© 1992 Oxford University Press Nucleic Acids Research, Vol. 20, No. 7 1711–1715

In *Euglena*, spliced-leader RNA (SL-RNA) and 5S rRNA genes are tandemly repeated

M.Keller, L.H.Tessier, R.L.Chan, J.H.Weil and P.Imbault*
Institut de Biologie Moléculaire des Plantes du CNRS, Université Louis Pasteur, 12 rue du Général Zimmer, 67084 Strasbourg-cedex, France

Received November 22, 1991; Revised and Accepted March 13, 1992 EMBL accession nos X63153 and X63154

ABSTRACT

In *Euglena gracilis*, a 26 nucleotide leader sequence (spliced leader sequence = SL) is transferred by trans-splicing to the 5′ end of a vast majority of cytoplasmic mRNAs (8). The SL originates from the 5′ extremity of a family of closely related snRNAs (SL-RNAs) which are about 100 nucleotide long. In this paper we present the nucleotide sequences of two SL-RNA genes, confirming the sequences previously established by sequencing purified SL-RNAs. Although some SL-RNA genes are dispersed throughout the genome, we show that the majority of SL-RNA genes are located on 0.6 kb repeated units which also encode the cytoplasmic 5S rRNA. We estimate that the copy number of these repeated units is about 300 per haploid genome. The association of SL-RNA and 5S rRNA genes in tandemly repeated units is also found in nematodes but paradoxically does not exist in trypanosomes which are phylogenically much closer to *Euglena*. We also show that a high number of sequences analogous to the 26 nucleotide SL are dispersed throughout the genome and are not associated with SL-RNAs.

INTRODUCTION

Several years ago it was demonstrated that a trans-splicing mechanism occurs in the maturation of cytoplasmic mRNAs in two types of organisms the trypanosomes (for a review see 1, 2) and some nematodes (for a review see 3, 4). A leader sequence ('spliced leader sequence' = SL), which constitutes the 5′ end of a small RNA (SL-RNA), is post-transcriptionally added to the 5′ end of the pre-mRNAs (for a review see 5). In trypanosomes, all the mRNAs analyzed so far are multicistronic and use a trans-splicing mechanism which involves a SL unique to each trypanosome species. In nematodes, only 10–20% of the mRNAs are trans-spliced (6) but at least two different SLs can be transferred to the pre-mRNAs in *Caenorhabditis elegans* (7). Different SLs can be transferred to closely related mRNAs but nothing is known about the mechanism which selects the SL.

We have recently demonstrated (8) that the cytoplasmic pre-mRNAs are trans-spliced in a third type of organism the Euglenoids which are phylogenically close to the Trypanosomatids. By comparing the 5′ ends of several cDNAs and by *in vitro* arrested translation experiments, we have shown that the vast majority of the mRNAs in *Euglena gracilis* have the same (or nearly the same) 26 nucleotide long SL. By comparing the *rbcS* gene encoding the ribulose-1,5-bisphosphate carboxylase/oxygenase with the *rbcS* mRNA sequence (9), we have also shown that the SL is not encoded together with the remaining part of the mRNA. The 26 nucleotide SL was found to constitute the 5′ extremities of the 100 nucleotide SL-RNAs. These SL-RNAs are not unique as in trypanosomes, but form a multi-gene family whose closely related members differ by the sequences of their 74 nucleotide mini-introns (part of the SL-RNA downstream of the SL). As compared to the trypanosomes, the diversity of the SL-RNAs able to transfer identical SLs led us to study the genes encoding SL-RNAs in *Euglena gracilis*.

In *Trypanosoma brucei*, the genes encoding the SL-RNA are localized in a 1.4 kb unit tandemly repeated about 200 times (10, 11). In *C. elegans* the SL1-RNA gene is located in a 1 kb unit which is repeated 100 times (12) and also contains the 5S rRNA gene (13) in the opposite orientation. In an other nematode, *Brugia malayi*, the SL is found both dispersed in the genome and associated with the repeated unit coding for the 5S rRNA, but the SL-RNA and the 5S rRNA are transcribed in the same direction (14). In this paper we present the nucleotide sequence of two SL-RNA genes from *Euglena gracilis*. While some SL-RNA genes are dispersed in the genome, we show that the majority are associated with the cytoplasmic 5S rRNA gene in 0.6 kb tandemly repeated units.

