Eukaryotic Cell

Genome-Wide Analysis of C/D and H/ACA-Like Small Nucleolar RNAs in *Leishmania major* Indicates Conservation among Trypanosomatids in the Repertoire and in Their rRNA Targets

Xue-hai Liang , Avraham Hury, Ehud Hoze, Shai Uliel, Inna Myslyuk, Avihay Apatoff, Ron Unger and Shulamit Michaeli
***Eukaryotic Cell* 2007, 6(3):361. DOI: 10.1128/EC.00296-06.**
Published Ahead of Print 22 December 2006.

EUKARYOTIC CELL, Mar. 2007, p. 361–377 Vol. 6, No. 3
1535-9778/07/$08.00+0 doi:10.1128/EC.00296-06

Genome-Wide Analysis of C/D and H/ACA-Like Small Nucleolar RNAs in *Leishmania major* Indicates Conservation among Trypanosomatids in the Repertoire and in Their rRNA Targets▽†

Xue-hai Liang,‡§ Avraham Hury,‡ Ehud Hoze, Shai Uliel, Inna Myslyuk, Avihay Apatoff, Ron Unger, and Shulamit Michaeli*

The Mina & Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan 52900, Israel

Received 13 September 2006/Accepted 7 December 2006

Small nucleolar RNAs (snoRNAs) are a large group of noncoding RNAs that exist in eukaryotes and archaea and guide modifications such as 2′-*O*-ribose methylations and pseudouridylation on rRNAs and snRNAs. Recently, we described a genome-wide screening approach with *Trypanosoma brucei* that revealed over 90 guide RNAs. In this study, we extended this approach to analyze the repertoire of the closely related human pathogen *Leishmania major*. We describe 23 clusters that encode 62 C/Ds that can potentially guide 79 methylations and 37 H/ACA-like RNAs that can potentially guide 30 pseudouridylation reactions. Like *T. brucei*, *Leishmania* also contains many modifications and guide RNAs relative to its genome size. This study describes 10 H/ACAs and 14 C/Ds that were not found in *T. brucei*. Mapping of 2′-O-methylations in rRNA regions rich in modifications suggests the existence of trypanosomatid-specific modifications conserved in *T. brucei* and *Leishmania*. Structural features of C/D snoRNAs, such as copy number, conservation of boxes, K turns, and intragenic and extragenic base pairing, were examined to elucidate the great variation in snoRNA abundance. This study highlights the power of comparative genomics for determining conserved features of noncoding RNAs.

In eukaryotes (2, 9, 17, 42) as well as in archaea (10, 29), the two major rRNA modifications, 2′-O-methylation (Nm) and pseudouridylation, are guided by small RNAs, which in eukaryotes are small nucleolar RNAs (snoRNAs). The 2′-O-methylations are guided by C/D box snoRNAs, which are named after short motifs known as the C box (RUGAUGA [R designates purines]) and the D box (CUGA). These boxes, together with the short sequences near the 5′ and 3′ ends of the RNA, which have the potential to form a K-turn structure, are essential for processing, localization, and stabilization of these molecules (7, 19, 40, 44). The K-turn motif is the binding site for the 15.5-kDa/Snu13 snoRNP protein (26). Such motifs exist in C/D snoRNAs, which guide modification, and also in U3 snoRNP (26) and U4 snRNA (28). Most of the guide RNAs carry internal boxes related to the C and D boxes, known as C′ and D′ boxes. The recognition of the target by the guide RNA is based on complementarity of 10 to 21 nucleotides (nt) between these two molecules, located upstream of the D and D′ sequences. The methylation site is situated 5 nt upstream from the D and D′ boxes, within the domain of interaction between the snoRNA and the substrate (9, 18).

In most eukaryotes, the snoRNAs that guide pseudouridylation consist of two hairpin domains connected by a single-stranded hinge, the H (AnAnnA) domain, and by a tail, the ACA box. Two short rRNA recognition motifs of the snoRNA that base pair with rRNA sequences flanking the uridine to be converted to pseudouridine have been characterized. The pseudouridine is always located 14 to 16 nt upstream from the H box or the ACA box of the snoRNA (14, 36). The two hairpin loops share structural and functional similarities, and functional pseudouridylation pockets are found with equal frequencies at the 5′ and 3′ ends of the molecule; in many cases, the RNAs direct pseudouridylation of rRNA at two different sites (14).

Most of the C/D and H/ACA box RNAs characterized to date are from humans, *Saccharomyces cerevisiae*, plants, and archaea (2, 14, 33, 37, 39). Only recently has information on these RNAs in unicellular eukaryotic organisms, such as the amoeba *Dictyostelium discoideum* (1) and the diplomonad *Giardia lamblia*, become available (46).

Trypanosomatids are unicellular parasitic protozoa which are the causative agents of several infamous parasitic diseases, such as African trypanosomiasis, caused by *Trypanosoma brucei*; Chagas' disease, caused by *Trypanosoma cruzi*; and leishmaniasis, caused by *Leishmania* species. The genomes of these three trypanosomatids were recently published, and their DNA sequences highlight many unique features of these organisms (5, 12, 16). Trypanosomatids are unique because they engage in two unusual RNA processing events, namely, *trans*-splicing (21) and mitochondrial RNA editing (34). In addition, the large subunit rRNA undergoes trypanosome-specific cleavages during rRNA maturation, yielding two large rRNA molecules and four small RNAs, ranging in size from 76 to 226 nt (43).

* Corresponding author. Mailing address: The Mina & Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan 52900, Israel. Phone: 972-3-5318068. Fax: 972-3-5351824. E-mail: michaes@mail.biu.ac.il.

‡ X.-H.L. and A.H. contributed equally to this study.

† Supplemental material for this article may be found at http://ec.asm.org/.

§ Present address: Department of Biochemistry and Molecular Biology, University of Massachusetts, Amherst, MA.

▽ Published ahead of print on 22 December 2006.

Most relevant to this study are the *Leishmania* parasites. *Leishmania* spp. are obligatory intracellular parasites that cause a spectrum of human diseases, with an annual incidence of 2 million cases in 88 countries (16). The parasite cycles between two hosts, namely, the phagolysosomes of mammalian macrophages and the midguts of sand flies. In the insect host, *Leishmania* parasites grow as flagellated extracellular promastigotes; in the mammalian host, they proliferate as aflagellated intracellular amastigotes. The *Leishmania* species are divided into Old World *Leishmania* spp., including species such as *Leishmania infantum*, *Leishmania donovani*, and *Leishmania major*, and New World *Leishmania* spp., including species such as *Leishmania mexicana* and *Leishmania braziliensis*. The gene order and sequence are highly conserved among the 30 known *Leishmania* species, and diagnostic tools are in demand to distinguish between these related species (16).

While relatively little is known about snoRNAs in trypanosomatids, some unique features were found for snoRNAs in these species (11, 38). Most, if not all, trypanosome H/ACA RNAs are composed of a single hairpin RNA and carry an AGA box instead of an ACA box (22, 38). The first discovered trypanosome H/ACA-like RNA, the spliced leader-associated RNA 1 (SLA1), guides modification on a short-lived RNA (25), the spliced leader RNA (SL RNA). This RNA is the donor of the spliced leader sequence to all trypanosome mRNAs (21). Silencing of the pseudouridine synthase (CBF5) by RNA interference in *T. brucei* provided evidence for the role of SLA1 in *trans*-splicing (4). We proposed that SLA1 has a unique chaperone function and escorts the SL RNA early in its biogenesis until it is assembled with Sm proteins (4).

Most recently, using bioinformatics and experimental tools, we performed a genome-scale analysis of snoRNAs that guide methylations and pseudouridylations on rRNAs in *T. brucei* (24). Our data suggested that most snoRNAs are clustered in reiterated repeats that carry a mixed population of C/D and H/ACA-like RNAs. Predicting the modifications guided by these RNAs and using partial mapping data, we identified 84 2′-*O*-methyls and 32 pseudouridines on rRNA, suggesting a high occurrence of Nms compared to pseudouridines on rRNA (24). Many of these modifications are species specific and increase modifications at domains which are already modification-rich. About 40% of the trypanosome-specific modifications are situated in unique positions outside the highly conserved domains of the rRNA (24).

In this study, we expanded our analysis to *Leishmania* species. By searching homologues of the *T. brucei* snoRNA genes and then evaluating the clusters, we identified 23 clusters carrying 62 C/D snoRNAs and 37 H/ACA-like RNAs. Several snoRNAs that were not revealed in *T. brucei* were identified here. However, in general, the pattern of Nm modifications is highly conserved between *L. major* and *T. brucei*. The *L. major* clusters are highly repeated compared to those of *T. brucei*. Surprisingly, the expression of snoRNAs in high-copy-number clusters is not necessarily abundant. Factors that may affect the abundance of C/D snoRNAs, such as copy number, the potential to form the K-turn motif, intragenic and/or extragenic base pairing, and conservation of the boxes, were examined. Sequence conservation of H/ACAs among *Leishmania* species and other trypanosomatids allowed us to identify structural features which are conserved in these RNAs and to highlight the high degree of phylogenetic conservation among related species for structure-function analysis of these guide RNAs.

MATERIALS AND METHODS

Oligonucleotides. All oligonucleotides were specific to *L. major* and included the following (name, sequence, description, positions of complementarity): *LM14Cs1C1*, 5′-ATGCGGGGCATCTAAAGGCAT-3′, antisense sequence complementary to snoRNA LM14Cs1C1, positions 43 to 64; *LM14Cs1C2*, 5′-GTGGGACAGCTATGCTTGTA-3′, antisense sequence complementary to snoRNA LM14Cs1C2, positions 60 to 78; *LM20Cs1C1*, 5′-GGATTCGGTTACTCAAAGTG-3′, antisense sequence complementary to snoRNA LM20Cs1C1, positions 73 to 93; *LM20Cs1C2*, 5′-ATCGGCCCTGGTATTAGGTT-3′, antisense sequence complementary to snoRNA LM20Cs1C2, positions 76 to 96; *LM20Cs1C3*, 5′-AAGCAAAAGGTGTGTGGGGT-3′, antisense sequence complementary to snoRNA LM20Cs1C3, positions 56 to 76; *LM20Cs1C4*, 5′-CGGAAGGGCTCGGACTCATTA-3′, antisense sequence complementary to snoRNA LM20Cs1C4, positions 32 to 52; *LM20Cs1C5*, 5′-TGCACAACCGCGTCTCTT-3′, antisense sequence complementary to snoRNA LM20Cs1C5, positions 47 to 65; *LM23Cs1C1*, 5′-ACAATCGACAGTCGTTCCGT-3′, antisense sequence complementary to snoRNA LM23Cs1C1, positions 50 to 70; *LM23Cs1C2*, 5′-GTAGTAGTATTGTTCCTCG-3′, antisense sequence complementary to snoRNA LM23Cs1C2, positions 40 to 59; *LM23Cs1C3*, 5′-GATTGGAAACAGTAAGGTC-3′, antisense sequence complementary to snoRNA LM23Cs1C3, positions 61 to 80; *LM26Cs1C1*, 5′-ATGTAGGTAGCTTTGGGTAC-3′, antisense sequence complementary to snoRNA LM26C1C1, positions 60 to 80; *LM26Cs1C3*, 5′-CAAGCCACAGCGAGAATACG-3′, antisense sequence complementary to snoRNA LM26C1C3, positions 70 to 90; *LM35Cs2C1*, 5′-ATAGCGTCATGGTCGGAGTG-3′, antisense sequence complementary to snoRNA LM35Cs2C1, positions 42 to 61; *LM26C1H1*, 5′-TCACGACAAGGCCGGATG-3′, antisense sequence complementary to snoRNA LM26C1H1, positions 43 to 60; *LM26C1H4*, 5′-ACGTCTCCCAAGAGCAAGTC-3′, antisense sequence complementary to snoRNA LM26C1H4, positions 49 to 58; *LM26C1H8*, 5′-GCATCTCACGCACCTCAGGA-3′, antisense sequence complementary to snoRNA LM26C1H8, positions 47 to 66; *LM26C1H9*, 5′-CTCTCGCGCAGTTTCGGCAC-3′, antisense sequence complementary to snoRNA LM26C1H9, positions to 49 to 68; *LM33C1H1*, 5′-GTCTCTCGGTCTTGACCATT-3′, antisense sequence complementary to snoRNA LM33C1H1, positions 51 to 70; *LM33C2H1*, 5′-GTTGCGAGATAGGGTCGTAC-3′, antisense sequence complementary to snoRNA LM33C2H1, positions 40 to 61; *LSU3-AS1217*, 5′-GAACGCTTGGCCGCCACAAG-3′, antisense sequence complementary to rRNA large subunit beta, positions 1217 to 1236; *LSU5-AS1238*, 5′-ACCAGCTATCCTGAGGGAAA-3′, antisense sequence complementary to rRNA large subunit alpha, positions 1238 to 1257; and *4139*, 5′-TTGCCGGAAGACGGGTTCCGGA-3′, antisense sequence complementary to the spliced leader RNA, positions 52 to 73.

Prediction of targets in rRNA. The potential targets (for 2′-O-methylation) in rRNA were determined using the computer program BestFit (from the GCG package), which searches for the optimal local alignment between two sequences. For this study, the program was used to search for complementarity to rRNA that complies with the +5 guiding rule. Additionally, the targets were also predicted based on the data available from their yeast homologues. To predict the pseudouridines guided by H/ACA RNAs, the secondary structure of H/ACA RNA was predicted using the mfold program (http://www.bioinfo.rpi.edu/applications/mfold/old/rna/form1.cgi), and the sequences from the internal loop were used to search for complementarity with rRNA, based on the guiding rules established for yeast, mammals, and plants (http://www.bio.umass.edu/biochem/rna-sequence/Yeast_snoRNA_Database/snoRNA_DataBase.html; http://bioinf.scri.sari.ac.uk/cgi-bin/plant_snorna/conservation).

Mapping of modified nucleotides. 2′-O-Methylations on rRNA were mapped by primer extension with different concentrations of deoxynucleoside triphosphates (dNTPs), as described by Xu et al. (45), using primers specific to the relevant region of rRNA. Primer extension products were analyzed in 6% polyacrylamide–7 M urea gel, next to sequencing reactions performed using the same primer.

