5S rRNA genes. This observation also holds when the *C. elegans* sequence is added to the comparison (data not shown). The perfect sequence conservation of the SL among the four nematodes is in contrast to the SL of protozoa of the family Kinetoplastidae, whose 39-nucleotide sequence varies at two to seven positions among related organisms (7, 9).

Some transcripts of *O. volvulus* contain internal copies of the SL sequence. In a Northern blot of total *O. volvulus* RNA (data not shown), the 22-mer probe appeared to recognize many RNA species, resulting in a faint smear of hybridization from which weak bands appear. Therefore, a cDNA library of adult *O. volvulus* poly(A)$^+$ RNA (10) was probed with the 22-mer oligonucleotide to identify specific RNAs that possess this sequence. During screening of about 200,000 phages of an amplified library, the probe hybridized to the DNA of about one recombinant phage in 17,000, or 0.006%. Four of these cDNAs were selected at random and subcloned into plasmid IBI30 for further characterization. The complete sequences of these cDNAs, OV3 (490 bp), OV9 (656 bp), OV5 (520 bp), and OV1 (905 bp), were determined and are shown in Fig. 3.

None of the four cDNAs contains the SL at the 5′ end. Each, however, has one internal copy of the perfect SL sequence (Fig. 3, solid underline), which is preceded in each case by three A residues, for a total conserved sequence of 25 nucleotides. These three A's do not precede the SL in the 5S rRNA gene repeat. Likewise, none of the four cDNA sequences immediately after the SL resembles the sequence downstream of the SL in the 5S rRNA gene repeat. The SL in two of the cDNAs is followed by the consensus dinucleotide, GT, of an intron donor site, whereas in the other two cDNAs it is not. None of the four cDNAs contains a continuous open translation reading frame in any of the six potential frames. Only OV9 has a long poly(A)$^+$ sequence at the 3′ end of the strand containing the SL sequence, although the 3′ ends of the complementary strands of the other