MATERIALS AND METHODS

Conditions for growth of *Euglena gracilis* Z strain, RNA and DNA extraction and analysis, genomic library construction in the phage Charon 40 were performed as previously described (8).

The genomic library was screened using the SL*a* and SL*b* oligonucleotide probes (see figure 1) complementary to the region 43–62 of the R2 SL-RNA (8) ('*a*' sub-family) and the region 50–69 of the R5 SL-RNA (8) ('*b*' sub-family). Using each of these probes several phages were selected. Among the positive phages, one clone (*a1* clone) selected using SL*a* probe and two clones (*b1* and *b2* clones) selected using SL*b* probe were shown

* To whom correspondence should be addressed

by restriction analysis to have, as inserted DNA, a series of 0.6 kb repeated units. The 0.6 kb DNA fragments generated by Sca 1 digestion of each *a1*, *b1* and *b2* phages were cloned in the Eco R5 site of the Bluescript plasmid. The 0.6 kb insert of five randomly selected sub-clones obtained from each phage were purified and sequenced.

To determine the number of 0.6 kb Sca 1 fragments repeated in the genome, Sca 1 digests of *Euglena* genomic DNA (0.2 μg to 6 μg), and Kpn 1-Sac 1 digests of recombinant plasmid containing one copy (0.7 ng to 20 ng) were submitted to electrophoresis, transferred to nitrocellulose and hybridized to labeled 5S rRNA probe. The relative intensities were determined by i) scanning the autoradiograms ii) plotting the number of cpm's bound.

The sequences of the oligonucleotide probes used are
SL*a* 5′ GGCTTCCTTGTCCACATGGG 3′
SL*b* 5′ CCATGGAGACTTCCTTGTCC 3′
Complementary to 5S rRNA 5′ ACCAGGCTTAACTTCTG-AAATCGAAC 3′
Oligo L 5′ CGAAAAAAAATAGACACTCAGAAA 3′

RESULTS

Repeated units encode SL-RNAs and 5S rRNA genes

All the 100 nucleotide SL-RNAs presently known in *Euglena* have the same 26 nucleotide SL and form a family whose different members have closely related sequences (8). Two sub-families can be distinguished the '*a*' sub-family illustrated by the R1, R2, R3 and R4 SL-RNAs and the '*b*' sub-family illustrated by the R5 and R6 SL-RNAs (8). Figure 1 shows the alignment of the sequences of R2 and R5 SL-RNAs, members of each sub-family. To clone the genes coding for each of these two sub-families, a *Euglena* genomic library in the phage Charon 40 was screened using the SL*a* and SL*b* oligonucleotides specific for *a* and *b* sub-families respectively (see figure 1).

The *a1* clone (from the *a* sub-family) and the *b1* and *b2* clones (from the *b* sub-family), were found to contain 0.6 kb repeated units which were sub-cloned as described in Materials and Methods. The units originating from a given phage exhibit identical nucleotide sequences (>99% homology). The units from phages *b1* and *b2* have the same sequences but are different from the units originating from the phage *a1* (68% homology). Figure 2 shows the nucleotide sequences from the 0.6 kb Sca 1 fragments cloned from the phages *a1* and *b1-b2*. Comparison of the gene sequences with the SL-RNA sequences (8) shows that the *a1* sequence encodes a SL-RNA (position 251−348) corresponding closely to the R2 SL-RNA. Three differences are found between the *a1* gene sequence and the sequence of the R2 SL-RNA A in position 66, and CG in position 94 and 95 of R2 are deleted in the RNA (compare figure 1 and figure 2). We attribute these differences to polymorphism within the multi-gene family. Similarly, the *b1-b2* sequence encodes a SL-RNA corresponding to the R5 SL-RNA with two changes (T to C transition in position 49 in the RNA and the deletion of G in position 98). The first two 5′ nucleotides of the SL, which were not determined in the RNA sequence are AC in the two genes.