H/ACA snoRNA folding and conservation in trypanosomatids. The *L. major* H/ACA snoRNA secondary structure was predicted using the mfold program (http://www.bioinfo.rpi.edu/applications/mfold/old/rna/form1.cgi). Homologues for the selected H/ACA snoRNA were screened from the *T. brucei* genome (http://www.genedb.org/genedb/tryp/), the *L. braziliensis* genome (http://www.genedb.org/genedb/lbraziliensis/), and the *L. infantum* genome (http://www.genedb.org/genedb/linfantum/) and were subjected to multiple alignment (http://www.ebi.ac.uk/clustalw/).

RNA preparation and primer extension analysis. RNAs were prepared from *T. brucei* cells, using TRI Reagent (Sigma). Primer extension analysis was performed as described previously (22, 45), using 5′-end-labeled oligonucleotides specific to target RNAs, as indicated in the figure legends. The extension products were analyzed in 6% polyacrylamide–7 M urea gels and visualized by autoradiography.

Northern analysis. Total RNA (20 μg/lane) was fractionated in a 10% polyacrylamide gel containing 7 M urea. RNAs were transferred to a nylon membrane (Hybond; Amersham Biosciences) and probed with [^{32}P]ATP-labeled oligonucleotides.

RESULTS

Search for *L. major* snoRNAs and their genomic organization. To perform a whole-genome search for both C/D and H/ACA RNAs in *Leishmania*, we screened the *L. major* genome against the sequences of *T. brucei* C/D and H/ACA RNAs (http://www.genedb.org/genedb/tryp/). After identifying the genes, we defined clusters and their repeat numbers. From this search, we identified 23 clusters (Fig. 1) encoding a total of 62 C/D snoRNAs and 37 H/ACA-like RNAs. Homologues of 24 snoRNAs were not identified in *T. brucei* (14 C/Ds and 10 H/ACAs) in these clusters. About half of the clusters carry only a portion of the snoRNAs present in the *T. brucei* homologue clusters. Four of the *L. major* clusters (clusters 1, 6, 9, and 20) are composed of snoRNAs that, in *T. brucei*, are present in two separate clusters. One interesting case is LM5Cs1 (cluster 1), which seems to originate from a fusion between two separate clusters in *T. brucei*, namely, TB11Cs2, which encodes SLA1, and TB9Cs3. This fusion is a *Leishmania*-specific event, since in *T. brucei* and *T. cruzi* these clusters are present in two separate loci. In the case of clusters 7 and 11, both the content and the order of genes are entirely conserved, while other clusters are conserved in content but not in order. Interestingly, novel snoRNA gene clusters that were not yet identified in *T. brucei* were revealed in *Leishmania* (clusters 3 and 10). These clusters were identified because, besides the novel snoRNA species, they also carry snoRNAs that were previously identified in *T. brucei*. For instance, Lm36Cs2C3 is located in cluster 22 together with three other snoRNAs. This *Leishmania*-specific snoRNA was identified based on its vicinity to its *T. brucei* snoRNA homologue. In addition, since our recent publication on *T. brucei* snoRNAs (24), we reinspected the published clusters, especially large intergenic regions. This analysis led to the discovery of seven new *T. brucei* snoRNAs that were overlooked in our initial analysis. These novel *T. brucei* snoRNAs led to the identification of their *L. major* homologues.

The majority of the clusters in *L. major* are repeated 1 to 18 times (Fig. 1). In contrast, in *T. brucei*, we found clusters that were repeated a maximum of 7.5 times (24). We also found that repeated clusters could appear at a second location within the same chromosome. Examples include clusters 3, 4, and 5, in which only part of the cluster is repeated, as well as clusters 11, 14, and 17, where the entire cluster is repeated (Fig. 1). As in *T. brucei*, several of the clusters are single-copy genes, and most of the clusters seem to carry both C/D and H/ACA snoRNAs. Only six clusters harbor exclusively C/D snoRNAs.

To examine the validity of the copy numbers of the different clusters, we searched the database of *L. major* containing the raw sequencing data obtained from the shotgun clones (ftp://ftp.sanger.ac.uk/pub/databases/L.major_sequences/shotgun_READS/), using BLASTN. This analysis enabled us to determine the relative abundance of each snoRNA in the genome. The number of independent clones found in the shotgun library should reflect the number of times this gene is repeated in the genome. Indeed, a correlation of 0.77 was found between the information in the shotgun library and the data found in the published genome, suggesting that the information in the genome was valid for determination of the copy numbers of the snoRNAs within each cluster (not shown). Although in certain cases the information in the published genome may not reflect the exact repeat structure, a comparative analysis among different repeats can be used to estimate the relative copy number of each of the snoRNAs we identified. Inspecting the chromosomal locations of the snoRNAs suggests that their genes reside mainly on chromosomes 14 to 36, except for one cluster that was found on chromosome 5, carrying the SLA1 locus.

The clusters were always found flanked by protein coding genes. As observed in *T. brucei*, the distance between the cluster and the upstream open reading frame (>650 bp) is generally larger than the distance between the 3′ end of the cluster and the downstream open reading frame (~200 bp). This observation may suggest the existence of a regulatory region upstream of the cluster that may have promoter-like activity, as recently described for the monogenetic trypanosomatid *Leptomonas collosoma* (23).

Repertoire and properties of *L. major* C/D and H/ACA RNAs and their potential targets. The repertoire of the C/D snoRNAs is illustrated in Table 1. All of the C/D snoRNAs range in size from 72 to 148 nt. Their potential targets were predicted using bioinformatic tools (BestFit from GCG software) and are presented in Figure S1A in the supplemental material. Of the 62 C/D snoRNAs, 20 have the potential to guide two modification sites. The percentage of snoRNAs that can potentially guide two modifications (double guiders) is slightly lower than that in *T. brucei*, where there are 27 double guiders out of 57 snoRNAs. We were able to predict targets for all C/D snoRNAs except four. In addition, we identified 14 novel C/D snoRNAs whose homologues were not identified in *T. brucei*. Of the 62 snoRNAs, 31 have homologues in other eukaryotes (Table 1). As already published for the *T. brucei* snoRNAs, we found that the 5′ end of C/D snoRNA is located 1 to 5 nt upstream from the C box (24). We therefore provide the sequence of the 5 nt upstream of the C box. Based on the experimental data for *T. brucei*, the 3′ ends of C/D molecules were usually found 1 to 3 nt downstream from the D box (11); we therefore provide the sequence of the 3 nt downstream from this box (Table 1).

As in the case of *T. brucei*, all of the H/ACA-like RNAs in *Leishmania* are smaller than the C/D snoRNAs (63 to 91 nt) (Table 2), except for the *Leishmania* homologue of snR30 (LM5Cs1H1), which functions in rRNA processing (4). The potential targets guided by these RNAs are presented in Fig. S1B in the supplemental material. Identification of H/ACA-like RNA termini was based on experimental data provided by several of these molecules, including SLA1 (23, 25, 30). In all H/ACA snoRNAs known so far for trypanosomatid species, the terminal box sequence is AGA, not ACA, and is located 3 nt upstream from the 3′ end. In addition, all of the H/ACA RNAs described here, except for LM5Cs1H1, can form only a

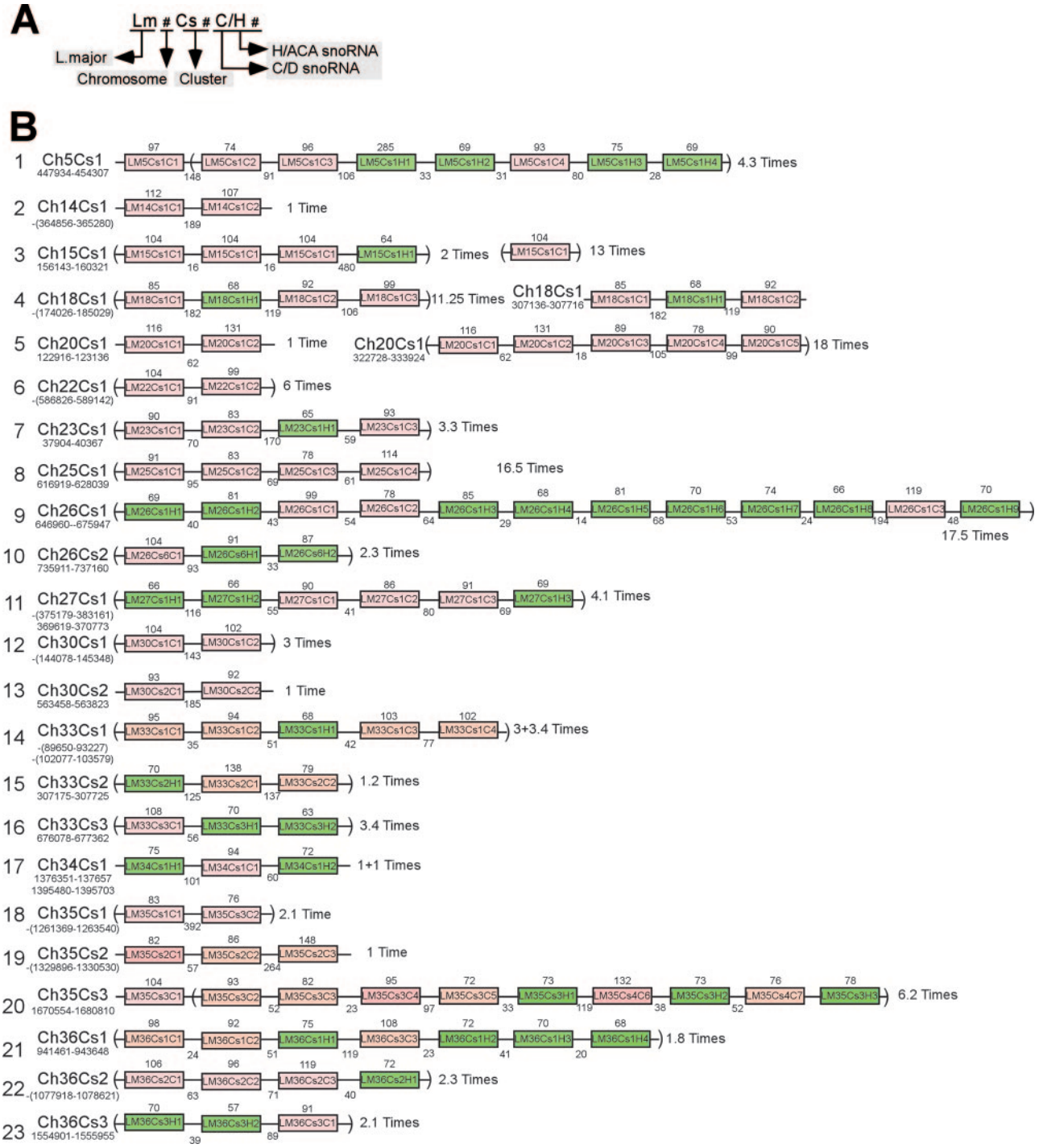


FIG. 1. Schematic representation of snoRNA clusters in *L. major*. (A) Nomenclature of *L. major* snoRNAs. (B) snoRNA clusters. The C/D snoRNAs are shown as pink boxes, whereas H/ACA-like snoRNAs are shown as green boxes. Thin lines indicate intergenic regions, and their lengths are indicated below the lines. The lengths of the snoRNA genes (±3 bp) are indicated above the boxes. The number on the right side of each cluster indicates the number of times the cluster is repeated in the genome database. The small numbers below the name of each cluster indicate the position of the cluster in the genome database in *L. major* GeneDB, release 5 (http://www.genedb.org/genedb/leish). A minus sign indicates the chromosomal location on an opposite strand. Parentheses indicate tandem repeats of the cluster.

single stem-loop structure. The 5′ ends of H/ACA-like RNAs are usually situated 1 to 3 nt upstream from the stem. Of 37 H/ACA-like RNAs, only 9 have homologues in other eukaryotes. We failed to predict the targets of seven H/ACA RNAs on rRNAs or snRNAs. In addition, we were unable to identify small Cajal-body RNAs in these clusters.