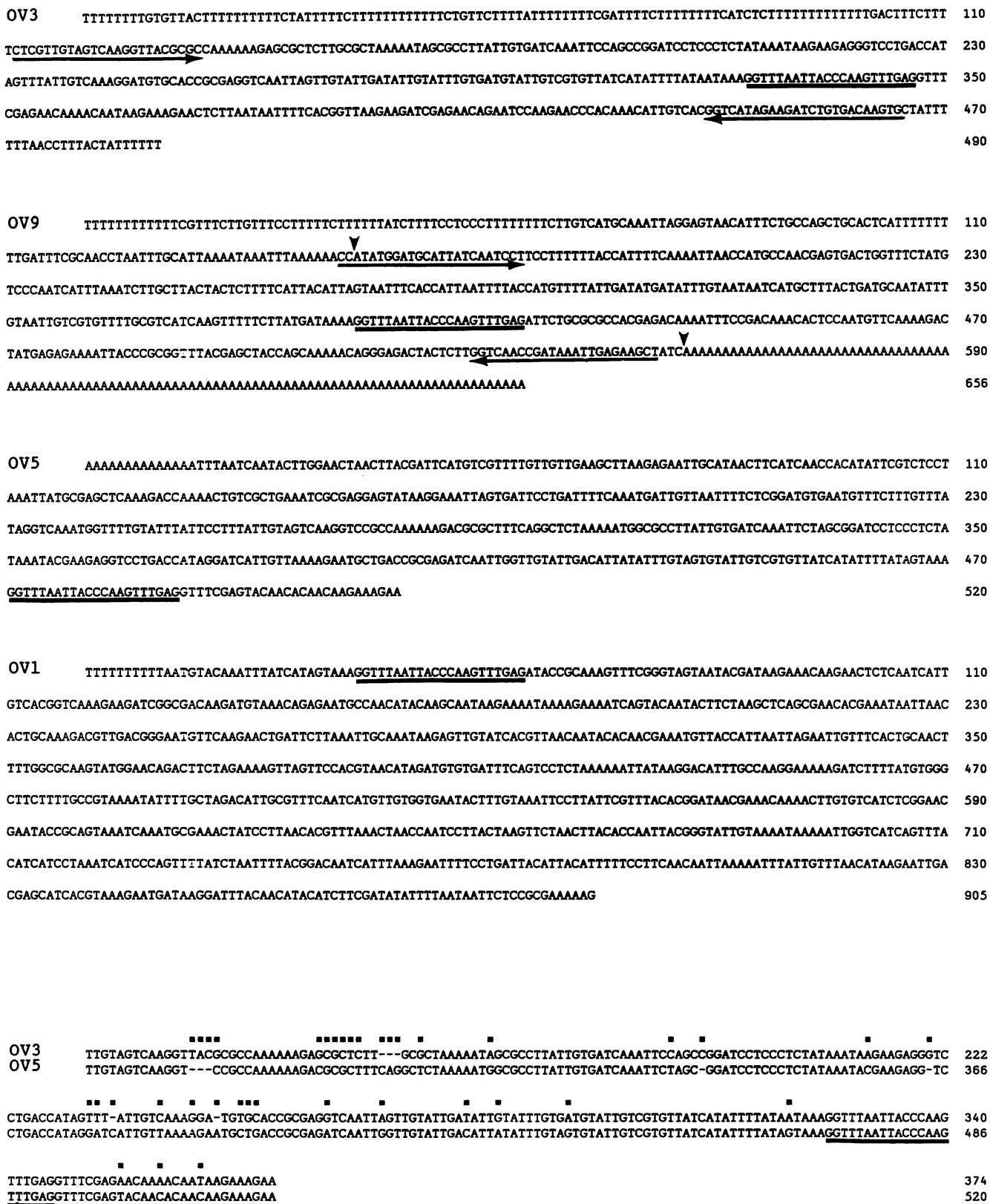

FIG. 3. Complete sequences of four different cDNA clones, OV3, OV9, OV5, and OV1, each of which contains an internal copy of the conserved SL sequence (solid underline). Horizontal arrows in the OV3 and OV9 sequences indicate oligonucleotides used in the PCR experiment shown in Fig. 5. The sequence of the region between the two vertical arrowheads in the OV9 sequence was also determined in an OV9 genomic DNA clone by using oligonucleotide primers indicated by the arrows and by oligonucleotides corresponding to the two strands of the SL sequence. At the bottom of the figure is a comparison of the similar regions of the OV3 and OV5 cDNA sequences. Symbols: –, spaces inserted to maximize similarity; ■, positions that are not identical.

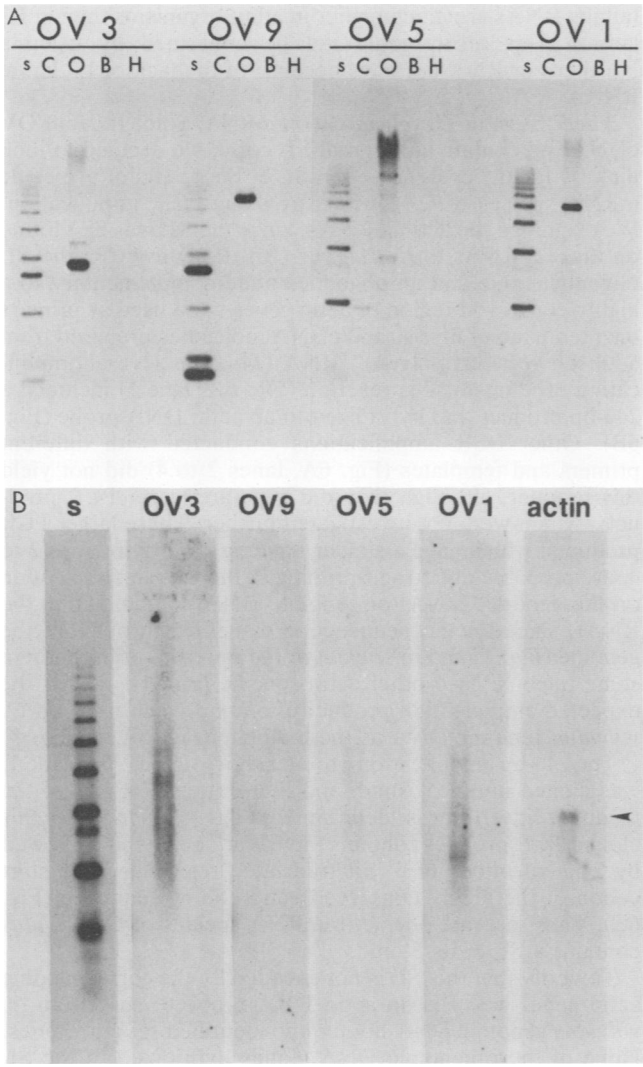

FIG. 4. (A) Autoradiograms of Southern blots in which 5 μg of DNA from *C. elegans* (lanes C), *O. volvulus* (lanes O), *B. malayi* (lanes B), and humans (lanes H) was digested with *Bam*HI and probed with OV3, OV9, OV5, and OV1 cDNAs. (B) Autoradiograms of Northern blots in which 20 μg of *O. volvulus* total RNA was probed with the four OV cDNAs and with an actin DNA probe. The arrowhead indicates the 1.9-kb actin mRNA. Lanes labeled s contain a ladder of 1-kb markers.

three cDNAs terminate in short stretches of A's. Curiously, the OV3 and OV9 cDNA strands shown in Fig. 3 each contain a 5′ T-rich segment that is periodically interrupted by other nucleotides. Similar T- or A-rich cDNA regions have been detected in other *O. volvulus* cDNAs (31) and may be a general feature of this organism whose significance is not clear.