Outside of the SL-RNAs coding sequences, two regions indicated in dashed lines in figure 2 are identical in the two genes (positions 1−44 and 555−629). An oligonucleotide probe specific for the second 74 nucleotide long region (complementary to nucleotides 590−615) hybridizes only to an RNA of about 120 nucleotides in length (data not shown) transcribed from the same DNA strand as the SL-RNAs. This transcript was identified as *Euglena* cytoplasmic 5S rRNA by a search through the Genbank database. The first region conserved in *a1* and *b1-b2* corresponds to the 3′ portion of 5S rRNA and the second region to the 5′ portion of the 5S rRNA indicating that these two regions

```
                       .     .     .     .    50
a (R2 SL-RNA) NNUUUCUGAGUGUCUAUUUUUUUUCGGUAUAUACUCGUUCCACCCAUGUG
b (R5 SL-RNA) ..........................................UU..C.

                       .     .     .     .   101
a (R2 SL-RNA) GACAAGGAAGCCCAAAGUGACUUCCAAUUUUGGAGGCGGCAACGCCAUCC
b (R5 SL-RNA) ..........U.UCC.UG.-........................-CAG..-
```

Figure 1. Nucleotide sequences of SL-RNAs (8) from '*a*' sub-family (R2 SL-RNA) and '*b*' sub-family (R5 SL-RNA). The whole sequence of R2 SL-RNA is shown, while in R5 SL-RNA only the nucleotides differing from those of R2 SL-RNA are indicated. (−) indicates deletion introduced to enhance the nucleotide alignment. The spliced-leader sequence (SL) is in bold. The regions complementary to the SL*a* (in R2 SL-RNA) and to the SL*b* (in R5 SL-RNA) oligonucleotide probes are underlined.

```
              .         .         .         .         .         60
a1     ACTGAGGTGGGCGACCACTTGGGAACACTGGGTGCTGTACGCTTTTTTGCTGCGTTTTTC
b1-b2  ............................................GC.CC.TCGG.GT

              .         .         .         .         .        120
a1     TCCCCCCCCCCGGCCCAAAAAAAAACCCTCCGTGCGCAAGATTTTCAGAATGTGTCCGTC
b1-b2  AAAGGT.....CTTTG.----------TG.GTGAG.GCG..CG.......GA......C.

              .         .         .         .         .        180
a1     GTGGGTTATAAAGTCCGTTCCACCCTTTTTTTCGTGAACGTGGCCGTGGGCCATCACCGT
b1-b2  ...AA...C........CCA........CCCG.AC.GT..C.TGACACC...-------.

              .         .         .         .         .        240
a1     CGCACCGCCCGCCCACCTCGGTGAATGCCCGGTGATGCTCCTCGTCCGCCCCGGTCCTTC
b1-b2  G....GTT...........C.......-....G..A...............GCTG..CAT

              .         .         .         .         .        300
a1     AACTGCGGACACTTTCTGAGTGTCTATTTTTTTTCGGTATATACTCGTTCCACCCATGTG
b1-b2  GG.....T.............................................TT....

              .         .         .         .         .        360
a1     GACAAGGAAGCCCAAGTGACTTCCAATTTTTGGAGGCGGCAACCATCCTTCAAAT-GAAA
b1-b2  ..........T.TCCA..G..................................C..AC...

              .         .         .         .         .        420
a1     TTAAAAATAATTGAAGCCTTGAAGCCAAACCGCTCTCG-----GCCCTCCCCCCCCTGTG
b1-b2  AAC....A..AGCG..GGGC..GT.TCTG..T......CTCGAT..TG.TGGT.T.CT.C

              .         .         .         .         .        480
a1     CGCTCTCACCCTGGGGCACCCTCGAAGAACGGACTGAAACCCCCAGCTTCAAACACCGAA
b1-b2  ..A...GT.T..CT.T.T.T...CCTTG.AA.TGG.TCGTTTTTTTTC.TTGG..A.T.G

              .         .         .         .         .        540
a1     AATTTTCTTTTTTTTCCCCTTCTCCAGCTCAGCAGCAGCGCCCCCTTTATGCACGGAAGA
b1-b2  ..AAAG..GG.GAAA..GAC..C..TC.C.CC.CC.CTG.GG..TCA.GC...GC.G.C-

              .         .         .         .         .        600
a1     TGCATAGGCTGTTGCGTACGGCCATACTACCGGGAATACACCTGAACCCGTTCGATTTCA
b1-b2  -...........................................................

              .         .   629
a1     GAAGTTAAGCCTGGTCAGGCCCAGTTAGT
b1-b2  .............................
```