Sequence conservation among H/ACA RNAs in different trypanosomatid species. Phylogenic conservation is an excel-

TABLE 1. *L. major* C/D snoRNAs[a]

Name	Sequence	Length (nt)	Trypanosome homologue	Other homologues	Target
LM5Cs1C1	AUCGGUGACUUCUCCUCGACACGCUAAGC ACAUGGAGAACUUGACACGAGUAUCAG GAGCCCCCUGACUGAGAGACUAUCAUG CACCACUCUGAGGU	97	TB9Cs3C1	U32/snR40	G1623-S
LM5Cs1C2	ACAGAUGAUGAUUGACUGUAACAUCACA GACUUUGAGUCGCGAUGAUAGCAACCA UGUCGCCCCAGUCUGACGU	74	TB11Cs2C1	U31/snR67	U1833-S G1045-LB
LM5Cs1C3	CGCGUUGACAUGAUCGACACCUAGGCUGA UGUAAAGCCGUCGCAGAUGGACGUUG AGUGGCGCACGCCGUGAAACCAACCCC GCCUCGUCUGAUCG	96	TB11Cs2C2		
LM5Cs1C4	GCGGGGGAUGACCGACAACACGACAACGA AGCACAGUUUGAGCUGAUGCAUCGCUG AUGGACAACGUCACACAUGAGCUAACU CUGCUGACGC	93	TB9Cs3C2		A2021-S
LM14Cs1C1	CGUUGUGAUGUCAACCUCACUGGACCAU UAUUGCAAUAUACGAAUGCCUUUAGA UGCCCCGCAUCGAGGAAGAGACCCCAA UCGAGAAAUCCGCAGAGAAGGACUGAA CGCU	112			U1777-S
LM14Cs1C2	CUCCAUGACGACACACCACUAUUUGCACG UCAGUUUGACUUAUAAUUUACUCGGCA UGAUACAAGCAUAGCUGUCCCACCCCC ACCCACACGCACAUGCACUCACCAC	107	TB9Cs5C2	snR72	A927-LA
LM15Cs1C1	CGCCUCCAUGGACCUUGGAUCUGCGGUG GAUGGGCGGGCAACGAGUGCUGAGACG GAUGCUGACGACGAGCCGCUAUCGCCG UCGUUCGGGCUCGGCCCGAAGC	104			C113-LA
LM18Cs1C1	CGCUAUGAUGACAAACAUCUCCUUAACGA CGAGCGGACCGACACGCACCCAUGCGG ACAGUUCGAAUGUGUACAAAUGUUG AGCC	85	TB10Cs4C3	U43/snR70	U2048-S C2059-S
LM18Cs1C2	GCGCAGGAUGAGAGACAUUCAACGGAUCC CAAUCAGAGCACUGUGCACUGAAACUC CGUCCUGAUGCUAAUCCGCAAAGCAAC GACUGACGC	92			C442-LB
LM18Cs1C3	GAAGCUGAUGCCAAACACUCGUAGUGCGC GUCUGACGCUGCGUGACUGAUACAAGAC GAUGCGGAGCCGGGAAGACAUCCCGCG CCUCUUUGCUGACCA	99	TB10Cs4C4		G1865-S
LM20Cs1C1	GCGUGUGAUGAUACACCAAAGCAGGGAA UCUCUUGUCGAGCGGCCGGCGCGCGUC GCCGGCGUGGUGACAUGACACACUUU GAGUAACCGAAUCCAUUCAUUGAAUU GCCUGACGC	116	TB8Cs1C3		A669-LB
LM20Cs1C2	CCAAAUGACGAAUGUGGCAUCUCUCCACG GACACCGCAUUCGCACUGCGGCGCGUG ACAUCACUGUUGCCUUCGAACCUAAUA CCAGGGCCGAUGUGGGCGUGGGUGUGG GUGUGGGUGUGCAGCUGAAGA	131			A17-U6 snRNA
LM20Cs1C3	CUGCGUGAUGAGCUACAGGUCUACGACG AACAACUCGUCUGAUCACCCUUGUUGA UGACCCCACACACCUUUUGCUUCCUCA CUGACUG	89	TB8Cs1C1		G1626-LA G2151-S
LM20Cs1C4	GCAGGAUUACCACACCAACGAGUUUCUGA UGUAAUGAGUCCGAGCCCUUCCGCGGC GACGAGUCGAUAACGCUGUCAA	78			A104-S
LM20Cs1C5	CACCUGGAUGAACUUGUCCUCCGCAGAGU GCCGCUGCUGCUCUCUGAAGAGACGCG GUGUGCACGAGGCUGCAUGAAAACAUC CUGAUGA	90			C467-S
LM22Cs1C1	GGCUGUGACAAUGACCCACUUACGACGGU CUUAUGACCACGAUCCCGGCCACGGAUG AAGCACGAGUGCUACGUCUGUACAACA AGGGGGCGCGCCGCCGAUUG	104	TB9Cs2C1	U35/snR73	C1396-LB
LM22Cs1C2	GUACGUGAUGAGAAUUCAAGCUUAGGAC ACCUUUGGAUGCGGCGCGUGCCGCAGG UGAAAAUAGCGAGUUUGUUACACAUUC GAUUCGUACUCUGACGG	99	TB5Cs1C1	AtsnoR10	U1076-LB C1158-LB
LM23Cs1C1	GCACAUGAUUACGAUCAAGGACAUCUGAG UCAAAACCAAAGUCACUGUCUGAACGG AACGACUGUCGAUUGUGUCCCCCGUGU CUGAGCU	90	TB8Cs3C2		G1550-S

Continued on following page

TABLE 1—*Continued*

Name	Sequence	Length (nt)	Trypanosome homologue	Other homologues	Target
LM23Cs1C2	GUGAA**UGAUGA**UGCGUGUAUUCUGCAAU G**CGGA**CUCCAC**UGACGA**GGAACAAUACU ACUACUAUGACUAUUGGCAU**CUGA**CGC	83	TB8Cs3C3	AtsnoR39	G75-5.8 C18-S
LM23Cs1C3	GAUUC**UAAUGA**AACAUCUUUGCUACAG**GC GA**UGCGCGCUGCGGCGUGCG**AGAUGA**C CUUACUGUUUCCAAUCUCCAACCAAGA UCA**AUGA**GCU	93	TB8Cs3C1	U24	C1527-LA A1539-LA
LM25Cs1C1	GCGAC**UGAUGA**GAAACCUCAUGUUACUG ACACCUCUAGU**CUGA**CACUGUUGG**UGA CGA**GCACACAUUUGUAUAGUUAUCCCU GU**CUGA**ACG	91	TB11Cs1C2	snR48 snR38	G1228-LB G1252-LB
LM25Cs1C2	CCACG**UGAUGA**GCCAAUCAUGUUUCACUC UUGU**CUGA**ACCACCCGUUG**UGAUUA**CAC GCACACCCGCGAGGUGUGG**CUGA**CCC	83			G1578-S
LM25Cs1C3	CCACA**UGAUGU**GGAACACACGCAACACUC CGAUACCG**CUGA**CCGUGCUUG**UGACUGA** CACCACAUUGCGGA**CUGA**CAC	78	TB11Cs1C3		A955-LA
LM25Cs1C4	ACACG**UGACGA**GUAUGCAACACACAACUA GGUAAGAGCCGU**CUGA**ACGACAUACCU CCGUCCUGUGGAACACCAACGCCUGGG GUCUGCCACCGCAAUGGCUGUGCC**CUG A**GUC	114			
LM26Cs1C1	CUGCA**UGAUGC**GAAAAUCGAGGCAUUUG U**AUGA**GCCCGCAACUGUGAUU**UGAACU**C CGUGUACCCAAAGCUACCUACAUACUA UACACGGGA**CUGA**GCG	99	TB6Cs1C3	snR13	A526-LB
LM26Cs1C2	ACGCG**UGAUGG**CAUUGCGAUUUAAUUAU ACAC**CCGA**GGCGCGCACAGCGCCGU**GGA CGA**GAACCGGAGUGA**CUGA**CUG	78	TB6Cs1C2		
LM26Cs1C3	GCGCG**UGAUGA**UGCUGCAGAGUCUGAGU GUACCUUUC**UUGA**GCGUGCGCUGGUGC GCGCGUGGUGAGA**UGACGU**AUUCUCGC UGUGGCUUGCGCCGCGUGCGCUGGCC GAGU**CUGA**UGC	119	TB9Cs1C1	U34/snR62	G70-LB
LM26Cs2C1	GUGAA**UGAUGGA**CUGACCUGUGGAAUCU ACUGCGUUCUUU**CUGA**GCGUUGCGG**CG AUGC**GAGCUUCAUUGAGUUGCAAUCUU CAACGAAA**CCUA**UGUCUGACUG	104			A43-5.8S G36-5.8S
LM27Cs1C1	GCGCG**UGAUGA**AUACUAUUCACAUAGUU UCCUGUCAGU**CCGA**GCACUGCA**UGAGG U**CAUAGCAUUUUCUCAACUCAGCUCAC A**CUGA**CGC	90	TB11Cs4C1	U29/snR71	A1371-LB A1383-LB
LM27Cs1C2	CUGAA**UGAUGC**ACUCAUCUAUGUUGGUC CAAACAGU**UUGA**CUGGCGCA**UGAGAA**GG UACACAUACAAAUCACCUUUCGG**CUG A**CGU	86	TB11Cs4C2	U51/snR39	U847-LA A858-LA
LM27Cs1C3	GGCCC**UGAUGG**AGAAACAAUCCUUUUUG UGCGCGAGCGAGGCU**AUGA**GCGCACAU UU**UGAGCC**AACAAACAAAACUCAACAG GU**CUGA**GAC	91	TB11Cs4C3		G654-LB
LM30Cs1C1	CCACG**UGAUGA**GUCACCGUGCAUUCCUGU GGUAU**CUGA**UGCAUCCU**GCAUGA**ACAC GACAACUGCAACGAGCGCCCUUCGCGA CGAAGUGGUGCGAA**CUGA**GUG	104	TB6Cs1C1		C1247-LB
LM30Cs1C2	CGUGG**UGAUGU**CACACACGAUUCUAUUA CCUGUGGGAU**CUGA**UCAAGUGUUGUA U**UGAUGA**AAAGCAGAACGAUGCACGCG UGCAGCGUGUUAGG**CUGA**GGC	102			C1759-LB
LM30Cs2C1	UGACA**GGAUGA**UAUCUAUUCUUUGAAUG UUCUUCG**AUGA**CGGGCCUCGGCCUG**CG AUGA**GAAGCAUAAUAGAUGAACCCUUU AACU**CUGA**AAC	93	TB6Cs2C1	U36/snR47	A668-S
LM30Cs2C2	CGCAC**UGAUGA**AUUCGAUUCUUUGAAUG UUCC**UUGA**CCGCAGUCUUUGGCUG**UGA UGA**AAGGCUUUAUCGAUGAACCCUUUA ACU**CUGA**GCU	92	TB6Cs2C1A	U36/snR47	A668-S
LM33Cs1C1	CUGCA**UGAUUA**CAUUGCGUCUUCCACCUA **ACGA**CCGU**CGAUGA**UUACAUGCGAUGCA AGUGACUACUAUACUACA**CCGA**UAUCC GAGGAUGACGU	95	TB10Cs3C5	AtsnoR161	U709-LB G702-LB
LM33Cs1C2	CGCUG**UGAUGA**CUGGAUGAGUUAUUGUU UGCACAUUUA**CCGA**GCCUUCGUGGCCC G**UGAGCG**ACCGUAUAACUUCUAAGAAC ACCAG**CUGA**UUC	94	TB10Cs3C4	AtsnoR41Y	U1107-LA U1071-LA

Continued on facing page

TABLE 1—*Continued*

Name	Sequence	Length (nt)	Trypanosome homologue	Other homologues	Target
LM33Cs1C3	GCGUG**UGAUGA**UGUACAGAGUGUGCGAA GACGCAUGUCGCGUAGCCGCCGGCUGU GCGCGCC**UGACGA**ACUACUGUAUUUCA CGGGUGAACAAUCU**CUGA**UGC	103	TB10Cs3C1	snR52	U1358-LB
LM33Cs1C4	GGCGA**UGAUGC**AGACAGAUCAAUACUCA CU**CUGA**CGCGUGUGUGGCGGUGGACCG CUGCGCACUGUGCGGAUACAAGGAUAA CAUGCACUACCAG**CUGA**UGG	102	TB10Cs3C2	U33/snR55	A235-LA U1621-S
LM33Cs2C1	UUUGC**UGAUGA**UGCGUGAGGAGCGCUCCU CUCAUCCUCCGAAAGAGCAGAAUUAACC AA**CUGA**CGCUUUCACGGCGCCGUUGCCG CCGCGCCCCUCUGCCCCCCCCGGCCGCCCU CGCAGGCGUCGGUGCUG**CUGA**GCG	138	TB10Cs2C1		U8-S
LM33Cs2C2	AACCG**AGAUGA**AGCUGAUAUGGUCCGUG UUU**CAGA**UCCGUGUGACGCACAUUUCU CUCUUAGUACAUAUCUU**CUGA**GUU	79	TB10Cs2C2	U18	A681-LA
LM33Cs3C1	ACACA**UGAUGA**GACUCUUUUUUCGCUGC C**ACUGA**GCGCCCCAUGCGUGGGGCGGC ACG**UGAUGC**ACACAUGGAUACACUACA CUGUAGAACUGUCGUUUCU**CUGA**CGG	108			A512-S A603-LB
LM34Cs1C1	GUUUA**UGAUGA**UAAAGUCCUCUCUACUA CUACUAGCGU**CCGA**UGCUUCGC**UGACGA** UUACCGUUUGAACUUCAUCUUUCGCU CCGAU**CUGA**GGG	94	TB9Cs2C5	U80/snR60	G959-LA
LM35Cs1C1	CAUCU**GCAUGA**UUUCAUCUUAGCGUCAA GAAACAUCG**CUGA**UUGUUCUUU**GGAUU U**UAAACGCAAGCCGUAAGCAGU**AUG A**UCG	83	TB9Cs4C3		C358-LB
LM35Cs1C2	CUGAA**UGAUGA**UCCACGUUUAACACAAAU UCGACACUGAGU**CGGA**GCAAUAC**CCAU GA**CGCACCGUGUC**CUGA**CCG	76	TB9Cs4C2		A162-5.8S
LM35Cs2C1	CGAGC**UGAUGA**UGAUCAACAAACUUCGU AGGGUAAAAAU**CUGA**UCACUCCGA**CCAU GA**CGCUAUGGAAGUACUGCA**CUGA**AAG	82	TB9Cs3C3		U1418-LB
LM35Cs2C2	CCGUC**UGAUAG**CCGCUGUGUAGUGCCGU AUCGCUU**GUGA**GUGUCAU**CAAUGA**AUU GUGGAUAGUGCCGGAAGUGGUAGG**CU GA**GGC	86			C451-S
LM35Cs2C3	GCACG**UGAAGC**AGGUAAAAUAAGCAACCU GUUUAUCGGUGCCGCCCAUGUGGUCU AUCAUCAA**CUGA**UGGAA**UGAACA**GGUC UUCAAGCUCUGUAAGAGCCCGCGUCGA UGCCCCCCAUCAUUGCCAUUAGCUUGG ACUGG**CUGA**CUU	148			
LM35Cs3C1	CGCUG**UGAUGC**GAAUACCAACACGAACGA GAUUAACAGCAACAU**CUGA**UGCGUGCG CCAAAGCGCAAAC**UCAUGA**CGCGAACUC GUUUUUACAACAG**CUGA**UGC	104	TB9Cs2C6	AtsnoR13	U661-S
LM35Cs3C2	CGUCG**UGAUGA**CAUCAACCGACUGGAGUC U**CUGA**CAGGCCUCU**UGAUCGA**AAACAUC GCGUUGGAAUACUCUUCUUCUUUCCCC GU**CUGA**GGC	93	TB9Cs2C7	snR190 U52/snR78	G640-LB U666-LB
LM35Cs3C3	GCGAG**UGAUGA**AACUGCCGAAUCCGUGU UUCAG**CUGA**GCCGUGC**UGCUGA**CAGCG UUUCUUGAUCUGUUUGACUU**CUG A**CGC	82	TB9Cs2C2	snR58	C695-LA A678-LA
LM35Cs3C4	CCGCG**UGAUGC**ACAACAAUUCUCAUGGUC CAAACAGGUA**CUGA**CUGUCUGUUG**UGA UGA**UCACGGAGUAUGCGAAUCACCUUU CGGGA**CUGA**GGG	95	TB9Cs2C3	snR39B	G856-LA U845-LA
LM35Cs3C5	GCGCG**UGAUGC**ACAGCUGUUUGACCAUC AGAUGCCCCCAUGAGGACACCCAAACA CAGACCUGCA**CUGA**UGG	72	TB9Cs2C4	U25/snR56	G1829-S
LM35Cs3C6	ACCGC**UGAUGA**GGACCCACUUACGACGGU CUU**AUGA**CACAACACAUCCCGACCAU**GG AUUGA**AGUCAAGUGUUCAGCAAACAAC UAGGGGCGCGCCACAGGAUUUCCAGCC GGCGCGCCCUGUGA**CUGA**GGU	132	TB9Cs2C1	U35/snr73	C1396-LB
LM35Cs3C7	CGACG**UGAUGA**GGCUGACCCUGCUCCGCU A**CUGA**CAUGUGAAACCG**UGAUCC**GAGA CUCUUCAGGGCAA**CUGA**CGG	76	TB9Cs4C1	U38/snR61	G1190-LA
LM36Cs1C1	CACGC**UGAUGA**UCUGUCUAUUCGCAUCGC GACACG**CUGA**CACACACAACCC**GAGUGA** CAGGCGAACGGAACCAACAAUCGAGAA AGAACAG**CUGA**GUG	98	TB10Cs1C3		U7-5.8S U1599-S