OV3, OV9, and OV1 cDNAs have no significant sequence similarity other than the conserved 25-mer sequence and this T-rich segment. However, OV3 and OV5 contain substantial sequence similarity within the sequenced regions surrounding the SL (Fig. 3, bottom). This comparison, in which the sequences are aligned to maximize their similarity, demonstrates that the two cDNAs come from related genes, a result also indicated by genomic Southern blots (Fig. 4A). On the 5′ sides of the SL in the comparison, the similarity between these two cDNAs extends to just after the T-rich segment at the beginning of OV3. On the 3′ sides, the similarity extends to the end of OV5. In neither cDNA doe these flanking sequences resemble the regions flanking the SL sequence in the 5S rRNA repeats.

Inspection of the OV9 chromosomal gene. Since the OV9 cDNA contains a 3′ poly(A)$^+$ tail of 100 nucleotides, as would be expected for an RNA that had undergone posttranscriptional polyadenylation, we selected this cDNA sequence to see whether a corresponding sequence exists in *O. volvulus* genomic DNA. A λEMBL3 library of *O. volvulus* genomic DNA was screened with the OV9 cDNA, and several genomic DNA clones were identified. A *Bam*HI fragment of one genomic clone was subcloned into a plasmid and served as the template for dideoxynucleotide sequencing with oligonucleotide primers whose sequences are either indicated by the two arrows in OV9 sequence of Fig. 3 or correspond to the two strands of the SL sequence itself. The sequencing information obtained with these four primers demonstrated that the cloned genomic DNA sequence corresponds to the cDNA in the region between the vertical arrowheads. Beyond the 3′-most vertical arrowhead [where the poly(A)$^+$ tail begins in the cDNA], the genomic DNA sequence continues with a normal distribution of the four nucleotides, indicating that this location is indeed a site of posttranscriptional polyadenylation. A putative polyadenylation signal of GATAAA lies about 15 nucleotides in front of this site. Thus, in this example the cDNA, including its internal 25-mer sequence, is derived from a specific gene sequence in the genomic DNA. In addition, preliminary screening of the genomic DNA library with the other three cDNAs indicated that genomic DNA clones corresponding to these sequences are present in the library.

Southern and Northern blots with the cDNAs as probes. The four OV cDNAs were used as probes in Southern and Northern blots (Fig. 4). When used as Southern probes under conditions of high stringency (Fig. 4A), three of the four cDNAs (OV3, OV9, and OV1) hybridize predominantly to single unique *Bam*HI fragments in *O. volvulus* DNA but do not hybridize to DNA fragments from *C. elegans*, *B. malayi*, or humans. OV5 cDNA is different in that it appears to contain a repetitive element within its sequence that hybridizes to a smear of high-molecular-weight *Bam*HI fragments of *O. volvulus*. In unrelated experiments, this filter was reprobed with an actin cDNA, which indicated that the smear observed in Fig. 4A is not the result of a partial restriction enzyme digest. Note also that OV5 hybridizes weakly to the same *O. volvulus* fragment as the one that OV3 hybridizes to strongly, as would be expected because of their partial sequence similarity (Fig. 3, bottom). Likewise, OV3 hybridizes weakly to the smear of large *Bam*HI fragments containing OV5-like sequences. Of the four cDNAs, only OV5 hybridizes weakly to a *Bam*HI fragment (ca. 0.95 kilobase [kb]) in *C. elegans* DNA.

In Northern blots (Fig. 4B), none of the four cDNAs hybridizes to a clearly discrete RNA species, as does the actin DNA probe in the right panel. Instead, each cDNA hybridizes to a smear of RNAs, out of which loom bands of the more prevalent RNA sizes. In the OV3 lane, for example, are RNA bands of 2.8 and 3.5 kb, whereas in the OV1 lane an RNA band of 1.2 kb is present. Weaker signals were generated by the OV9 and OV5 probes. The presence of the 1.9-kb actin mRNA species indicates that general degradation of RNA molecules of this size class had not occurred during RNA isolation, but, of course, for larger RNAs this could be a factor. Thus, the significance of the RNA smears observed with the OV cDNAs is not clear; it may reflect the