Figure 2. Nucleotide sequences of 0.6 kb Sca 1 fragments cloned from phages *a1* and *b1-b2*. The whole sequence of *a1* fragment is shown, while in *b1-b2* fragment only the nucleotides differing from those of *a1* are indicated. (−) indicates deletion introduced to enhance the nucleotide alignment. The SL-RNA genes are underligned by a continuous line in bold and the 5S rRNA gene by a discontinuous line in bold.

are contiguous in the genomic DNA. The Sca 1 repeat units are therefore oriented in the same direction in the genomic DNA. The 5S rRNA gene sequence corresponds to the RNA sequence previously determined (15) except that there is one G at the 5′ end (instead of two shown by these authors). The few modifications found later (16) in the *Euglena* 5S rRNA sequence are not confirmed by the gene sequence that we present here.

Organization of the repeated units encoding the SL-RNAs and 5S rRNA

Since the Sca 1 site is located inside the 5S rRNA sequence, we propose that the units are tandemly repeated as shown in figure 3A. To determine whether this tandem organization is typical of all the 5S rRNA and SL-RNA genes, genomic *Euglena* DNA was digested by restriction enzymes Sca 1, Dde 1 and Sau 3A.

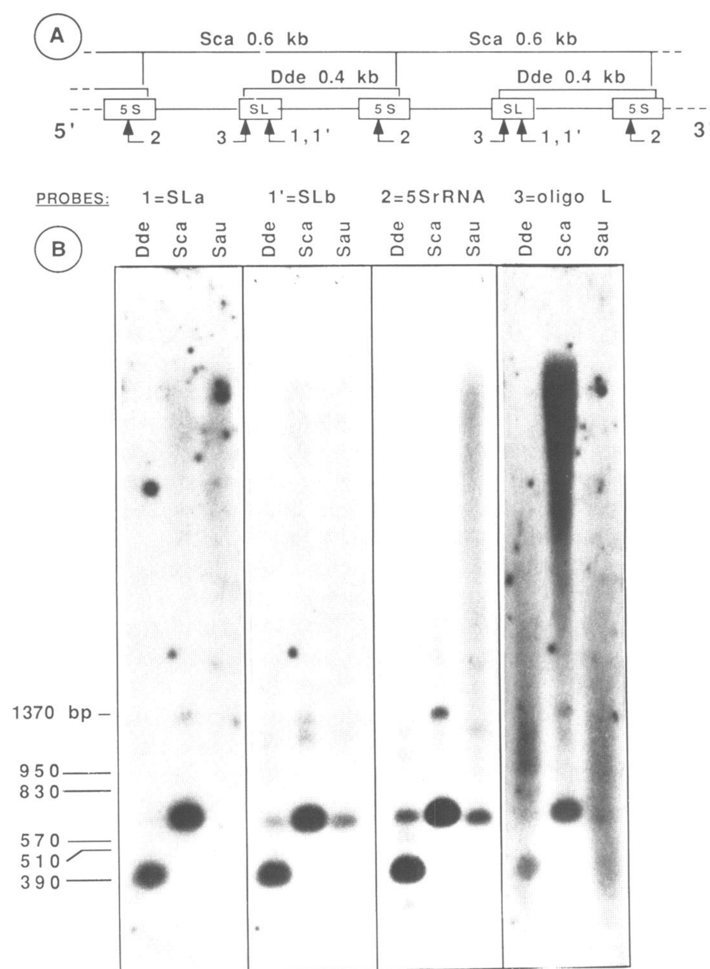

Figure 3. Organization of the SL-RNA and 5S rRNA genes in the *Euglena* genome. A—Map showing the tandemly repeated units in *a1* and *b1-b2* clones containing SL-RNA (SL) and 5S rRNA (5S) genes. The positions where SL*a* (1), SL*b* (1′), 5S rRNA (2) and oligo L (3) probes anneal are indicated by arrows. The positions of 0.6 kb Sca 1 and 0.4 kb Dde 1 fragments which hybridize to the probes are indicated. The Sau 3A enzyme cuts only *b1-b2* clone in two adjacent sites located downstream from the SL-RNA gene. B—14 μg of genomic DNA were digested using Dde 1, Sca 1 and Sau 3A restriction enzymes,electrophoresized on a 0.6% agarose gel and transferred on nylon membrane. The blot was successively hybridized by SL*a* (1), SL*b* (1′), 5S rRNA (2) and oligo L (3) probes which anneal the DNA at the positions indicated in figure 3A. Before hybridization, the probe was removed by washing the blot 30 min. at 45°C in 0.4 N NaOH and 5 min. in boiling water. The probe cleaning up was then tested by autoradiography.