Continued on following page

TABLE 1—*Continued*

Name	Sequence	Length (nt)	Trypanosome homologue	Other homologues	Target
LM36Cs1C2	CGGUG**UGAUGA**GUCUGCUAAGAUGUACG CAACAUUUACAAUUGGAAGACGAGU**CU GA**ACAC**UGAUGC**ACACACUGAUAAACGG CA**CUGA**CGC	92	TB10Cs1C4	snR75	G533-LB
LM36Cs1C3	ACCGC**UGAUGA**CAAUCCUUCACCACAAAU GUAACUCGGGAAAA**CUGA**ACCACCACG UGUGGG**UGAUGA**GCCACCGUACGGAAC AAUCGUUAGAUAGUAAGU**CUGA**GCU	108	TB10Cs1C1	U74/snR64	A590-LB C582-LB
LM36Cs2C1	GCGCC**UGAUGA**UCAAUGAGAACAAGCAUA UAU**CCGA**CUGCGCUGCGCGUCAGCUCG CUGUCGUGCGCCG**UCAUGA**GAUGCAAC CUCUAGUCCUUGAAAU**CUGA**CUG	106	TB8Cs2C1	U27/snR74 AtsnoR66	A28-S C38-S
LM36Cs2C2	ACGCG**UGAUGC**AUAAUUAAUCGACCAAU AAU**CGGA**GCGUGCCGCCUGCGCGGUGG CGUGCGUGAAGUCAAACCAAUGCAAUA UAUGUGG**CUGA**CUG	96	TB8Cs2C0		U1979-S
LM36Cs2C3	AAGCG**UGAUGC**AGAACAUACAACAAUCAC CUCGGGGGGCUGCCUCGCGGUGGCUCC CU**CUGA**AGCGCCGCUGCGUGCGGCCAC U**UGACUGA**ACCCCUGUUAAGCGUCAUC CG**CUGA**GUU	119			U1371-LA
LM36Cs3C1	UUCGA**UGAUGA**UGCACGAGUUAAUCCAU UCU**CUGA**GUCUCUGUGUGCGUGUGUA UGUGCGCAUGUGCACUGCCGAACUGCG GGC**GUGA**CGG	91	TB11Cs3C1		U559-LB

[a] The sequences of the C/D snoRNAs were obtained from the *L. major* genome database (http://www.genedb.org/genedb/lies/BLAST.jsp). Potential C/D and C′/D′ boxes are shown in bold. The coding region was predicted based on the positions of the C/D boxes. In cases where the termini of the molecules were not experimentally determined, we provide sequences spanning a region 5 nt upstream of the C box and 3 nt downstream of the D box. The predicted size of each snoRNA is indicated. The sequences predicted to interact by base pairing with their targets are underlined. The *T. brucei* homologue and homologues from other organisms are provided with their designated names (yeast homologues are designated sn*x*, human homologues are designated U*x*, and *Arabidopsis* homologues are designated Atsno*x*). The potential targets on rRNAs are indicated as follows: S, rRNA small subunit; LA/LB, rRNA large subunit (5′ and 3′ halves, respectively); and 5.8S, 5.8S rRNA. All sequences can be found in the *L. major* genome database (http://www.genedb.org/genedb/lies).

lent approach to evaluate the importance of particular domains for the function of RNA molecules. To this end, we inspected the sequences and secondary structures of most *L. major* H/ACA snoRNAs and indicated changes in homologues of *T. brucei*, *T. cruzi*, and *L. infantum* RNAs. The results for five such molecules are presented in Fig. 2. More molecules were analyzed, and the results indicate that H/ACA-like RNAs share several conserved features. All H/ACA-like RNAs, except for LM5Cs1H1 (snR30 homologue), have an AGA instead of an ACA motif. In most cases, an A exists 1 nt upstream of stem I (94% in *T. brucei* and 67% in *L. major*). In *L. major*, C can also appear in this position (about 30% of the time), while G and U are rarely found. Stem I is usually perfect and can range from 4 to 8 nt in length. In most cases, stem I is 6 to 7 bp long, and compensatory changes are often found to support the integrity of the stem (Fig. 2). The pseudouridylation pocket varies in size from 12 to 17 nt. Stem II also varies in size, but a prefect stem of 4 to 7 nt exists immediately adjacent to the pseudouridylation pocket. The conserved length of the stems is supported by the presence of compensatory changes. A consensus structure for trypanosomatid H/ACA-like RNAs is shown based on these findings (Fig. 2).

Conservation of snoRNAs among different *Leishmania* species. The existence of full genome sequences for a variety of *Leishmania* species enabled us to compare the sequences of snoRNA genes among *Leishmania* species from the Old World (*L. major* and *L. infantum*) with those from *Leishmania* species from the New World (*L. braziliensis*). The analysis was performed on cluster Lm34Cs1, and the results (Fig. 3) indicate 87% identity in the coding sequence between the New World species *L. braziliensis* and Old World species *L. major* and *L. infantum*. There is almost complete identity (97%) between the two Old world *Leishmania* species. This conservation also lies outside the coding sequences. On the other hand, only 35% identity was found in the intergenic regions between Old and New World *Leishmania* species, suggesting that comparative analysis between Old and New World *Leishmania* species can be used to identify only the more conserved coding sequences of the RNAs. Based on these data, the intergenic regions can potentially be used diagnostically to differentiate related *Leishmania* species from *L. major*. Additional clusters were analyzed in the same way, and similar results to those presented in Fig. 3 were obtained. However, in several cases, high conservation in the intergenic regions made it difficult to determine the exact boundaries between the coding and intergenic regions.

Most modifications mapped on rRNA are shared between *L. major* and *T. brucei*. The finding of snoRNA genes that exist in *L. major* but not in *T. brucei* raised the possibility of the existence of *L. major*-specific modifications. We therefore mapped the Nms on *L. major* and *T. brucei* rRNAs, using primer extension with different dNTP concentrations. We chose the large subunit (LSU) regions that are rich in modifications. At low dNTP concentrations, the reverse transcriptase pauses 1 nt before the Nm. To map the exact positions where the primer extension stops, the extension products were separated next to DNA sequencing reaction mixtures, using the same primer on the *T. brucei* rRNA gene fragment. The results (Fig. 4A and B) indicate that most of the modifications are present in both *T. brucei* and *Leishmania*. The species-specific

TABLE 2. *L. major* H/ACA-like RNAs[a]

Name	Sequence	Length (nt)	Trypanosome homologue	Other homologue	Target
LM5Cs1H1	CACTGGCCCTGTCACCTTGAGGGCGCCCGTGCCGCTT GTCGGCGCATGCTGGGCGTACGGCCTCGAGTTGCT GCGGCGCGGGCGCACGTGCGTGCATGTGCCGTCTG CGCTGCGATTCGGGGGGCCACAGAGCAAGGGGTG CGAGCGCGCGGTGTGATGCTCGGCGTGCTGCTCGC GCGCCTGACGCACCGCCGCGCGCCGCGGTTCCCGT TACGCATATCCTGTGTGCGCTCCAGCCCCGTCGGG GCCATGACGAGTTCGCAACCAACCATGCGTTACAC ATTG	285	TB11Cs2C3	snR30	
LM5Cs1H2	ACGCGCUCACUGUAGUGCGCGUGCCGCAAGACCCA CAGGCACCGUGCAUGAAACUGGAGCGCU**AGA**UGG	69	TB11Cs2H1	SLA	SL-RNA
LM5Cs1H3	GC**A**GUACGAUUAAACACGGACGCGCCGCUGCAUAC GCUUGGUGCUGGCGCCGUCCGCCAGUGUCGUAC G**AGA**UGU	75	TB9Cs3H1	snR34	1263-LB
LM5Cs1H4	GC**A**GUGCCGUCAGGUAAAGUGACGAUGGUGUGUC CCAUCCGCGCCACUGGUUCAAGCGGCACG**AG** **A**GCG	69	TB9Cs3H2	Atsnor77	870-LA
LM15Cs1H1	AG**A**GGAGUCAGGCGCGUAUGCGGCACAGAGAGACG AUGUGUACGCCGUUGGGCUCUGC**AGA**GUG	64			1754-S
LM18Cs1H1	CG**A**GGAGAACCGUAGCCGCUGCAACUCCGCUAGGU GAGAUGCCGCGGGUGAUACUUCUCCA**AGA**GCUC	68	TB10Cs4H4		455-S
LM23Cs1H1	CG**A**UGGGGUUCGAAUGCGCUGGUCGAAACCGGUU UGCCCUAGCGUGGGAAACGCCCCAA**AGA**GCC	65	TB8Cs3H1		1374-S
LM26Cs1H1	UC**A**GCGCACGCACCAGCCGCGCGCAGUUGAUGUUG GUUGCGCAUCCGGCCUUGUCGUGACGCG**AGA**GAC	69			607-S
LM26Cs1H2	CC**A**GCGUGCUCCGUGGGUGAGGCAGCGGACUCUAC UACCCUUGGUGCCGUACAUUGCGCCUCAAUCGG GCUCGCG**AGA**GCU	81	TB6Cs1H2		1395-LA
LM26Cs1H3	C**A**CCUUCCUCCCAGCCGCACGUGGUGUCGGCUGCG GCUGCCCACUGCUGUGUGCGUGCGCGUGGCUUG AAGGGUGGCCG**AGA**GAU	85			558-S
LM26Cs1H4	A**A**CCAGGAGUAGCACCGUUGCGGAACAGGACCCAU UGAUCUAUCGGCGGACUUGCUCUUGGG**AGA**CGU	68	TB6Cs1H3		609-S
LM26Cs1H5	CC**C**GCCGCGCAUUCAAGGUCCCUUGGCGGCGAUUA UAUACGCUCGAUGCUGCUUCACAGGAUACCCAGC GUGGCC**AGA**GCU	81	TB6Cs1H4		562-LA
LM26Cs1H6	**C**GCGCAGUUCAAGAAGCCGGUGCAUGCGUGUGACU GUGUGCUGCACCGGUGCCGAGACUGCGCAA**G** **A**GCC	70	TB9Cs1H1	SnR34	941-S
LM26Cs1H7	GGC**A**GGUCAGCGAAAACACAGUCGAAGUGGUAUUG ACCCCGUCAUUUCGCCUGCACCUGUGCUGUCCG**A** **GA**GAG	74	TB6Cs1H1		1412-LB
LM26Cs1H8	CC**A**ACGCACUAUUGCUUCCGCUGCUGCGUGUCGAU GGGCAGCGGGAUCCUGAGGUGCGUG**AGA**UGC	66	TB9Cs1H3		1970-S
LM26Cs1H9	UG**A**GCGCAGCCUAGGAAACCAGUGCAUGCAAGCCC UUGUGCUGCUCUGGUGCCGAAACUGCGCG**AG** **A**GCU	70	TB9Cs1H1A		1317-LB
LM26Cs2H1	GUG**C**AUGCGUACGCGCCUGCUCAACUCCCGUCUGU UCACGCUGCAGGCUGUGGGAGUCGGUGGUGCGG CCGGCAGGCCAGCAUGA**AGA**AGC	91			
LM26Cs2H2	AA**A**GCAGCUCGUGUUACCGCUCGAAUGGAUGCGCU GUGCAUCUGCUCCCGUAUUCGAGGUUUUGUGUG UCAAAGGGCUGCA**AGA**CGA	87		SnR35	1543-S
LM27Cs1H1	AC**C**UGUCUGGCACCUGUCUUUACACCUGUCAGUGA CAGAACACGCCCGAAAGUUCAAGAG**AGA**CAA	66	TB11Cs4H1		1402-LB
LM27Cs1H2	GC**C**GCACAACUCCCUUGGAAUGGUGGGUGCCUCAU CCCACGCAAGUCCUUGUCCUGUUGUGCC**AGA**GCG	69	TB11Cs4H3		610-LB
LM27Cs1H3	GAA**C**GGAGUGCAUUGAGUCCCACCGCUCCCCACCA GUACGCGCGUGGGAGCAUCAACUCUCCU**AGA**UGC	69	TB11Cs4H2		1841-S
LM33Cs1H1	**U**GCACCUUCCGGGCAGGAUGACGCCCUUGUGCCCU UGGCGACUGUGUCCAAUCGGUGGUGCG**AGA**UGC	68	TB10Cs3H2	SnR32	436-LB
LM33Cs2H1	A**C**GGUGUUGUCGGUUGGGCGUGACUCCUUGCACC UCGUAGGUACGACCCUAUCUCGCAACCCCC**AG** **A**UUC	70	TB10Cs2H1	Atsnor93	1212-LB
LM33Cs3H1	CG**A**GGUCUUCGUGUACCCCGUGCGCCCCCGCUCGU GGCGCACACAUGGGGAAUGGUCAAGACCG**AG** **A**GAC	70	TB9Cs4H1		2048-S
LM33Cs3H2	CCAUUUCGCUCGUCCCGCUCCGCACCGUUGAGCUG CCGUCGAUGAACAAUGGGCCGAGACGU	63			
LM34Cs1H1	GCG**A**AACGUCAGCUACGCAGUGUGAGCUGAUAGUG UCUGUUCGGCUACGUGCACUACCAAAGCGGCGU G**AGA**GCG	75			1533-LA
LM34Cs1H2	GUG**A**CGUCACUCCAACCGCUGCUCAUACGUCGUUG CCCAUCGUAUAGCGGUGAUAUCUGUGGACGG**AG** **A**UGG	72			1529-LA