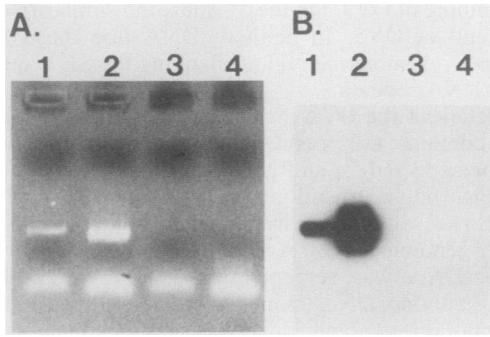

FIG. 5. (A) Agarose gel, stained with ethidium bromide, containing PCR amplification products obtained with OV9-specific primers (lanes 1 and 3) or OV3-specific primers (lanes 2 and 4) and DNA templates extracted from adult (lanes 1 and 2) or L3 larval-stage (lanes 3 and 4) *O. volvulus* cDNA libraries. The lower bands in all four lanes are the oligonucleotide primers. (B) Autoradiogram of a Southern blot in which the DNAs shown in panel A were probed with the SL oligonucleotide.

degradation of much larger RNAs during isolation, or it could indicate processing or instability of the RNAs in vivo.

The OV lanes in the Northern blot were exposed to X-ray film for about 20 times longer than the actin lane. Thus, the OV RNAs are relatively rare mRNAs compared with the abundant actin RNA species, which is consistent with their scarcity in the cDNA library (0.006%). Efforts to obtain significantly longer versions of the four OV cDNAs from the cDNA library were not successful.

Stage-specific expression of internal SL-containing RNAs. Since the four OV cDNA sequences do not contain long open translation reading frames (Fig. 3) and occur within a range of RNA sizes in *O. volvulus* adults (Fig. 4B), their biological roles are not clear. Therefore, we examined these RNAs further by determining whether they are present at a different developmental stage of the life cycle of the parasite. For this investigation we conducted PCR amplifications in which the template DNAs were total phage DNAs extracted from (i) the adult cDNA library and (ii) another cDNA library constructed from RNA of the infective L3 larval stage of *O. volvulus*. The L3 stage is the final developmental stage of the parasite in the blackfly and is the stage that is transmitted to humans, in whom the adult parasites develop. Primers for the PCR amplifications were the OV3 and OV9 oligonucleotides, indicated by horizontal arrows in Fig. 3. These two sets of primers gave the expected PCR products when phage DNA from the adult cDNA library was used as the template (Fig. 5A, lanes 1 and 2) but no visible product when DNA from the L3 library was used (Fig. 5A, lanes 3 and 4). The two phage DNA templates had been normalized in preliminary experiments to give equal-intensity PCR bands when actin-specific oligonucleotides were used as primers (data not shown). Thus, the results in Fig. 5A suggest that more OV3 and OV9 RNA is present in adult parasites than in the L3 larvae. To verify that the OV9 and OV3 PCR products in lanes 1 and 2 contained the SL sequence and to see whether there were very small amounts of these products in lanes 3 and 4, the DNAs in Fig. 5A were probed in a Southern blot with the SL oligonucleotide. The autoradiogram in Fig. 5B demonstrates that the PCR products of the adult cDNA library do indeed contain the SL sequence and that no detectable OV3 or OV9 sequences are present in the L3 cDNA library. This result strongly suggests that the steady-state levels of these two internal SL-containing RNAs are much higher in adult organisms than in L3 larvae. This result could not be confirmed by using a Northern or RNA dot blot, because of the lack of L3 larval RNA.

The 5′ SL of an *O. volvulus* actin mRNA. Since the four OV cDNAs all contain an internal SL copy, we decided to look directly for the presence of an SL at the 5′ end of a specific mRNA species of *O. volvulus* by using PCR amplification. We examined the actin mRNAs, since the SL was first found on these mRNAs in *C. elegans* (16). For the PCR, the SL oligonucleotide and an oligonucleotide complementary to a highly conserved region of actin genes were used as primers on a template of first-strand cDNA molecules prepared from adult *O. volvulus* poly(A)$^+$ RNA (24). The several amplification products of this reaction (Fig. 6A, lane 1) included a 240-bp product that hybridized to an actin DNA probe (Fig. 6B). Other PCR amplifications conducted with different primers and templates (Fig. 6A, lanes 2 to 4) did not yield this product, although they did generate larger PCR products, which were not investigated further. These larger PCR products, which gave a similar banding pattern in lanes 2 to 4, are probably due to self-priming of the SL oligonucleotide on the genomic DNA (or genomic DNA contaminating the RNA), since the SL sequence is dispersed throughout the genome (Fig. 1) and may exist in the opposite orientation in some regions. Two other actin-specific primers yielded the expected amplification product of 550 bp when used on *O. volvulus* genomic DNA as the template (Fig. 6A, lane 5).

The 240-bp amplification product shown in Fig. 6A, lane 1, was cloned into a plasmid, and the sequence of one of the resultant inserts was determined, The sequence of this cloned PCR product contains the SL at one 5′ end, followed by 24 nucleotides of 5′ nontranslated region and the start codon of the highly conserved actin coding sequence (Fig. 6C). Thus, at least one actin mRNA species of *O. volvulus* contains a SL at its 5′ end.