Figure 3B shows that the SL*a*, SL*b* and 5S rRNA probes hybridize mainly to 0.6 kb Sca 1 and 0.4 kb Dde 1 fragments as predicted in figure 3A. These results show that SL-RNA and 5S rRNA genes are arranged on repeated fragments and confirm that the model in figure 3A concerns not only the three cloned genes (*a1*, *b1* and *b2*) but almost all 100 nucleotide long SL-RNA genes. Only a very small fraction of 5S rRNA and SL-RNA genes, revealed by very faint bands on figure 3B (probes 1′ and 2), do not follow this model. Furthermore, the presence of Sca 1 and Dde 1 restriction sites on the genes themselves (Sca 1 on the 5S rRNA gene and Dde 1 on the 5S rRNA and SL-RNA genes) rules out the hypothesis that there could be two types of repeated fragments, one encoding the 5S rRNA genes, the other encoding the SL-RNA genes. If this were the case, the Sca 1 and Dde 1 sites would have to be located at exactly the same positions in each of the two types of repeats (outside the genes) as in the cloned genes, in order to get the pattern shown in figure 3B, and this is very unlikely.

The number of 0.6 kb Sca 1 repeats present in the *Euglena* genome was estimated using a 5S rRNA probe and known amounts of plasmid as a standard. Assuming that the *Euglena* genome contains 1.36×10^9 base pairs (17) we calculated that about 300 Sca 1 repeats are present per haploid genome (see Materials and Methods).

The hybridization signals (0.6 kb fragments) obtained using 5S rRNA and SL*b* probes for Sau 3A digest are much weaker than the signals obtained after Sca 1 and Dde 1 digestions. This result shows that although some repeated units contain a Sau 3A site (like *b1-b2* gene), others (like *a1* gene) do not contain this restriction site and generate a smear when hybridized with SL*a* and 5S rRNA probes.

In figure 3B, we present hybridizations of the same digests with the oligo L, strictly complementary to the 26 nucleotide SL. In the case of Dde 1 digestion, the hybridization signal is weaker because this enzyme cuts the DNA within the hybridization locus and reduces the number of base-pairs formed. In the cases of Sca 1 and Dde 1 digestions, we can see clear signals (0.6 kb and 0.4 kb fragments respectively) originating from the 100 nucleotide SL-RNA genes tandemly organized with the 5S rRNA gene as described above. The oligo L probe mainly hybridizes to heterogeneous fragments as evidenced by the smears observed. This result suggests that some SLs are dispersed throughout the genome not associated either with 5S rRNA genes or with SL-RNA genes from the *a* and *b* sub-families.

Some sequences corresponding to 5S rRNA, SL-RNA and SLs are dispersed throughout the genome

When screening the library using SLa, SLb and 5S rRNA probes, we found a few clones (about 12 when a haploid genome was plated) which do not contain SL-RNA and 5S rRNA on repeated units. Some clones contain only one 5S rRNA gene, some clones only one SL-RNA gene and in some clones these two genes are organized on a single, non-repeated unit. Sequence analysis shows that these clones contain modified SL-RNA or 5S rRNA genes (the total number does not exceed 10% of the repeated genes) with deletions and nucleotide substitutions.