Continued on following page

TABLE 2—*Continued*

Name	Sequence	Length (nt)	Trypanosome homologue	Other homologue	Target
LM35Cs3H1	GUG**C**GGCCCACAUUGGAGUUGUGUCCUGGCCCCAG CAGUGGCCCUGUCUGCACACUCGCGUGGGCCA**AG A**GAU	73	TB9Cs2H1		661-LB
LM35Cs3H2	**C**GUCCCACCAGCUACCACGCGGAGUGCAUACCCUU UCAGUCUUCGCGGUCGGUCGUUUUCUGGGCCA**A GA**UGC	73	TB9Cs2H2		1381-LB
LM35Cs3H3	UUGCGGUGA**A**GCGGCUUUAGCCUUGGCGCGGCUU UCACACCCCUGUGCCGUGUGCCAUCUUGGUGCGC CUCG**AGA**UCG	78	TB9Cs4H2	SnR37	1171-LA
LM36Cs1H1	CGUGGA**C**GGAACGUGGCAGGCUGCGGAGGGCACAC CCAUCCCUCAUACGCGGUUCAACACAUGAUCCC**A GA**GCUU	75	TB10Cs1H3		12-S
LM36Cs1H2	CGGC**A**UCCCUUGUUGAUAGCGAGCAGUCCACCCGC AUCCCCGCAAGCUGCUCGUCUACACCGAGGG**AG A**GAA	72	TB10Cs1H1		703-LB
LM36Cs1H3	A**A**GCGCUCAUUUUCAAGCCGCACGCGGGCGUGUUG CUGGUGCCCCUGCGGUCGGGUGAUGCGCG**AG A**GCG	70	TB10Cs1H2		672-LA
LM36Cs1H4	**AG**CAGUGUCGCUAUGCGUCUCCGUCUCACCUCGAU GCGGAGCAACGGUACGAAACACGUGCG**AGA**UGC	68	TB10Cs3H1		
LM36Cs2H1	GACCC**A**CCUGCACUUCGCUCGUCCCGCUCGCGUGU CGAUGAGUGGUGCCGAUGGAUGGUGUGAGGG**AG A**GCG	72	TB8Cs2H1		
LM36Cs3H1	UCUCACGAACAAUCCCGCCCGGAGUUUACCGUACU UGAUCUCUUGCGCGGACCUACGGGUGCCCAG ATGC	70			
LM36Cs3H2	UGC**A**GCUGCGCCAAAGGGGUCUUUGUCAAAUUCCU UGCAGCGGCUGGUGCG**AGA**UGG	57			

[a] The sequences (±3 nt at the 5′ end) were obtained from the *L. major* genome database. The size of the coding region was predicted based on the position of the AGA box and the secondary structure. The sizes of the RNAs are indicated. Homologues in *T. brucei* and other organisms are given. The potential targets on rRNAs are indicated as described in Table 1. All sequences can be found in the *L. major* genome database (http://www.genedb.org/genedb/lies).

modifications observed are only 2 nt apart. Additional mapping data for 20 Nms located in four conserved rRNA domains indicated that 95% of the modifications were conserved between *L. major* and *T. brucei*. As shown in Fig. 4A, although the positions of the modifications were conserved, their intensities differed. It was of special interest to examine whether there are *Leishmania*-specific modifications that are guided by *Leishmania*-specific snoRNAs. To this end, we identified five snoRNAs with no conserved target that are missing in *T. brucei* (Lm35Cs2C2, Lm25Cs1C2, Lm36Cs2C3, Lm26Cs2C1, and Lm20Cs1C5). We mapped the regions which are potentially guided by these snoRNAs, but no modification was detected in either *T. brucei* or *L. major*. It should be noted that these snoRNAs are poorly expressed and that their modification levels may be too low to be detected in our experiments, or alternatively, these snoRNAs may represent pseudogenes. Another appealing possibility is that these snoRNAs are developmentally regulated and might be expressed in the amastigote stage of the parasite.

Expression levels of snoRNAs. All of the clusters (except one) in *T. brucei* were found to be expressed (24). However, different snoRNAs present within the same cluster were not always expressed at the same level (22). To examine whether this was also the case for *L. major*, the levels of both C/D and H/ACA RNAs within the same cluster were examined. The results (Fig. 5A and B) demonstrate that the levels of both C/D and H/ACA RNAs differed substantially. No correlation was found between the level of snoRNA and the copy numbers of the genes. The expression level of each C/D snoRNA was determined by densitometric analysis, relative to that of snoRNA LM33Cs1H1, and is shown in Fig. 6C. The levels of several H/ACA RNAs were also investigated by Northern analyses (Fig. 5B), and the levels of snoRNAs present in the same cluster varied significantly.

The differential expression of snoRNAs encoded in the same cluster raises the question of which factors determine the level of snoRNAs present within a given cluster. The RNA level is most probably not affected by transcription because of the polycistronic nature of snoRNA transcription (11, 38, 45). Factors that can affect the steady-state level of snoRNAs include the efficiency of processing and the half-life of the RNAs. To this end, we examined, using mfold (http://www.bioinfo.rpi.edu/applications/mfold/old/rna/), the potential of sequences flanking the snoRNAs to form external and internal stems (Fig. 6A). As opposed to the case in *L. collosoma*, in which sequences flanking the snoRNAs form extended stems (45), the external stems flanking the *L. major* snoRNAs are only 6 to 20 nt long (Fig. 6A). To examine the possible correlation between the stability of the stem and snoRNA abundance, the ΔG values of the extragenic stems were calculated using the mfold program (Fig. 6C). The ΔG values ranged from -0.7 to -30 kcal/mole (average of -16.5 ± 7.8 kcal/mol). The ΔG values of the external stems suggest that such stems most probably exist, unlike the putative internal stems, whose ΔG values are higher (-3.9 ± 3.4 kcal/mol) and which most probably are not formed (Fig. 6C). Thus, we suggest that the external stem is formed and is a real structural feature of C/D snoRNAs in trypanosomes. In contrast, the stability of the internal stem is very weak, and this stem either does not form at all or exists only transiently. However, no direct correlation was found between the strength of the external stem and RNA abundance (Fig. 5A and 6C).

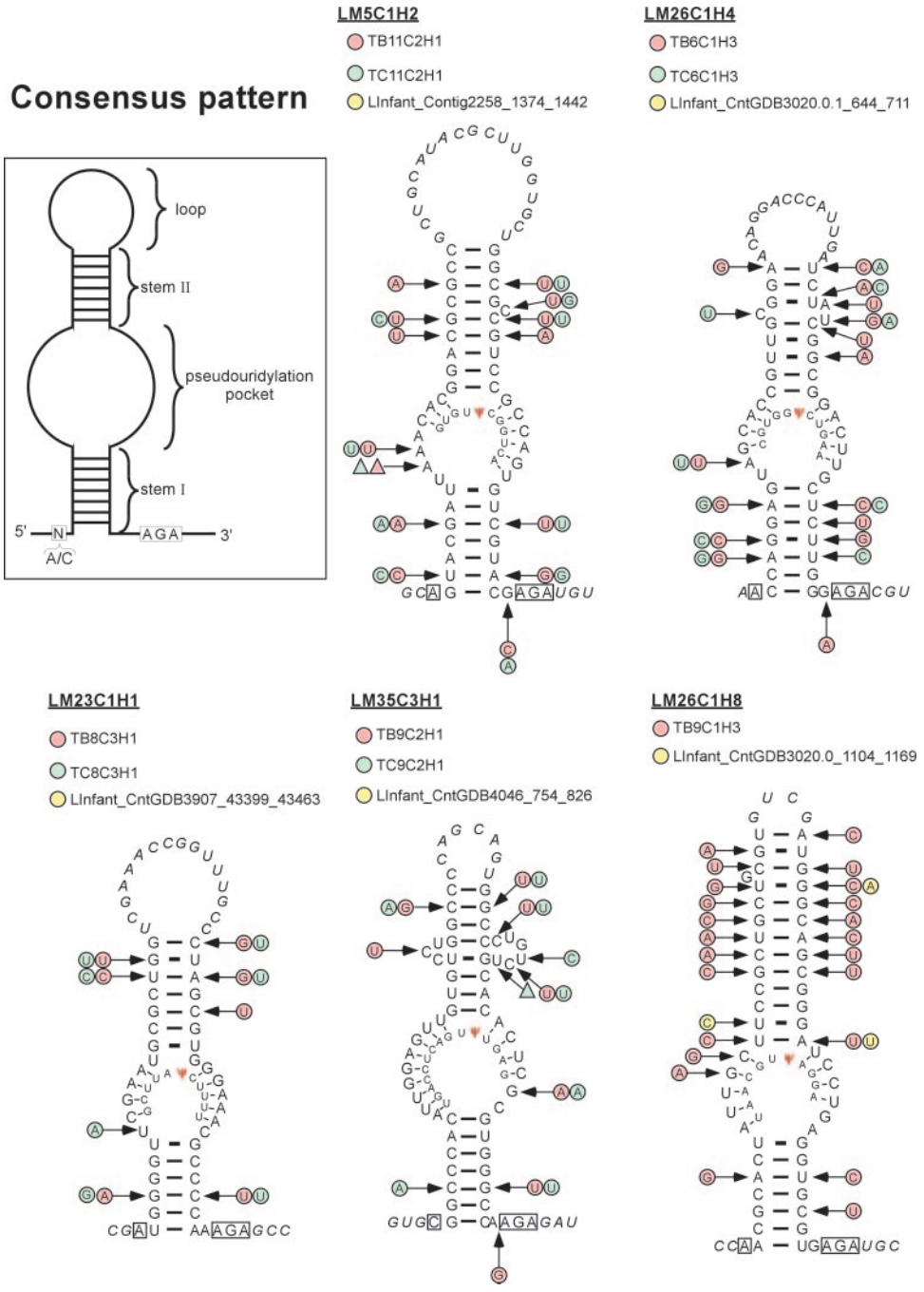


FIG. 2. Conservation of H/ACA secondary structures between different trypanosomatid species. The proposed H/ACA-like snoRNA secondary structure was determined by mfold, taking into consideration the complementarity with the target and the distance between AGA and Ψ. The sequences of *L. major* snoRNAs are given, and differences from the *L. major* sequences are marked in colored circles for *T. brucei* (red), *T. cruzi* (green), and *L. infantum* (yellow). Deletions are indicated by triangles. The species designations are as follows: TB, *Trypanosoma brucei*; TC, *Trypanosoma cruzi*; LM, *Leishmania major*; and Linfant, *Leishmania infantum*. The AGA box and the conserved nucleotide upstream from stem I are boxed. A schematic representation of the H/ACA consensus pattern is boxed, and the structural motifs are marked.

Another structural feature of the C/D snoRNAs that may affect the expression level is the kink-turn motif (K turn), formed by interactions between the C and D boxes located near the ends of the RNAs (35, 40). The putative canonical K-turn structure (32) is depicted in Fig. 6B, comparing the consensus K turn with the potential of LM20Cs1H1 to form an analogous structure. This motif is composed of stems I and II (Fig. 6B). Noncanonical base pairing, such as G-A and U-U pairs, have been shown to be important for the formation of the K turn and to determine the binding of Snu13 to box C/D (26). Efficiency of binding of Snu13 may affect the fate of snoRNA and determine whether the snoRNA will assemble to

