Cl gene cluster encoding several small nucleolar RNAs: a comparison amongst trypanosomatids

Paola Nocua, Carolina Gómez, Claudia Cuervo, Concepción Puerta/+

Laboratorio de Parasitología Molecular, Departamento de Microbiología, Facultad de Ciencias, Pontificia Universidad Javeriana, Carrera 7 43-82, Lab. 113, Bogotá, Colombia

*Small nucleolar RNAs (snoRNAs) are small non-coding RNAs that modify RNA molecules such as rRNA and snRNA by guiding 2′-*O*-ribose methylation (C/D box snoRNA family) and pseudouridylation reactions (H/ACA snoRNA family). H/ACA snoRNAs are also involved in* trans*-splicing in trypanosomatids. The aims of this work were to characterise the Cl gene cluster that encodes several snoRNAs in* Trypanosoma rangeli *and compare it with clusters from* Trypanosoma cruzi, Trypanosoma brucei, Leishmania major, Leishmania infantum, Leishmania braziliensis *and* Leptomonas collosoma. *The* T. rangeli *Cl gene cluster is an 801 base pair (bp) repeat sequence that encodes three C/D (Cl1, Cl2 and Cl4) and three H/ACA (Cl3, Cl5 and Cl6) snoRNAs. In contrast to* T. brucei, *the Cl3 and Cl5 homologues have not been annotated in the* Leishmania *or* T. cruzi *genome projects (http//:www.genedb.org). Of note, snoRNA transcribed regions have a high degree of sequence identity among all species and share gene synteny. Collectively, these findings suggest that the Cl cluster could constitute an interesting target for therapeutic (gene silencing) or diagnostic intervention strategies (PCR-derived tools).*

Key words: C/D snoRNA - H/ACA snoRNA - rRNA - *Trypanosoma rangeli* - trypanosomatid

Small nucleolar RNAs (snoRNAs) are a defined population of non-protein coding RNAs that guide post-transcriptional modifications of ribosomal RNA (rRNA) and some spliceosomal small nuclear RNAs (snRNAs) that are crucial for appropriate RNA folding, as well as RNA-RNA and RNA-protein interactions (Weinstein & Steitz 1999). These 60-300 nucleotide (nt)-long RNAs exist as snoRNA-protein complexes called snoRNPs. Most snoRNAs can be divided into two classes on the basis of their function and the presence of conserved sequence motifs; one contains two such motifs, box C (5' -RUGAUGA-3') and box D (5' -CUGA-3') along with two more degenerate internal copies of these elements, C' and D' boxes (Kiss-Laszlo et al. 1998); the other group contains the H (ANANNA) and ACA motifs (Maxwell & Fournier 1995). A few snoRNAs in each family are involved in pre-rRNA processing, but most are associated with nucleotide modification. Box C/D snoRNAs guide 2'-*O*-ribose methylation (Samarsky et al. 1998), while H/ACA snoRNAs guide pseudouridine formation (Balakin et al. 1996). The snoRNAs interact directly with their RNA targets through base complementarity and methylation or uridine isomerisation reactions are catalysed by core snoRNP proteins. Each box C/D snoRNP contains a single snoRNA harbouring the C and D elements required for snoRNA nucleolar localization, accumulation, maturation and protein association (Samarsky et al. 1998). Box C/D snoRNAs have one or two regions of sequence complementarity to specific region(s) in the rRNA (Bachellerie et al. 1995).

These 10-20 nt guide sequences reside immediately upstream of box D or D' and form duplexes with the target RNA, directing methylation of the fifth nucleotide (+5) that is paired upstream of the 5' end of box D or D'; this is known as the "+5 rule" (Decatur & Fournier 2003). The C/D snoRNAs of trypanosomatids such as *Leptomonas collosoma* (Levitan et al. 1998, Xu et al. 2001, Liang et al. 2004, Uliel et al. 2004), *Trypanosoma brucei* (Roberts et al. 1998, Dunbar et al. 2000a, b, Liang et al. 2005), *Leishmania tarentolae* (Roberts et al. 1998), *Leishmania major* (Uliel et al. 2004, Liang et al. 2007), *Trypanosoma cruzi* (Roberts et al. 1998, Uliel et al. 2004) and *Trypanosoma rangeli* (Morales et al. 2002) exhibit the same general structure and follow the +5 rule as well. Some trypanosomatid C/D snoRNAs are also able to guide methylation at two different sites in the same rRNA molecule or even in two different rRNAs, acting as double guides (Morales et al. 2002).