When screening the genomic library, using oligo L (complementary to the 26 nucleotide SL) as a probe, about 1/50 of the plated phages (about 2000–3000 phages when a haploid genome is plated) are positive, showing that sequences analogous to the SL are dispersed throughout the whole *Euglena* genome as suggested by the Southern experiments (figure 3B). In order

```
SLd1  GTTGCCTTTCTGAGTGTCTATTTTTTTTTC-GTCT---GCA--T-ATCATG 50
SLd2  C.GTG........................C....TCC...--.-.A....
SLd3  T.GT.........................-....-CCA..CG.GC.....

SLd1  TCGGACAAGTGAAAGCCCG--GT-GGGGTATATTTTGAAGCCCGAAGACA 100
SLd2  ..C------..........-G..C.....C-------....TTT.TT.AT
SLd3  G.C------..........AC..-.....C-----.....TT.AG.AGG.
```

Figure 4. Nucleotide sequences of the regions of SL*d1*, SL*d2* and SL*d3* phages carrying dispersed SLs. The whole sequence of SL*d1* is shown, while in SL*d2* and SL*d3* only the nucleotides differing from those of SL*d1* are indicated. The regions corresponding to the SL are in bold and the deletions introduced to enhance the nucleotide alignment are indicated (−).

to determine if dispersed SLs belong to other SL-RNA subfamilies, the DNA of three phages chosen at random (SL*d1*, SL*d2* and SL*d3*) was purified and analyzed. The sequences of the regions hybrizing to the oligo L (about 2000 nucleotide long for SL*d1* and SL*d3*, about 3000 for SL*d3*) were determined. Except for the 26 nucleotide SL itself, which is well conserved in the three clones, the surrounding regions have no significant homology. No repetitive structure was found in any of the three clones. Therefore, the organization of these dispersed SLs seems to be unrelated to the structure found in *Trypanosoma brucei gambiense* (18,19),*Crithidia fasciculata* (20) or *Trypanosoma cruzi* (21), where a retro-transposon interrupts a SL-RNA gene encoded by a repeated DNA structure. Figure 4 shows that some conserved sequence elements exist downstream of the SL, but these elements do not correspond to the 74 nucleotide SL-RNA mini-intron (which is part of the SL-RNA downstream of the SL). No transcribed RNA was found using probes surrounding the SLs. Therefore, the dispersed SLs we have characterized seem not to correspond to any functional RNA gene. Such dispersed SLs have also been characterized in the nematodes *Dirofilariae immitis* and *Onchocerca volvulus* (22), but in the latter some of the SLs were found inside expressed RNAs.

DISCUSSION

The results presented here show that a large proportion of the SL-RNAs which transfer by trans-splicing the SL to the *Euglena* mRNAs are encoded with the 5S rRNA in 0.6 kb tandemly repeated units. The SL-RNA gene sequences confirm the SL-RNA sequences obtained by polymerase-chain reactions on purified SL-RNAs (8) and have enabled us to determine the two 5′ nucleotides (AC) of the SL which were not sequenced on the RNAs. The sequence of the 5S rRNA gene confirms the accuracy of one of the sequences obtained by RNA analysis (15). Furthermore, our results are in good agreement with recent finding (23) showing that the 5S rRNA is not encoded by the 11.5 kb circular DNA containing the 19S, 25S and 5.8S rRNA genes as first predicted (24).

Although both SL-RNA and 5S rRNA are encoded by the same DNA strand, they are probably not co-transcribed, since in all organisms studied, the 5S rRNA is transcribed by RNA-polymerase III (like the U6 snRNA and the tRNAs), whereas SL-RNAs (like other capped snRNAs), are transcribed by RNA-polymerase II. In trypanosomes and in the case of a few nematode mRNAs, the transfer by trans-splicing of the SL to the pre-mRNAs provides a 5′ capped end to the mature mRNAs (25, 26). By analogy, we could suppose that *Euglena* SL-RNAs are also probably transcribed by RNA-polymerase II. In addition, the SL-RNA gene sequence has several stretches of T's (at the end of the 5S rRNA gene and in the SL itself). Such sequences are found to inhibit the transcription by RNA polymerase III (for a review see 27).

Nothing is known about *Euglena* RNA polymerases and the promoter elements they recognize. If *Euglena* SL-RNA and 5S rRNA genes are transcribed, as in the other organisms so far studied, the 0.6 kb Sca 1 fragment may contain the promoter elements recognized by both *Euglena* RNA polymerases II and III. A manual search in the 5′ flanking region of *a1* and *b1-b2* SL-RNA genes did not reveal any sequence close to the consensus RNA polymerase II-specific promoter sequences characterized in vertebrates (for a review see 28), *C. elegans* (29) or higher plants (30,31). Similarly the characteristic RNA polymerase III upstream promoter (for a review see 27) was not found upstream of the *Euglena* 5S rRNA gene. The analysis of other snRNA genes and site-directed mutagenesis experiments could help locating promoter elements recognized by *Euglena*. RNA-polymerases.