To verify that this SL is not encoded by the corresponding actin gene, we used an actin DNA probe to screen an *O. volvulus* genomic DNA library and identified several clones. None of these genomic DNA clones hybridized to the SL oligonucleotide. Restriction mapping of the clones indicated that several actin genes are present in the *O. volvulus* genome, a conclusion supported by the results of Southern blots (data not shown). Oligonucleotides specific for the N-terminal conserved coding sequence of actin were used to obtain partial sequences of these clones. One clone possessed a sequence that exactly matched the sequence of the cDNA (Fig. 6C). This clone did not contain the SL sequence, but had a consensus acceptor splice sequence, AG, at the location of the SL in the cDNA. Furthermore, PCR amplification of genomic DNA with an actin oligonucleotide and the SL oligonucleotide did not produce a product detectable with the actin probe (Fig. 6A and B, lanes 3), indicating that the actin genes and the SL sequences are not closely linked in the genome. These combined results strongly suggest that *trans* splicing of the SL does occur in *O. volvulus*, as proposed for the other organisms in which an SL has been detected (6, 16, 22, 23, 27, 28).

By analogy with *C. elegans*, *B. malayi*, and *A. lumbricoides* (6, 23, 28), it seems likely that a transcript from the SL RNA genes within the 5S rRNA gene repeats provides the SL sequence that is present at the 5′ end of this actin mRNA. However, attempts to use Northern blots of total *O. volvulus* RNA to identify potential primary SL transcripts from this region of the genome were not successful (data not shown). We suspect that this negative result is because small