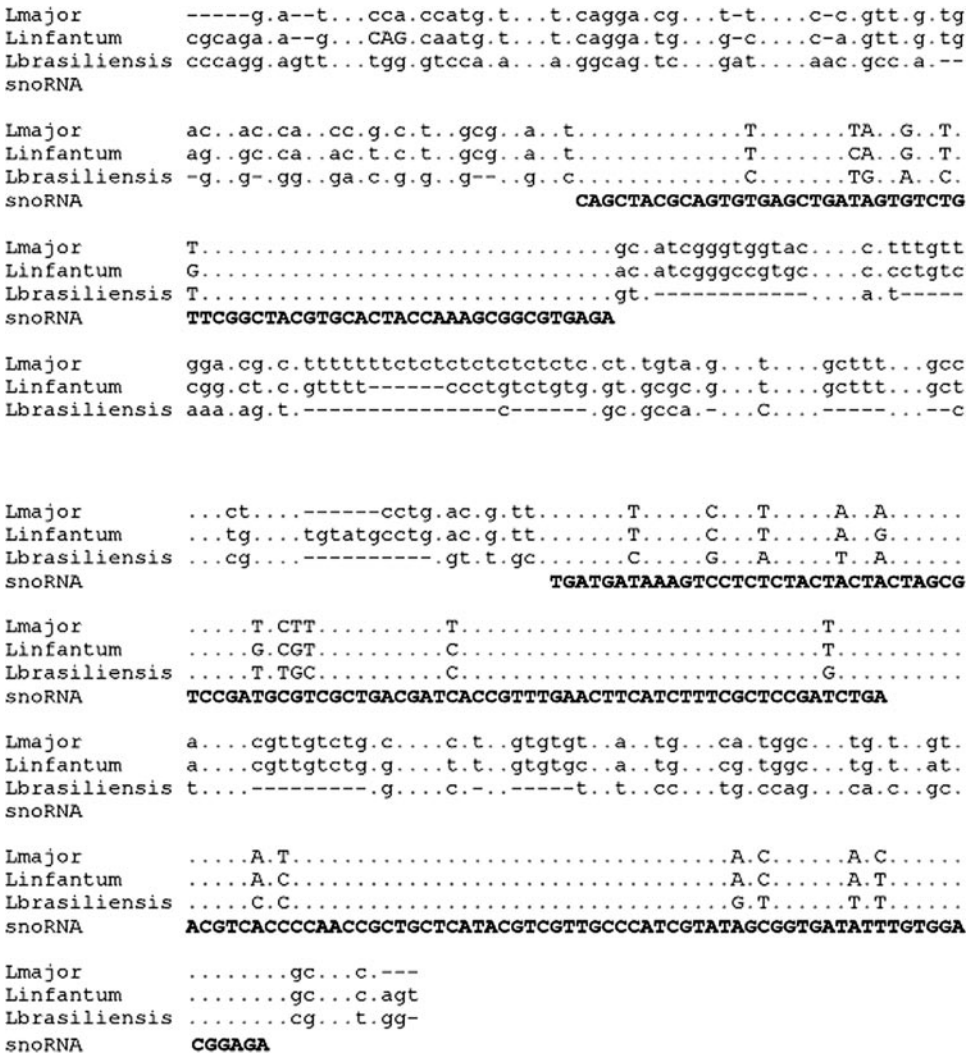


FIG. 3. Sequence alignment of snoRNA clusters from different *Leishmania* species. Multiple alignment of the various *Leishmania* species was performed using CLUSTAL W on the entire LM34Cs1 cluster. The conserved nucleotides are shown as dots, and gaps are indicated by hyphens. The snoRNA consensus sequence is marked below in bold. The coding regions are shown in capital letters, while intragenic regions are marked in lowercase letters. The species designations are as follows: Lmajor, *Leishmania major* (LmjF34.snoRNA.0003); Linfantum, *Leishmania infantum* (LinJ34_20050901_V2.0); and Lbrasiliensis, *Leishmania braziliensis* (brazil865f07.p1k).

snoRNP or be degraded by the exosome during processing (8). We therefore examined whether there is a correlation between the strength of conservation of the K-turn motif and the level of the snoRNA. The potential of several snoRNAs to form a K turn is indicated in Fig. 6A and was qualitatively assessed as shown in Fig. 6C. Other factors that may affect snoRNA processing and binding of Snu13 are the C and D box sequences (26, 40). To this end, we inspected the conservation of the C and D boxes, i.e., CUGA and UGAUGA, respectively. Those that are conserved are marked with plus signs, and those that deviate by at least 1 nt are marked with minus signs (Fig. 6C). We found that if the two boxes deviate from the consensus, the snoRNA is poorly expressed. However, high conservation of the boxes is not sufficient to ensure high expression, as in the case of Lm35Cs2C1 (Fig. 6C). Inspection of the data in Fig. 6A and C suggests that there is no single structural feature that fully correlates with the abundance of the snoRNA. For instance, the K turns of all molecules inspected agree with the canonical structure mostly in stem II (A-G pairing), but there are cases of K turns which fit closely with the consensus and yet the snoRNAs harboring them are poorly expressed (Fig. 6C).

DISCUSSION

In this study, we expanded our understanding of the exceptionally rich repertoire of snoRNA genes in several trypanosomatid species. We used the genome collection of snoRNAs recently described for *Trypanosoma brucei* to search for homologues in the *Leishmania major* genome. Our analysis led to the identification of over 99 snoRNAs in this organism. The *L. major* snoRNA genes identified in this study are organized into 23 clusters encoding 62 C/Ds that can potentially guide 79 methylations. In addition, 37 H/ACA-like RNAs are found in these clusters that can potentially guide pseudouridine forma-

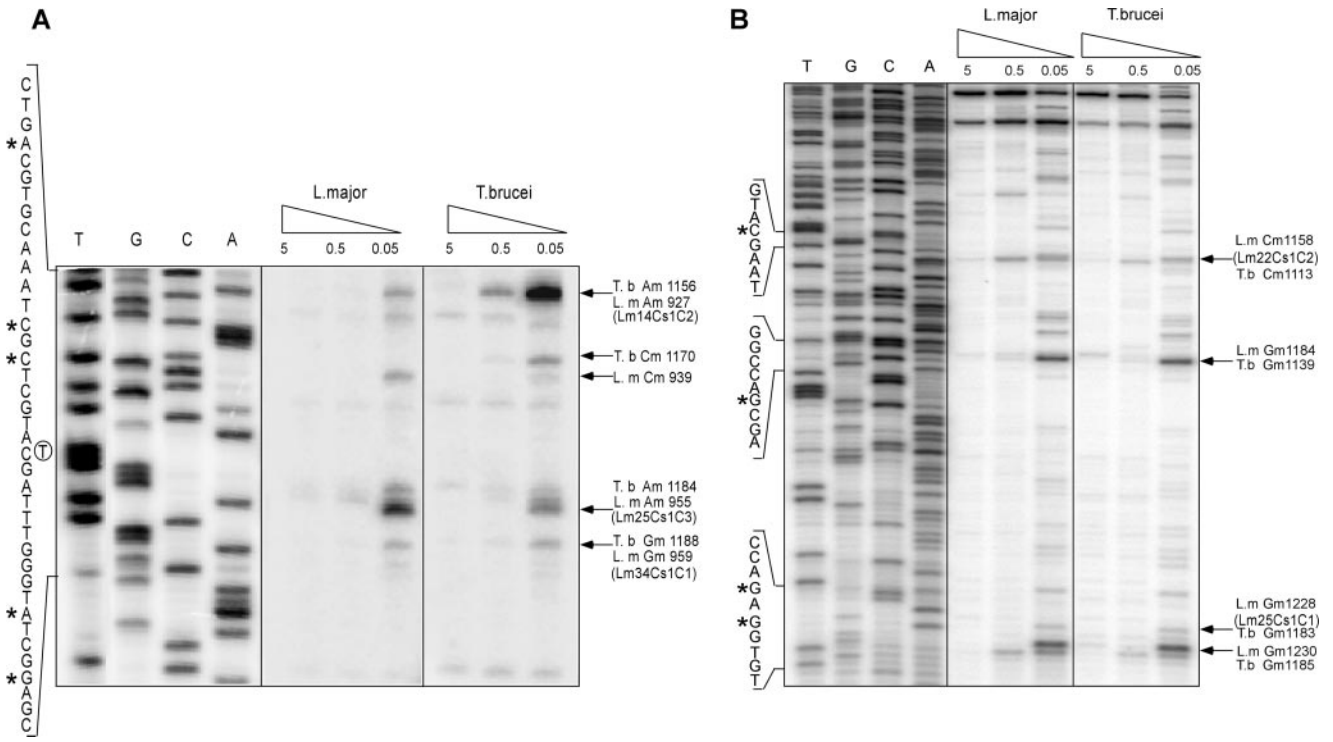


FIG. 4. Conservation of modified Nm nts on *T. brucei* and *L. major* rRNAs. Total RNAs from *L. major* and *T. brucei* were subjected to primer extension using different concentrations of dNTPs (0.05, 0.5, and 5 mM) to detect the Nm modification. The primers used for this mapping were specific to 5′ LSU rRNA (A) and 3′ LSU rRNA (B) (using oligonucleotides 1238 and 1217, respectively). Extension products were analyzed in a 6% polyacrylamide–7 M urea gel next to DNA sequencing products. Sequencing was performed using plasmids containing *T. brucei* rRNA as a template, with the same oligonucleotides as those used for primer extension. A partial DNA sequence is given on the left, and the modified nts are denoted by asterisks. The nts that differ in *L. major* from *T. brucei* are circled. The stops representing 1 nt before the modified nts are indicated on the right by arrows. The positions of the modified nts in *T. brucei* and *L. major* are indicated. The names of *L. major* snoRNAs that potentially guide the modification are given in parentheses.

tion at 30 sites, suggesting the existence of a rich repertoire of modifications and guide RNAs. This study led to the identification of 10 novel H/ACA RNAs and 14 C/D RNAs that were not identified in *T. brucei*. Mapping of Nms in rRNAs suggested the existence of trypanosomatid-specific modifications which are 95% conserved between *T. brucei* and *L. major*. For the most part, the snoRNA clusters are syntenious, but the number of cluster repeats is higher in *L. major* than in *T. brucei*. Phylogenic analysis of the individual H/ACA-like sequences identified conserved structural features among these entire molecules.

Comparative genomics between *L. major* and *T. brucei* snoRNA clusters. For the most part, the repertoire of snoRNA genes is conserved between *T. brucei* and *L. major*. However, the snoRNA sequences described here and previously described for *T. brucei* represent only part of the estimated repertoire. Mapping of Nms in *T. brucei* revealed the existence of an additional 52 novel Nms that were found mainly in domains already rich in modifications. These modifications were reduced in cells depleted of fibrillarin by RNA interference, suggesting that they are guided by C/D snoRNPs (S. Barth and S. Michaeli, unpublished data). Among these modifications are those conserved in plants, yeast, and mammals, suggesting that snoRNAs that guide these modifications most probably exist. Why were these snoRNAs not detected in the entire genomes of both *L. major* and *T. brucei?* Our search for *T. brucei* snoRNAs was based on the identification of repeats, using the SnoScan program and comparative genomics with *T. cruzi* (24). This suggests that our screen was biased towards the identification of repeated snoRNAs, but snoRNAs which are not encoded in repeated clusters may also exist.

Finding these missing snoRNAs by either experimental or computational strategies turned out to be complicated. Examining the scores of individual snoRNAs by SnoScan (http://lowelab.UCSC.edu/snoscan) suggested that many known snoRNAs in trypanosomes receive low scores, explaining why it is very difficult to find additional snoRNAs by performing a genome search using this program. The finding of novel C/D snoRNAs using such a genome-scale search seems to be a complicated task because the conservation of these RNAs is relatively low and exists only in the boxes and in the regions of complementarity to the targets (24, 38). Thus, we cannot rule out the possibility of the existence of a novel family of guide RNAs that contain consensus boxes which strongly deviate from the C/D consensus and guide modifications present on rRNAs. A program to search for H/ACA-like molecules in the *T. brucei* genome is being developed (I. Myslyuk, Y. Horesh, S. Michaeli, and R. Unger, unpublished results). There is great hope that the identification of novel H/ACA RNA-containing clusters will reveal some of the missing C/D snoRNAs as well, since most snoRNA genes identified so far are present in clusters that contain both H/ACA and C/D snoRNA genes.

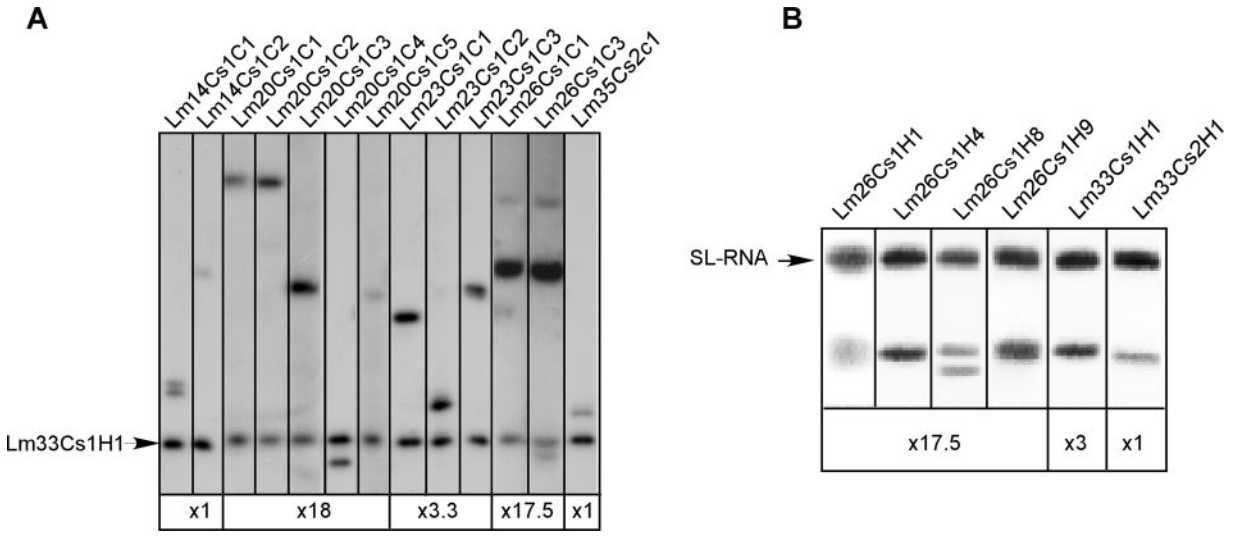


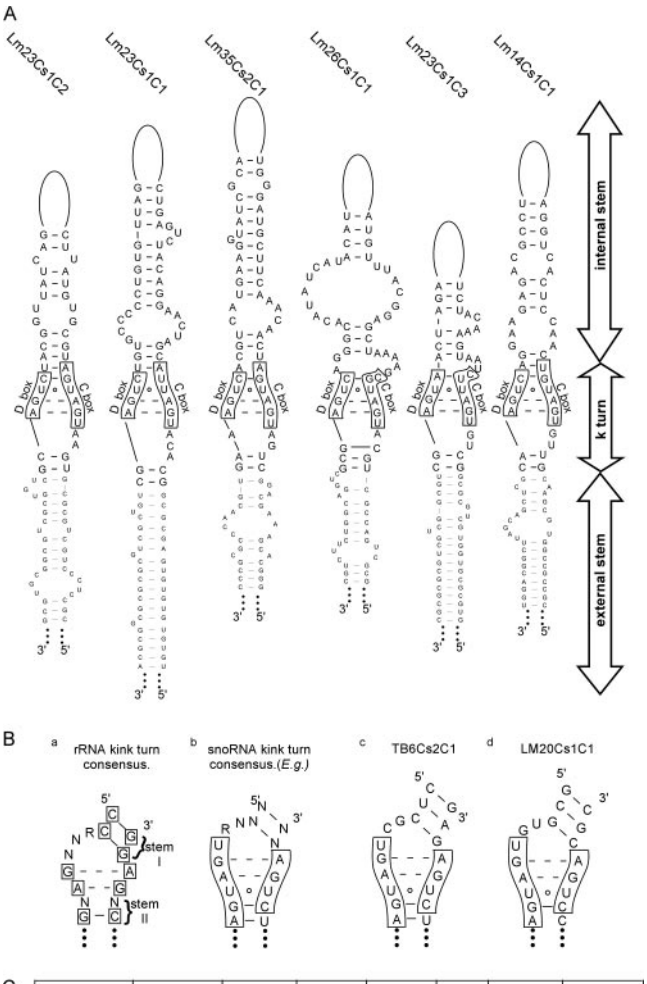
FIG. 5. Expression of selected snoRNAs. (A) Levels of C/D snoRNAs. Total RNA from *L. major* was subjected to primer extension using primers specific to LM14Cs1C1, LM14Cs1C2, LM20Cs1C1, LM20Cs1C2, LM20Cs1C3, LM20Cs1C4, LM20Cs1C5, LM23Cs1C1, LM23Cs1C2, LM23Cs1C3, LM26Cs1C1, LM26Cs1C3, and LM35Cs2C1. The level of snoRNA LM33Cs1H1 was used to control the amount of RNA in each sample. Extension products were analyzed in a 6% polyacrylamide–7 M urea gel. (B) Levels of H/ACA-like snoRNAs. Total RNA from *L. major* was separated in a 10% polyacrylamide–7 M urea gel. The RNAs were electroblotted and hybridized with probes complementary to LM26Cs1H1, LM26Cs1H4, LM26Cs1H8, LM26Cs1H9, LM33Cs1H1, and LM33Cs2H1. The level of SL RNA was used to control for equal loading. The numbers at the bottom indicate the copy numbers of the clusters.