The presence of a SL-RNA gene in the repeated DNA unit coding for the 5S rRNA has already been shown in several nematodes. In the parasitic nematodes *B. malayi* (14), *A. lumbricoides* (32), *O. volvulus* and *D. immitis* (22) the two genes are on the same DNA strand but in the free-living *C. elegans* (12), the two genes are on the complementary strands. In *Euglena*, we have shown that the SL-RNA genes are associated with the 5S rRNA gene on the same strand of a 0.6 kb DNA fragment which is repeated about 300 times in the genome. In *Euglena*, at least two different 0.6 kb DNA repeated fragments, each coding for one type of SL-RNA, can be distinguished by the mini-intron and the intergenic sequences. In a given repeated units series, the sequences of the repeats are remarkably conserved. The 0.6 kb tandemly repeated DNA fragments contain the vast majority of the SL-RNA and 5S rRNA genes, but Southern experiments as well as screening of the genomic library have shown that sequences containing these two genes are also dispersed in the *Euglena* genome. We also found phages carrying non-repeated tandem of SL-RNA and 5S rRNA genes, but the analysis of the sequences of these genes shows that they are much less conserved (nucleotide substitutions, deletions) than the repeated ones. All the modified forms were selected using the probe specific for the *a* sub-family and none using the *b* sub-family specific probe. At least some of the modified forms are not functional and the *a* sub-family may be a degenerate form of the *b* sub-family.

The 26 nucleotide SL is found about 2000−3000 times scattered through the haploid genome. Sequencing and hybridization experiments suggest that these dispersed SLs seem to belong neither to SL-RNA genes nor to other transcribed genes. Their origin and their function(s), if any, remain unknown.

The phylogenic position of the Euglenoids (33) is close to the position of the Kinetoplastids (Bodonids and Trypanosomatids) if we consider their nucleus and nuclear division (closed mitosis, intracellular spindle, lack of metaphase plate formation), the 5S rRNA structure (15, 34), the fragmented structure of the rRNAs of the large ribosomal sub-unit (19, 35), the transfer by trans-splicing of a SL to a vast majority of mRNAs (8). The analysis of the protist phylogeny shows that Euglenoids and Trypanosomatids both diverged very early from the other eukaryotes (36, 37) and at about the same time (38). Our results demonstrate that, unlike the trypanosomes, *Euglena* has SL-RNA

and 5S rRNA genes associated in tandemly repeated units. This organization is found in some nematodes, which are multicellular eukaryotes that diverged much later during evolution. A possible explanation could be that in the ancestor common to *Euglena*, trypanosomes and nematodes, SL-RNA and 5S rRNA genes were associated and that this association was lost in trypanosomes. This suggests that the trans-splicing of SL sequences could exist in organisms less evolved than *Euglena* or trypanosomes.

ACKNOWLEDGEMENTS

We wish to thank Mrs Hoeft for the synthesis of oligonucleotides. R.L.C. was supported by a grant from CONICET-Argentina.