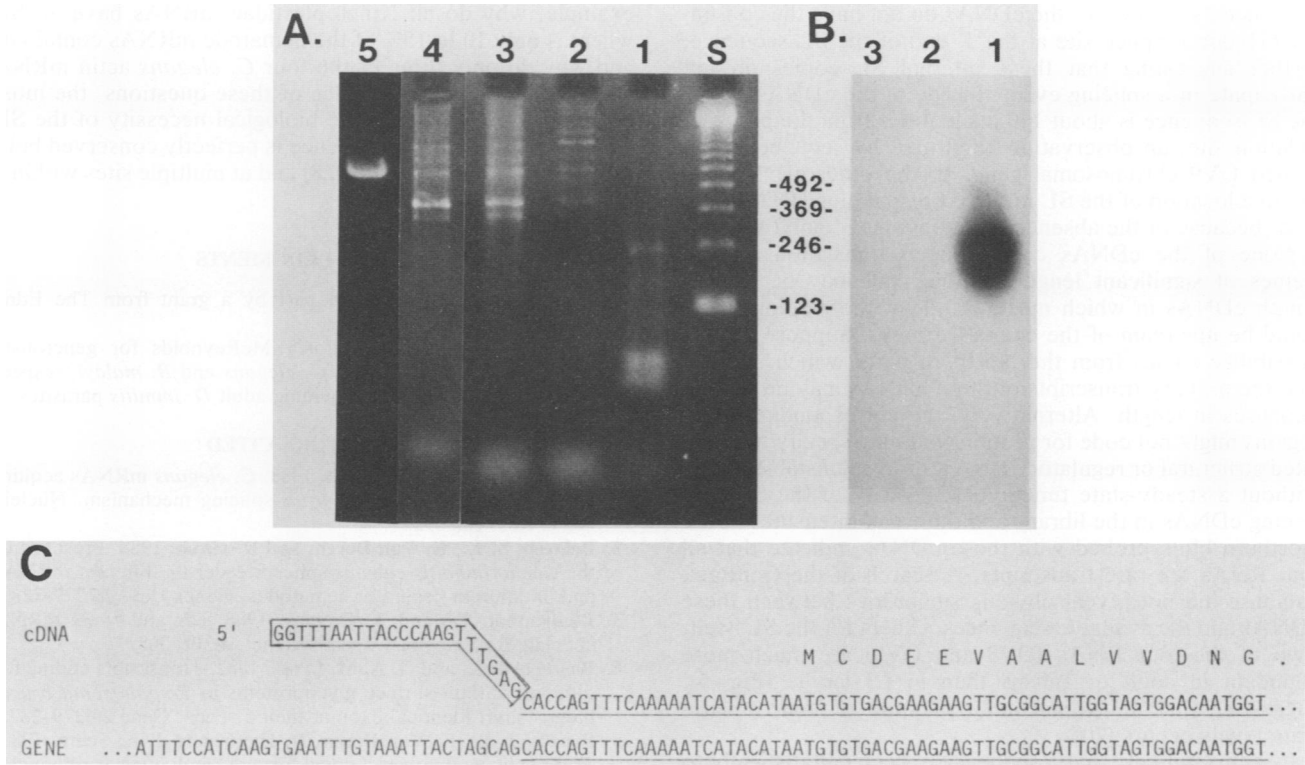

FIG. 6. (A) Agarose gel, stained with ethidium bromide, containing the products of different PCR amplifications. Templates for the reactions were first-strand cDNA (lane 1), poly(A)$^+$ RNA (lane 2), and genomic DNA (lanes 3 to 5) of *O. volvulus*. Primers were the SL oligonucleotide and a complementary actin oligonucleotide (lanes 1 to 3), the SL oligonucleotide alone (lane 4), and two different actin-specific oligonucleotides (lane 5). Lane S contains a 123-bp ladder of DNA fragments, some of whose sizes are indicated. (B) Autoradiogram of a Southern blot in which DNAs from lanes 1 to 3 in panel A were blotted to nitrocellulose and probed with an actin DNA probe. (C) Comparison of the 5′ sequences of an *O. volvulus* actin cDNA cloned from the PCR product shown in panels A and B, lanes 1, and its corresponding gene cloned from the *O. volvulus* genome. The boxed sequence is the SL, and the underlined sequences are the same in the cDNA and the gene. N-terminal actin amino acids are shown above their codons.

RNAs are lost in the procedure we used to simultaneously extract RNA and DNA from adult *O. volvulus* (10), since relatively little 5S rRNA was recovered. Unfortunately, *O. volvulus* cannot be grown in conventional laboratory animals or in culture, so adult organisms can be obtained only by surgery on volunteer onchocerciasis patients and L3 larvae can be obtained only by dissection from blackflies. This limits the amount of parasite material available and precludes our use of other RNA extraction procedures for this experiment.

DISCUSSION

Linkage between the 5S rRNA gene and the SL RNA gene within a DNA repeat has now been found in five different nematodes (16, 23, 28; see above). In the four parasitic nematodes (*O. volvulus*, *D. immitis*, *B. malayi*, and *A. lumbricoides*) the two genes are on the same strand of the DNA repeat, whereas in the free-living nematode (*C. elegans*) the genes are on complementary strands. Although we were unable to detect specific transcription of these repetitive SL RNA genes because of a lack of parasite material, it is likely that they are transcribed, because at least one *O. volvulus* actin mRNA species contains a 5′ SL that is not encoded by its gene (Fig. 6). In addition, the 3′-flanking regions of the repeated SL sequences in *O. volvulus* and *D. immitis* possess consensus sequences for both a donor splice site and a Sm antigen-binding site and can adopt a potential secondary structure similar to that proposed (5, 23) for the other nematode SL RNAs (data not shown). These 3′-flanking region are particularly highly conserved among the three parasitic nematodes, with more than 90 of the first 100 residues being identical. The parasitic nematode SL sequences and their 3′ downstream regions are both more highly conserved than are the corresponding repetitive SL RNA genes of the protozoa of the family Kinetoplastidae, suggesting that less sequence variation is tolerated in the nematode-splicing mechanism and/or SL function. In contrast, there is much more sequence variation in the nematode SL RNA 5′-flanking regions, where, presumably the SL RNA gene transcription is initiated (8, 28).

The presence of isolated orphan copies of the SL sequence in the *O. volvulus* and *D. immitis* genomes was initially detected by Southern blots (Fig. 1) and is consistent with a similar distribution of SL sequences in the DNA of *B. malayi* (28), but not in the *C. elegans* genome (16). In contrast to *B. malayi*, for which no evidence for transcription of these regions was detected (28), at least some of the isolated *O. volvulus* SL copies are transcribed to yield RNAs with perfect internal copies of the SL sequence, as represented by the four cDNAs described here. Three of these four cDNAs have no obvious properties in common except for the presence of this SL within a conserved 25-mer sequence. Likewise, these cDNAs are unrelated to the SL RNA gene sequence within the 5S rRNA gene repeats except for the SL

sequence itself. Two of the cDNAs do not have the consensus GT donor splice site at the 3′ end of the SL sequence, further suggesting that these internal SL copies do not participate in a splicing event. In one of the cDNAs, OV9, the SL sequence is about 140 nucleotides from the polyadenylation site, an observation confirmed by sequencing the cloned OV9 chromosomal gene. In the other cDNAs the relative location of the SL in the corresponding RNA is less clear because of the absence of an obvious 3′ poly(A)$^+$ tail.

None of the cDNAs contain open translation reading frames of significant length, although all may be partial-length cDNAs in which most, or all, of the reading frame could be upstream of the cloned segment. Support for this possibility comes from the Northern blots, which indicate that the primary transcripts of these cDNAs may be several kilobases in length. Alternatively, the corresponding DNA regions might not code for proteins but may specify untranslated structural or regulatory RNAs, or even unstable RNAs without a steady-state function. The rarity of the SL-containing cDNAs in the library and the weak intensities of the Northern blots probed with these cDNAs indicate that all four RNAs are rare transcripts. A search of the GenBank database did not reveal obvious similarities between these cDNAs and the available sequences, except for the SL itself. Two of the four RNAs (OV3 and OV9) are much more abundant in adult organisms than in L3 larvae (Fig. 5), suggesting that developmental regulation of their steady-state levels occurs (20).