It is interesting that some C/D and H/ACA RNAs were identified in *L. major* but not in *T. brucei*. Among these RNAs are two snoRNAs with no predicted targets (Lm25Cs1C4 and Lm35Cs2C3) and three snoRNAs that can potentially guide modifications conserved in other organisms (LM30Cs1C2, LM33Cs3C1, and LM18Cs1C2). Seven snoRNAs with no known targets in *T. brucei* (Lm14Cs1C1, Lm35Cs2C2, Lm25Cs1C2, Lm36Cs2C3, Lm26Cs2C1, Lm20Cs1C4, and Lm20Cs1C5) were also identified and were found to be poorly expressed. In addition, we could not detect the predicted modifications on the rRNA. These could represent pseudogenes, very weak modifications, or alternatively, snoRNAs which are developmentally regulated.

Factors that may affect the differential abundance of snoRNAs present in *Leishmania*. The data presented in Fig. 5 and 6 highlight the observation already reported that the levels of snoRNAs expressed from the same gene cluster can vary considerably (22, 23, 24). This raises a question regarding the factors that affect the abundance of a certain snoRNA within the cell. The transcriptional regulation of snoRNA genes cannot be disputed at this point, since a genomic region (~700 bp) upstream of a snoRNA gene cluster enhances the expression of snoRNA genes present on a multicopy plasmid (23). However, we could not identify consensus sequences upstream of the snoRNA clusters in either of these organisms. A rigorous experimental approach is needed to identify such sequences. Interestingly, we recently identified snoRNAs whose levels are enhanced in the bloodstream form of *T. brucei* compared to the procyclic state, suggesting that snoRNA gene expression might be developmentally regulated (Barth and Michaeli, unpublished data).

The expression of a snoRNA cluster can be influenced by its location near a "real" promoter. Recently, it was shown that promoter-like regions exist in *L. major* and are located in a strand-switching region between two long polycistronic transcripts (27). Indeed, in one special case, i.e., LM5Cs1, which carries SLA1, the cluster is located near an inflection point. This location enhances the transcription of SLA1 and accompanying snoRNAs, and indeed these RNAs are the most abundant snoRNAs (except for U3). It will therefore be of interest to examine the "strength" of this "promoter-like" element and to compare it to sequences present upstream of other snoRNA clusters which are not present at inflection points.

Most intriguing is the differential expression of snoRNA genes present in the same cluster, as clearly demonstrated in this study (Fig. 5 and 6). Transcription regulation cannot explain these differences, since these snoRNAs are processed from the same transcripts. At least two factors may influence the levels of such snoRNAs, namely, the efficiency of processing and the strength by which these snoRNAs bind their cognate binding proteins. To date, none of the factors that mediate snoRNA processing have been identified in trypanosomes. The only study performed with *L. collosoma* suggests that 10 extragenic flanking nucleotides are sufficient to govern efficient processing of the snoRNA (23, 45). Studies from mammals and yeast suggest that base pairing of the 5′-3′-terminal stem of the snoRNA coding sequence is required for the processing and accumulation of box C/D snoRNAs (7, 15, 41, 44). However, many intron-encoded box C/D snoRNAs expressed in mammalian cells or polycistronic snoRNAs in yeast lack the canonical 5′-3′-terminal helix (8). Studies on such snoRNAs indicated that the processing of these C/D snoRNAs is supported by external intronic stem structures that are fully or partially degraded during exonucleolytic cleavage (8). The trypanosome C/D snoRNAs resemble such snoRNAs, lacking the 5′-3′ stem but having, instead, extragenic flanking helices. However, the



C

snoRNA Name	Copy number	External Stem (ΔG)	C box consensus	D box consensus	K turn	Internal Stem (ΔG)	Expression
Lm14Cs1C1	1	-0.7	-	+	++	-5.4	4.6
Lm14Cs1C2	1	-	-	-	-	-	1.0
Lm20Cs1C1	18	-17.3	+	+	++	-3.5	8.0
Lm20Cs1C2	18	-	-	+	+	-11	16.8
Lm20Cs1C3	18	-18.2	+	+	++	-4.3	25.6
Lm20Cs1C4	18	-	-	-	-	-	4.9
Lm20Cs1C5	18	-	-	+	+	-	1.8
Lm23Cs1C1	3.3	-26.1	-	+	+	-5.1	13.3
Lm23Cs1C2	3.3	-11	+	+	++	-1.6	11.7
Lm23Cs1C3	3.3	-29.6	-	-	+	0.3	11.4
Lm26Cs1C1	17.5	-9.7	-	+	++	0.7	53.6
Lm26Cs1C3	17.5	-13.2	+	+	++	-2.0	69.6
Lm35Cs2c1	1	-9.4	+	+	-	-6.3	3.0

FIG. 6. Relationships among predicted secondary structure, K-turn motif, and structural features of C/D snoRNAs to their expression level. (A) C/D snoRNA secondary structures were predicted using the mfold program (http://www.bioinfo.rpi.edu/applications/mfold/old/rna/). Flanking sequences appear in small letters, and coding sequences appear in large letters. The conserved C and D sequences are boxed and indicated. The extragenic stem, K-turn motif, and intragenic stem are indicated by arrows on the right. Noncanonical base pairing of the K turn is indicated by a dashed line. (B) Schematic representation of K-turn motifs. (a) Consensus motif of rRNA; (b) consensus motif of *E. gracilis* snoRNA (*E.g.*) (32); (c) K turn from *T. brucei* snoRNA; (d) K turn from *L. major* snoRNA. (C) Table summarizing the structural features of C/D snoRNAs. The degree of expression is given in relative numbers. The ΔG values (kcal/mol) of the duplexes (external and internal stems) were calculated using mfold. The copy numbers of the genes are indicated. +, C/D boxes which agree with the consensus (cited in the text); −, boxes that deviate from the consensus. The degree of resemblance to the K-turn consensus was graded as follows: −, strong deviation from the consensus; +, partial resemblance to the canonical structure; and ++, cases that fully obey the consensus. Expression levels were calculated using densitometric analysis and are presented in arbitrary units (1 represents the snoRNA expressed at its lowest level).

data presented in Fig. 5A and 6C indicate that although the external stems are thermodynamically favorable, no significant correlation can be found between the abundance of the snoRNA and the stability of the external duplex. Other structural factors, such as the presence of a canonical K turn and conserved C and D boxes, also influence snoRNA abundance. The conclusion from our analysis suggests that efficient expression of trypanosomatid snoRNAs takes place if the extragenic flanking sequences of the snoRNAs form an extended stem and/or the snoRNAs can form a canonical K turn with conserved boxes. None of these factors exclusively affects snoRNA abundance. However, it is possible that regulatory sequences situated in the flanking sequences or other structural factors that were overlooked in this study affect snoRNA abundance. The level of snoRNA may be governed by combinatorial effects of all the factors discussed.

Even less is known about factors that affect the processing and stability of H/ACA RNAs (20). The phylogenetic analysis in Fig. 3 highlights the most important structural features of these RNAs, which are the conserved AGA box, the distance between the AGA box and Ψ, and the lengths of stem I and part of stem II. These are the structural features that were shown to dictate the in vitro binding of CBF5 to archaeal H/ACA RNAs (3). The differential expression of H/ACA RNAs present in the same cluster observed in this study may result from the differential binding of CBF5 to the H/ACA RNA molecule. In addition, the sequences flanking the coding region may also affect snoRNA processing.

Conservation of rRNA Nm modifications among the trypanosomatids and their biological role. For the most part, the modifications mapped in *L. major* coincide with the modifications present in *T. brucei* rRNAs, except for a few cases where the modification was slightly shifted. As in *T. brucei*, the number of Nms in *L. major* is almost double the number of pseudouridines. We previously suggested that the need for a large number of Nms may help the parasite to cope with growth at an elevated temperature in its mammalian host (37°C) compared to 26°C in the insect host (24). Support for this role of Nms in trypanosomes was recently obtained when we observed an elevation in the level of certain modifications in bloodstream-stage *T. brucei* parasites compared to the procyclic stage (Barth and Michaeli, unpublished data). It will be interesting to examine whether the level of modification is also higher in *L. major* amastigotes than in promastigotes. Large numbers of Nms were also reported for plants, and this finding was rationalized by the fact that plants are exposed to large temperature changes, during which the ribosomes must be produced and remain active. Interestingly, plants (6) and *Euglena gracilis* (31)

are the only eukaryotes studied so far that carry mixed clusters containing both C/D and H/ACA RNAs. The mixed snoRNA cluster may have developed independently at least twice in evolution, since there is no evidence for large-scale horizontal transfer of genetic material from plants to trypanosomes (13).

In sum, this study highlights the strengths of comparative genomics among related trypanosomatid species for the identification of snoRNA genes, and most probably other noncoding RNAs. The unique features of trypanosome snoRNAs, i.e., a greater number of C/D than H/ACA RNAs, single-hairpin H/ACA RNAs, and reiterated clusters composed of C/D and H/ACA RNAs, are all conserved in this family of parasites. The pattern of Nm modification is also highly conserved. Based on conserved features and deviations among the H/ACA RNAs, a canonical trypanosome H/ACA RNA was established that is currently in use for developing an algorithm to search for these RNAs on a genomic scale. The nonconserved snoRNA intergenic regions present in Old and New World *Leishmania* species could be used to differentiate *Leishmania*-related species and to identity novel noncoding RNAs in the *Leishmania* genome.

ACKNOWLEDGMENTS

This research was supported by a grant from the Israel Science Foundation, a grant from the Israeli Ministry of Science and Technology to S.M. and R.U., and an International Research Scholar's Grant from the Howard Hughes Foundation to S.M.