REFERENCES

1. Borst,P. (1986) *Ann. Rev. Biochem.* **55**, 701–732.
2. Laird,P.W. (1989) *Trends Genet.* **5**, 204–208.
3. Blumenthal,T. and Thomas,J. (1988) *Trends Genet.* **4**, 305–308.
4. Nilsen,T.W. (1989) *Exp. Parasitol.* **69**, 413–416.
5. Agabian,N. (1990) *Cell* **61**, 1157–1160.
6. Bektesh,S., Van Doren,K. and Hirsh,D. (1988) *Genes Dev.* **2**, 1277–1283.
7. Huang,X.Y. and Hirsh,D. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 8640–8644.
8. Tessier,L.H., Keller,M., Chan,R.L., Fournier,R., Weil,J.H. and Imbault,P. (1991) *EMBO J.* **10**, 2621–2625.
9. Chan,R.L., Keller,M., Canaday,J., Weil,J.H. and Imbault,P. (1989) *EMBO J.* **9**, 333–338.
10. De Lange,T., Liu,A.Y.C., Van der Ploeg,L.H.T., Borst,P., Tromp,M.C. and Van Boom,J.H. (1983) *Cell* **34**, 891–900.
11. Milhausen,M., Nelson,R.G., Sather,S., Selkirk,M. and Agabian,N. (1984) *Cell* **38**, 721–729.
12. Krause,M. and Hirsh,D. (1987) *Cell* **49**, 753–761.
13. Nelson,D.W. and Honda, B.M. (1985) *Gene* **38**, 245–251.
14. 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.
15. Delihas,N., Andersen,J., Andresini,W., Kaufman,L. and Lyman,H. (1981) *Nucl. Acids Res.* **9**, 6627–6633.
16. Kumazaki,T., Hori,H. and Osawa,S. (1982) *J. Mol. Evol.* **18**, 293–296.
17. Rawson,J.R.Y. (1975) *Biochim. Biophys. Acta* **402**, 171–178.
18. Aksoy,S., Lalor,T.M., Martin,J., Van der Ploeg,L.H.T. and Richards,F.F. (1987) *EMBO J.* **6**, 3819–3826.
19. Carrington,M., Roditi,I. and Williams,R.O. (1987) *Nucl. Acids Res.* **15**, 10179–10198.
20. Gabriel,A., Yen,T.J., Schwartz,D.C., Smith,C.L., Boeke,J.D., Sollner-Webb,B. and Cleveland,D.W. (1989) *Mol. Cell. Biol.* **10**, 615–624.
21. Villanuava,M.S., Williams,S.P., Beard,C.B., Richards,F.F. and Aksoy,S. (1991) *Mol. Cell. Biol.* **11**, 6139–6148.
22. Zeng,W., Alarcon,C.M. and Donelson,J.E. (1990) *Mol. Cell. Biol.* **10**, 2765–2773.
23. Schnare,M.N., Cook,J.R. and Gray,M.W. (1990) *J. Mol. Biol.* **215**, 85–91.
24. Curtis,S.E. and Rawson,J.R.Y. (1981), *Gene* **15**, 237–247.
25. Laird,P.W., Kooter,J.M., Loosbroek,N. and Borst,P. (1985) *Nucl. Acids Res.* **13**, 4253–4266.
26. Zwierzynski,T.A. and Buck,G.A. (1990) *Nucl. Acids Res.* **18**, 4197–4206.
27. Geiduschek,E.P. and Tocchini-Valentini,G.P. (1988) *Ann. Rev. Biochem.* **57**, 873–914.
28. Parry,H.D., Scherly,D. and Mattaj,I.W. (1989) *Trends Biochem. Sci.* **14**, 15–19.
29. Thomas,J., Lea,K., Zucker-Aprison,E. and Blumenthal,T. (1990) *Nucl. Acids Res.* **18**, 2633–2641.
30. Vankan,P. and Filipowicz,W. (1989) *EMBO J.* **8**, 3875–3882.
31. Waibel,F and Filipowicz,W. (1990) *Nucl. Acids Res.* **18**, 3459–3466.
32. Nilsen,T.W., Sharnbaugh,J., Denker,J., Chubb,G., Fraser,C., Putnam,L. and Bennet,K. (1989) *Mol. Cell. Biol.* **9**, 3543–3547.
33. Kivic,P.A. and Walne,P.L. (1984) *Origins of Life* **13**, 269–288.
34. Lenardo,M.J., Dorfman,D.M., Reddy,L.V. and Donelson,J.E. (1985) *Gene* **35**, 131–141.
35. Schnare,M.N. and Gray,M.W. (1990) *J. Mol. Biol.* **215**, 73–83.
36. Qu,L.H., Perasso,R., Baroin,A., Brugerolle,G., Bachellerie,J.P. and Adoutte,A. (1988) *Biosystems* **21**, 203–208.
37. Sogin,M.L., Gunderson,J.H., Elwood,H.J., Alonso,R.A. and Peattie, D.A. (1989) *Science* **243**, 75–77.
38. Sogin,M.L., Elwood,H.J. and Gunderson,J.H. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 1383–1387.