It seems unlikely that a sequence expected statistically to occur only once in more than 10^{13} bp would, in fact, appear with perfect conservation at more than 30 different sites in the 1.5×10^8-bp *O. volvulus* DNA without being of use to the organism. One explanation for this scattered distribution is that the multiple SL genomic locations are the result of fairly recent duplications of an SL region within the genome. The similarity of the SL flanking regions in the OV3 and OV5 cDNAs supports this possibility, but other unique SL-containing genomic regions, represented by OV1 and OV9 cDNAs, are quite different except for the presence of the SL sequence. Alternatively, the presence of the T-rich regions upstream of the SL in three of the cDNAs suggests that the scattered genomic distribution of the SL could be the result of retroposonlike activity. This is an attractive possibility, but if it were the case, one might expect to find multiple copies of the entire cDNA sequence in the genome, such that the OV1 and OV9 cDNAs would hybridize to more genomic restriction fragments than they do in Fig. 4A. Another possibility is that the internal SL sequences might be evolutionary vestiges for *O. volvulus* and the other parasitic nematodes, whose function is no longer necessary because of the interaction between these nematode parasites and their insect or primate hosts.

Still another, perhaps more speculative, interpretation of the data is that the parasitic nematodes differ from *C. elegans* in the means by which the SL arrives at some of its RNA destinations but not in the function(s) it performs once there. This explanation would imply that the presence, rather than the location, of the SL is the most important factor for some RNAs and/or that the SL provides different functions for different RNAs. For instance, in one group of RNAs it may contribute to the processing, stability, or translation of the RNAs, whereas in other groups it might serve to influence the cellular location or membrane transport of the RNAs. Indeed, the metabolic need for the SL in the RNAs of both protozoa of the family Kinetoplastidae and nematodes continues to be its most mysterious feature. For example, why do all Kinetoplastidae mRNAs have an SL whereas only 10 to 15% of the nematode mRNAs contain it, and why do only three of the four *C. elegans* actin mRNA species require it? In the face of these questions, the most compelling argument for the biological necessity of the SL continues to be that its sequence is perfectly conserved both among different nematodes (28) and at multiple sites within a nematode species.

ACKNOWLEDGMENTS

This work was supported in part by a grant from The Edna McConnell Clark Foundation.

We thank X. Huang and Larry McReynolds for generously providing genomic DNAs from *C. elegans* and *B. malayi*, respectively, and John McCall for providing adult *D. immitis* parasites.

LITERATURE CITED

1. **Bektesh, S. L., and D. Hirsh.** 1988. *C. elegans* mRNAs acquire a spliced leader through a *trans*-splicing mechanism. Nucleic Acids Res. **16**:5692.
2. **Bektesh, S. L., K. Van Doren, and D. Hirsh.** 1988. Presence of the *Caenorhabditis elegans* spliced leader on different mRNAs and in different genera of nematodes. Genes Dev. **2**:1277–1283.
3. **Blumenthal, T., and J. Thomas.** 1988. *Cis* and *trans* mRNA splicing in *C. elegans*. Trends Genet. **4**:305–308.
4. **Boothroyd, J., and G. A. M. Cross.** 1982. Transcripts coding for different variant surface glycoproteins in *Trypanosoma brucei* have a short identical exon at their 5′-ends. Gene **20**:279–287.
5. **Bruzik, J. P., K. Van Doren, D. Hirsh, and J. A. Steitz.** 1988. *Trans* splicing involves a novel form of small nuclear ribonucleoprotein particles. Nature (London) **335**:559–562.
6. **Campbell, D., D. Thornton, and J. Boothroyd.** 1984. Apparent discontinuous transcription of *Trypanosoma brucei* surface antigen genes. Nature (London) **311**:350–355.
7. **Cook, G. A., and J. E. Donelson.** 1987. Mini-exon gene repeats of *Trypanosoma* (*Nannomonas*) *congolense* have internal repeats of 190 base pairs. Mol. Biochem. Parasitol. **25**:113–122.
8. **De Lange, T., A. Liu, L. Van der Ploeg, P. Borst, M. Tromp, and J. Van Boom.** 1983. Tandem repetition of the 5′ mini-exon of variant surface glycoprotein genes. A multiple promoter for VSG gene transcription? Cell **34**:891-900.
9. **Donelson, J., and A. C. Rice-Ficht.** 1985. Molecular biology of trypanosome antigenic variation. Microbiol. Rev. **49**:105–125.
10. **Donelson, J. E., B. O. L. Duke, D. Moser, W. Zeng, N. Erondu, R. Lucius, A. Renz, M. Karam, and G. Z. Flores.** 1988. Construction of *Onchocerca volvulus* cDNA libraries and partial characterization of the cDNA for a major antigen. Mol. Biochem. Parasitol. **31**:241–250.
11. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**:6–13.
12. **Gill, L. L., N. Hardman, L. Chappell, L. H. Qu, M. Nicoloso, and J. P. Bachellerie.** 1988. Phylogeny of *Onchocerca volvulus* and related species deduced from rRNA sequence comparisons. Mol. Biochem. Parasitol. **28**:69–76.
13. **Graham, R. W., K. Van Doren, S. Bektesh, and E. P. M. Candido.** 1988. Maturation of the major ubiquitin gene transcript in *Caenorhabditis elegans* involves the acquisition of a *trans*-spliced leader. J. Biol. Chem. **263**:10415–10419.
14. **Hattori, M., and Y. Sakaki.** 1986. Dideoxy sequencing method using denatured plasmid templates. Anal. Biochem. **152**:232–238.
15. **Huynh, T. V., R. A. Young, and R. W. Davis.** 1985. Construction and screening cDNA libraries in λgt10 and λgt11, p. 49–78. *In* D. M. Glover (ed.), DNA cloning: a practical approach, vol. I. IRL Press Ltd., Oxford.
16. **Krause, M., and D. Hirsh.** 1987. A *trans*-spliced leader sequence on actin mRNA in *C. elegans*. Cell **49**:753–761.
17. **Laird, P., J. Zomerdijk, D. de Korte, and P. Borst.** 1987. In vitro labelling of intermediates in the discontinuous synthesis of mRNAs in *Trypanosoma brucei*. EMBO J. **6**:1055–1062.