Box H/ACA snoRNAs have two conserved elements: the H domain (located in a hinge region joining two functionally similar hairpin domains) and the ACA triplet located 3 nt upstream from the 3' end. Two short rRNA recognition motifs within the snoRNA pair with rRNA sequences flanking the target uridine, which is always 14-16 nt upstream of the snoRNA H or ACA box (Ganot et al. 1997). The two internal loops within the H/ACA snoRNA usually contain complementary rRNA domains (Ni et al. 1997). These snoRNAs form a single hairpin in trypanosomatids ending in the AGA triplet instead of ACA (Liang et al. 2001, 2002, 2004, 2007, Barth et al. 2005).

Arranged differently from mammalians, most of trypanosomatid snoRNAs are clustered in tandem repeats and carry a mix of both snoRNA families (Dunbar et

Financial support: Vicerrectoría Académica, PUJ (1707)
+ Corresponding author: cpuerta@javeriana.edu.co
P Nocua and C Gómez contributed equally to this work.
Received 8 October 2008
Accepted 3 March 2009

al. 2000a, Morales et al. 2002, Liang et al. 2007). This genomic organisation resembles that of plants in which snoRNAs are also clustered and transcribed as polycistronic snoRNA precursors (Brown et al. 2003).

T. rangeli is a hemoflagellate protozoan parasite that, in contrast to *T. cruzi* (the etiologic agent of Chagas disease), is considered to be non-pathogenic to mammalian hosts (D'Alessandro & Saravia 1999). Despite its lack of human pathogenicity, *T. rangeli* is a serious concern for human Chagas disease epidemiology and diagnosis (Guhl & Vallejo 2003). Recent studies have classified *T. rangeli* into KP1(+) and KP1(-) strains, which are related to transmission-vector ability (Vallejo et al. 2003, 2007).

The Cl gene cluster (coding for several snoRNAs) was characterised in this work in a representative strain from each *T. rangeli* group and compared to those from *T. cruzi*, *T. brucei*, *L. major*, *Leishmania infantum*, *Leishmania braziliensis* and *L. collosoma*.

MATERIALS AND METHODS

Parasites - Epimastigotes from the *T. rangeli* KP1(+) H14 strain (MHOM/Hond/H14) (Acosta et al. 1991) were used in this study. They were characterised by using S35/S36/KP1L PCR (Vallejo et al. 2002). Parasites were grown at 26ºC in modified LIT medium supplemented with 15% (v/v) heat-inactivated foetal bovine serum.

PCR amplification, cloning and nucleotide sequence - PCR was performed using TrF1 (5' - CGC CCC GTC TTG CCC TGT-3') and TrR2 (5' - CGC AGC AAG GAC AGG AGG GA-3') primers, which, based on the *T. rangeli* KP1(-) C23 (MAOT/CO/82/C23) (Zuñiga et al. 1997) strain, amplify a 620 bp fragment exclusively in *T. rangeli*. A 25 µL reaction contained 100 ng purified genomic DNA, 1X reaction buffer (10 mM Tris-HCl pH 9.0, 50 mM KCl and 0.1% Triton X-100), 1.25 units of Expand High Fidelity enzyme (Roche, Branford, USA), 1.5 mM $MgCl_2$, 0.2 mM of each deoxynucleoside triphosphate and 20 pmol of each primer. Reactions were carried out in an MJ Research PTC-100 DNA thermal cycler using the following method: 94ºC /5 min, 15 cycles of 95ºC/30 s, 63ºC/1 min, 72ºC/30 s and 20 cycles of 95ºC/30 s, 60ºC/1 min, 72ºC/30 s and a final incubation of 72ºC for 5 min. Amplified fragments were purified from an agarose gel using a GFX Gel Band Purification kit (Amersham Biosciences) and cloned into the pGEM-T Easy plasmid (Promega). Both cloned fragment strands were sequenced by the Sanger method (Sanger et al. 1977) in a 373 Automatic DNA sequencer (Pharmacia LKB) using universal and specific primers.