REFERENCES

1. **Aspegren, A., A. Hinas, P. Larsson, A. Larsson, and F. Söderbom.** 2004. Novel non-coding RNAs in *Dictyostelium discoideum* and their expression during development. Nucleic Acids Res. **32:**4646–4656.
2. **Bachellerie, J. P., J. Cavaillé, and A. Hüttenhofer.** 2002. The expanding snoRNA world. Biochimie **84:**775–790.
3. **Baker, D. L., O. A. Youssef, M. I. Chastkofsky, D. A. Dy, R. M. Terns, and M. P. Terns.** 2005. RNA-guided RNA modification: functional organization of the archaeal H/ACA RNP. Genes Dev. **19:**1238–1248.
4. **Barth, S., A. Hury, X. H. Liang, and S. Michaeli.** 2005. Elucidating the role of H/ACA-like RNAs in *trans*-splicing and rRNA processing via RNA interference silencing of the *Trypanosoma brucei* CBF5 pseudouridine synthase. J. Biol. Chem. **280:**34558–34568.
5. **Berriman, M., E. Ghedin, C. Hertz-Fowler, G. Blandin, H. Renauld, D. C. Bartholomeu, N. J. Lennard, E. Caler, N. E. Hamlin, B. Haas, U. Bohme, L. Hannick, M. A. Aslett, J. Shallom, L. Marcello, L. Hou, B. Wickstead, U. C. Alsmark, C. Arrowsmith, R. J. Atkin, A. J. Barron, F. Bringaud, K. Brooks, M. Carrington, I. Cherevach, T. J. Chillingworth, C. Churcher, L. N. Clark, C. H. Corton, A. Cronin, R. M. Davies, J. Doggett, A. Djikeng, T. Feldblyum, M. C. Field, A. Fraser, I. Goodhead, Z. Hance, D. Harper, B. R. Harris, H. Hauser, J. Hostetler, A. Ivens, K. Jagels, D. Johnson, J. Johnson, K. Jones, A. X. Kerhornou, H. Koo, N. Larke, S. Landfear, C. Larkin, V. Leech, A. Line, A. Lord, A. Macleod, P. J. Mooney, S. Moule, D. M. Martin, G. W. Morgan, K. Mungall, H. Norbertczak, D. Ormond, G. Pai, C. S. Peacock, J. Peterson, M. A. Quail, E. Rabbinowitsch, M. A. Rajandream, C. Reitter, S. L. Salzberg, M. Sanders, S. Schobel, S. Sharp, M. Simmonds, A. J. Simpson, L. Tallon, C. M. Turner, A. Tait, A. R. Tivey, S. Van Aken, D. Walker, D. Wanless, S. Wang, B. White, O. White, S. Whitehead, J. Woodward, J. Wortman, M. D. Adams, T. M. Embley, K. Gull, E. Ullu, J. D. Barry, A. H. Fairlamb, F. Opperdoes, B. G. Barrell, J. E. Donelson, N. Hall, C. M. Fraser, S. E. Melville, and N. M. El Sayed.** 2005. The genome of the African trypanosome *Trypanosoma brucei*. Science **309:**416–422.
6. **Brown, J. W., G. P. Clark, D. J. Leader, C. G. Simpson, and T. Lowe.** 2001. Multiple snoRNA gene clusters from Arabidopsis. RNA **7:**1817–1832.
7. **Cavaillé, J., and J. P. Bachellerie.** 1996. Processing of fibrillarin-associated snoRNAs from pre-mRNA introns: an exonucleolytic process exclusively directed by the common stem-box terminal structure. Biochimie **78:**443–456.
8. **Darzacq, X., and T. Kiss.** 2000. Processing of intron-encoded box C/D small nucleolar RNAs lacking a 5′,3′-terminal stem structure. Mol. Cell. Biol. **20:**4522–4531.
9. **Decatur, W. A., and M. J. Fournier.** 2003. RNA-guided nucleotide modification of ribosomal and other RNAs. J. Biol. Chem. **278:**695–698.
10. **Dennis, P. P., A. Omer, and T. Lowe.** 2001. A guided tour: small RNA function in archaea. Mol. Microbiol. **40:**509–519.
11. **Dunbar, D. A., S. Wormsley, T. M. Lowe, and S. J. Baserga.** 2000. Fibrillarin-associated box C/D small nucleolar RNAs in *Trypanosoma brucei*. Sequence conservation and implications for 2′-*O*-ribose methylation of rRNA. J. Biol. Chem. **275:**14767–14776.
12. **El-Sayed, N. M., P. J. Myler, D. C. Bartholomeu, D. Nilsson, G. Aggarwal, A. N. Tran, E. Ghedin, E. A. Worthey, A. L. Delcher, G. Blandin, S. J. Westenberger, E. Caler, G. C. Cerqueira, C. Branche, B. Haas, A. Anupama, E. Arner, L. Aslund, P. Attipoe, E. Bontempi, F. Bringaud, P. Burton, E. Cadag, D. A. Campbell, M. Carrington, J. Crabtree, H. Darban, J. F. da Silveira, P. de Jong, K. Edwards, P. T. Englund, G. Fazelina, T. Feldblyum, M. Ferella, A. C. Frasch, K. Gull, D. Horn, L. Hou, Y. Huang, E. Kindlund, M. Klingbeil, S. Kluge, H. Koo, D. Lacerda, M. J. Levin, H. Lorenzi, T. Louie, C. R. Machado, R. McCulloch, A. McKenna, Y. Mizuno, J. C. Mottram, S. Nelson, S. Ochaya, K. Osoegawa, G. Pai, M. Parsons, M. Pentony, U. Pettersson, M. Pop, J. L. Ramirez, J. Rinta, L. Robertson, S. L. Salzberg, D. O. Sanchez, A. Seyler, R. Sharma, J. Shetty, A. J. Simpson, E. Sisk, M. T. Tammi, R. Tarleton, S. Teixeira, S. Van Aken, C. Vogt, P. N. Ward, B. Wickstead, J. Wortman, O. White, C. M. Fraser, K. D. Stuart, and B. Andersson.** 2005. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. Science **309:**409–415.
13. **El-Sayed, N. M., P. J. Myler, G. Blandin, M. Berriman, J. Crabtree, G. Aggarwal, E. Caler, H. Renauld, E. A. Worthey, C. Hertz-Fowler, E. Ghedin, C. Peacock, D. C. Bartholomeu, B. J. Haas, A. N. Tran, J. R. Wortman, U. C. Alsmark, S. Angiuoli, A. Anupama, J. Badger, F. Bringaud, E. Cadag, J. M. Carlton, G. C. Cerqueira, T. Creasy, A. L. Delcher, A. Djikeng, T. M. Embley, C. Hauser, A. C. Ivens, S. K. Kummerfeld, J. B. Pereira-Leal, D. Nilsson, J. Peterson, S. L. Salzberg, J. Shallom, J. C. Silva, J. Sundaram, S. Westenberger, O. White, S. E. Melville, J. E. Donelson, B. Andersson, K. D. Stuart, and N. Hall.** 2005. Comparative genomics of trypanosomatid parasitic protozoa. Science **309:**404–409.
14. **Ganot, P., M. L. Bortolin, and T. Kiss.** 1997. Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. Cell **89:**799–809.
15. **Huang, G. M., A. Jarmolowski, J. C. Struck, and M. J. Fournier.** 1992. Accumulation of U14 small nuclear RNA in *Saccharomyces cerevisiae* requires box C, box D, and a 5′,3′-terminal stem. Mol. Cell. Biol. **12:**4456–4463.
16. **Ivens, A. C., C. S. Peacock, E. A. Worthey, L. Murphy, G. Aggarwal, M. Berriman, E. Sisk, M. A. Rajandream, E. Adlem, R. Aert, A. Anupama, Z. Apostolou, P. Attipoe, N. Bason, C. Bauser, A. Beck, S. M. Beverley, G. Bianchettin, K. Borzym, G. Bothe, C. V. Bruschi, M. Collins, E. Cadag, L. Ciarloni, C. Clayton, R. M. Coulson, A. Cronin, A. K. Cruz, R. M. Davies, J. De Gaudenzi, D. E. Dobson, A. Duesterhoeft, G. Fazelina, N. Fosker, A. C. Frasch, A. Fraser, M. Fuchs, C. Gabel, A. Goble, A. Goffeau, D. Harris, C. Hertz-Fowler, H. Hilbert, D. Horn, Y. Huang, S. Klages, A. Knights, M. Kube, N. Larke, L. Litvin, A. Lord, T. Louie, M. Marra, D. Masuy, K. Matthews, S. Michaeli, J. C. Mottram, S. Muller-Auer, H. Munden, S. Nelson, H. Norbertczak, K. Oliver, S. O'Neil, M. Pentony, T. M. Pohl, C. Price, B. Purnelle, M. A. Quail, E. Rabbinowitsch, R. Reinhardt, M. Rieger, J. Rinta, J. Robben, L. Robertson, J. C. Ruiz, S. Rutter, D. Saunders, M. Schafer, J. Schein, D. C. Schwartz, K. Seeger, A. Seyler, S. Sharp, H. Shin, D. Sivam, R. Squares, S. Squares, V. Tosato, C. Vogt, G. Volckaert, R. Wambutt, T. Warren, H. Wedler, J. Woodward, S. Zhou, W. Zimmermann, D. F. Smith, J. M. Blackwell, K. D. Stuart, B. Barrell, and P. J. Myler.** 2005. The genome of the kinetoplastid parasite, Leishmania major. Science **309:**436–442.
17. **Kiss, T.** 2002. Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. Cell **109:**145–148.
18. **Kiss-László, Z., Y. Henry, J. P. Bachellerie, M. Caizergues-Ferrer, and T. Kiss.** 1996. Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. Cell **85:**1077–1088.
19. **Lange, T. S., A. Borovjagin, E. S. Maxwell, and S. A. Gerbi.** 1998. Conserved boxes C and D are essential nucleolar localization elements of U14 and U8 snoRNAs. EMBO J. **17:**3176–3187.
20. **Lange, T. S., M. Ezrokhi, F. Amaldi, and S. A. Gerbi.** 1999. Box H and box ACA are nucleolar localization elements of U17 small nucleolar RNA. Mol. Biol. Cell **10:**3877–3890.
21. **Liang, X. H., A. Haritan, S. Uliel, and S. Michaeli.** 2003. *trans* and *cis* splicing in trypanosomatids: mechanism, factors, and regulation. Eukaryot. Cell **2:**830–840.
22. **Liang, X. H., L. Liu, and S. Michaeli.** 2001. Identification of the first trypanosome H/ACA RNA that guides pseudouridine formation on rRNA. J. Biol. Chem. **276:**40313–40318.
23. **Liang, X. H., A. Ochaion, Y. X. Xu, Q. Liu, and S. Michaeli.** 2004. Small nucleolar RNA clusters in trypanosomatid *Leptomonas collosoma*. Genome organization, expression studies, and the potential role of sequences present upstream from the first repeated cluster. J. Biol. Chem. **279:**5100–5109.
24. **Liang, X. H., S. Uliel, A. Hury, S. Barth, T. Doniger, R. Unger, and S. Michaeli.** 2005. A genome-wide analysis of C/D and H/ACA-like small nucleolar RNAs in *Trypanosoma brucei* reveals a trypanosome-specific pattern of rRNA modification. RNA **11:**619–645.
25. **Liang, X. H., Y. X. Xu, and S. Michaeli.** 2002. The spliced leader-associated

RNA is a trypanosome-specific sn(o) RNA that has the potential to guide pseudouridine formation on the SL RNA. RNA **8:**237–246.

26. **Marmier-Gourrier, N., A. Cléry, V. Senty-Ségault, B. Charpentier, F. Schlotter, F. Leclerc, R. Fournier, and C. Branlant.** 2003. A structural, phylogenetic, and functional study of 15.5-kD/Snu13 protein binding on U3 small nucleolar RNA. RNA **9:**821–838.
27. **Martinez-Calvillo, S., D. Nguyen, K. Stuart, and P. J. Myler.** 2004. Transcription initiation and termination on *Leishmania major* chromosome 3. Eukaryot. Cell **3:**506–517.
28. **Nottrott, S., K. Hartmuth, P. Fabrizio, H. Urlaub, I. Vidovic, R. Ficner, and R. Lührmann.** 1999. Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5′ stem-loop of U4 snRNA. EMBO J. **18:**6119–6133.
29. **Omer, A. D., T. M. Lowe, A. G. Russell, H. Ebhardt, S. R. Eddy, and P. P. Dennis.** 2000. Homologs of small nucleolar RNAs in archaea. Science **288:**517–522.
30. **Roberts, T. G., N. R. Sturm, B. K. Yee, M. C. Yu, T. Hartshorne, N. Agabian, and D. A. Campbell.** 1998. Three small nucleolar RNAs identified from the spliced leader-associated RNA locus in kinetoplastid protozoans. Mol. Cell. Biol. **18:**4409–4417.
31. **Russell, A. G., M. N. Schnare, and M. W. Gray.** 2004. Pseudouridine-guide RNAs and other Cbf5p-associated RNAs in *Euglena gracilis*. RNA **10:**1034–1046.
32. **Russell, A. G., M. N. Schnare, and M. W. Gray.** 2006. A large collection of compact box C/D snoRNAs and their isoforms in *Euglena gracilis*: structural, functional and evolutionary insights. J. Mol. Biol. **357:**1548–1565.
33. **Schattner, P., W. A. Decatur, C. A. Davis, M. Ares, Jr., M. J. Fournier, and T. M. Lowe.** 2004. Genome-wide searching for pseudouridylation guide snoRNAs: analysis of the *Saccharomyces cerevisiae* genome. Nucleic Acids Res. **32:**4281–4296.
34. **Stuart, K. D., A. Schnaufer, N. L. Ernst, and A. K. Panigrahi.** 2005. Complex management: RNA editing in trypanosomes. Trends Biochem. Sci. **30:**97–105.
35. **Szewczak, L. B., J. S. Gabrielsen, S. J. Degregorio, S. A. Strobel, and J. A. Steitz.** 2005. Molecular basis for RNA kink-turn recognition by the h15.5K small RNP protein. RNA **11:**1407–1419.
36. **Tollervey, D., and T. Kiss.** 1997. Function and synthesis of small nucleolar RNAs. Curr. Opin. Cell Biol. **9:**337–342.
37. **Torchet, C., G. Badis, F. Devaux, G. Costanzo, M. Werner, and A. Jacquier.** 2005. The complete set of H/ACA snoRNAs that guide rRNA pseudouridylations in *Saccharomyces cerevisiae*. RNA **11:**928–938.
38. **Uliel, S., X. H. Liang, R. Unger, and S. Michaeli.** 2004. Small nucleolar RNAs that guide modification in trypanosomatids: repertoire, targets, genome organisation, and unique functions. Int. J. Parasitol. **34:**445–454.
39. **Vitali, P., H. Royo, H. Seitz, J. P. Bachellerie, A. Hüttenhofer, and J. Cavaillé.** 2003. Identification of 13 novel human modification guide RNAs. Nucleic Acids Res. **31:**6543–6551.
40. **Watkins, N. J., A. Dickmanns, and R. Lührmann.** 2002. Conserved stem II of the box C/D motif is essential for nucleolar localization and is required, along with the 15.5K protein, for the hierarchical assembly of the box C/D snoRNP. Mol. Cell. Biol. **22:**8342–8352.
41. **Watkins, N. J., R. D. Leverette, L. Xia, M. T. Andrews, and E. S. Maxwell.** 1996. Elements essential for processing intronic U14 snoRNA are located at the termini of the mature snoRNA sequence and include conserved nucleotide boxes C and D. RNA **2:**118–133.
42. **Weinstein, L. B., and J. A. Steitz.** 1999. Guided tours: from precursor snoRNA to functional snoRNP. Curr. Opin. Cell Biol. **11:**378–384.
43. **White, T. C., G. Rudenko, and P. Borst.** 1986. Three small RNAs within the 10 kb trypanosome rRNA transcription unit are analogous to domain VII of other eukaryotic 28S rRNAs. Nucleic Acids Res. **14:**9471–9489.
44. **Xia, L., N. J. Watkins, and E. S. Maxwell.** 1997. Identification of specific nucleotide sequences and structural elements required for intronic U14 snoRNA processing. RNA **3:**17–26.
45. **Xu, Y., L. Liu, C. Lopez-Estraño, and S. Michaeli.** 2001. Expression studies on clustered trypanosomatid box C/D small nucleolar RNAs. J. Biol. Chem. **276:**14289–14298.
46. **Yang, C. Y., H. Zhou, J. Luo, and L. H. Qu.** 2005. Identification of 20 snoRNA-like RNAs from the primitive eukaryote, *Giardia lamblia*. Biochem. Biophys. Res. Commun. **328:**1224–1231.