18. **Lenardo, M. J., D. M. Dorfman, and J. E. Donelson.** 1985. The spliced leader sequence of *Trypanosoma brucei* has a potential role as a CAP donor structure. Mol. Cell. Biol. **5**:2487–2490.
19. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
20. **Maniatis, T., S. Goodbourn, and J. A. Fischer.** 1987. Regulation of inducible and tissue-specific gene expression. Science **236:** 1237–1244.
21. **Maxam, A. M., and W. Gilbert.** 1977. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
22. **Murphy, W., K. Watkins, and N. Agabian.** 1986. Identification of a novel Y branch structure as an intermediate in trypanosome mRNA splicing: evidence for *trans* splicing. Cell **47**:517–525.
23. **Nilsen, T. W., J. Sharnbaugh, J. Denker, G. Chubb, C. Faser, L. Putnam, and K. Bennet.** 1989. Characterization and expression of a spliced leader RNA in the parasitic nematode *Ascaris lumbricoides* var. *summ*. Mol. Cell. Biol. **9**:3543–3547.
24. **Rappolee, D. A., A. Wang, D. Mark, and Z. Werb.** 1989. Novel method for studying mRNA phenotypes in single or small numbers of cells. J. Cell. Biochem. **39**:1–11.
25. **Sharp, P.** 1987. *Trans* splicing: variation on a familiar theme. Cell **50**:147–148.
26. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
27. **Sutton, R., and J. Boothroyd.** 1986. Evidence for trans splicing in trypanosomes. Cell **47**:527–535.
28. **Takacs, A. M., J. A. Denker, K. G. Perrine, P. A. Maroney, and T. W. Nilsen.** 1988. The 22-nucleotide spliced leader sequence in the human parasitic nematode *Brugia malayi* is identical to the *trans*-spliced leader exon in *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA **85**:7932–7936.
29. **Thomas, J. D., R. C. Conrad, and T. Blumenthal.** 1988. The *C. elegans trans*-spliced leader RNA is bound to Sm and has a trimethylguanosine cap. Cell **54**:533–539.
30. **Thomas, P. S.** 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA **77**:5201–5205.
31. **Unnasch, T. R., M. Y. Gallin, P. T. Soboslay, K. D. Erttman, and B. M. Greene.** 1988. Isolation and characterization of expression cDNA clones encoding antigens of *Onchocerca volvulus* infective larvae. J. Clin. Invest. **82**:262–269.
32. **Van Doren, K., and D. Hirsh.** 1988. *Trans*-spliced leader RNA exists as small nuclear ribonucleoprotein particles in *Caenorhabditis elegans*. Nature (London) **335**:556–558.
33. **Walder, J., P. Eder, D. Engman, S. Brentano, R. Walder, D. Knutzon, D. Dorfman, and J. E. Donelson.** 1986. The 35 nucleotide spliced leader sequence is common to all trypanosome messenger RNAs. Science **233**:569–571.