Sequence analysis - Sequence of an 801 bp fragment from the *T. rangeli* KP1(-) C23 strain was retrieved from GenBank (accession AY028385) and included for comparative analyses. The *T. rangeli* KP1(+) H14 strain sequence produced in this study is available from the GenBank database (accession EF100612). Homology searches were performed in GenBank and parasite genome project databases (http://www.genedb.org) using the BLAST program (Stephen et al. 1997); pair-wise and multiple sequence alignments were carried out using LALIGN (Pearson 1990) and ClustalW (Thompson et al. 1994) software, respectively.

RESULTS

Analysing the Cl gene cluster from T. rangeli - BLASTN comparative analysis of the previously reported 801 bp fragment coding for snoRNA-Cl1 of the *T. rangeli* C23 strain (Morales et al. 2002) and *T. brucei* GeneDB version 4.0 revealed the presence of additional genes coding for five snoRNAs. This gene cluster repeat (named Cl) encoded six snoRNAs, three C/D (Cl1, Cl2 and Cl4) and three H/ACA (Cl3, Cl5 and Cl6), organised from the 5' end from Cl2, Cl3, Cl4, Cl5, Cl6 to Cl1. All C/D snoRNAs exhibited the consensus C, D, C' and D' boxes and regions having sequence complementarity to the methylation site on rRNA (Fig.1). Like other trypanosomatids, Cl gene cluster H/ACA snoRNAs formed

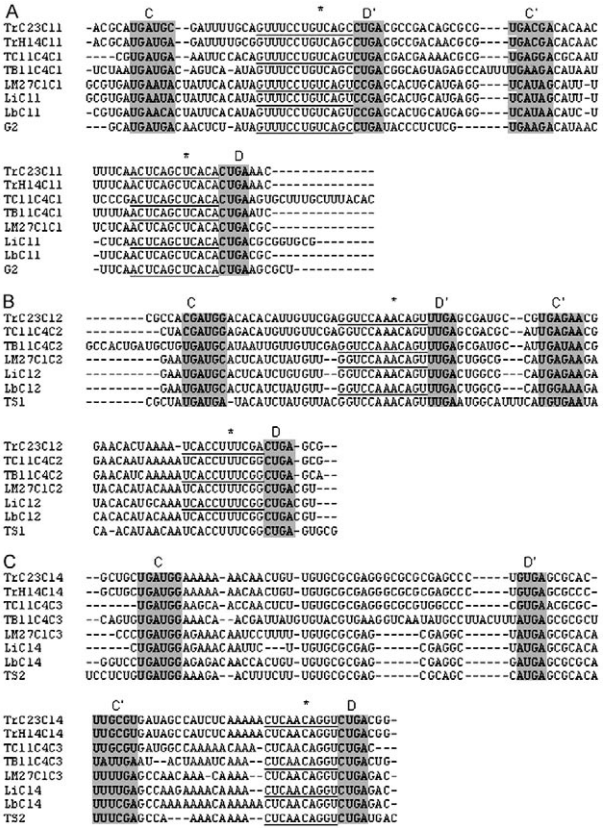

Fig. 1: CLUSTAL W multiple sequence alignment of C/D Cl-small nucleolar RNAs (snoRNAs) from different trypanosomatid species. A: snoRNA Cl1; B: snoRNA Cl2; C: snoRNA Cl4. The C, C', D and D' boxes are shaded, the rRNA complementarity regions are underlined and the gaps are shown as short lines. Asterisk denotes the methylation site. The species designations are as follows: TrC23: *Trypanosoma rangeli* C23 (GenBanK accession AY028385); TrH14: *T. rangeli* H14 (GenBanK accession EF100612); TC11: *Trypanosoma cruzi* (GeneDB contig 4406, Tc00.1047053487475.10, Tc00.1047053487475.20 and Tc00.1047053487475.30), TB11: *T. brucei* (GeneDB Tb11_snoRNA_0005, Tb11_snoRNA_0004 and Tb11_snoRNA_0002); LM27: *Leishmania major* (GeneDB LmjF27.snoRNA.0004, LmjF27.snoRNA.0001 and LmjF27.snoRNA.0002); Li: *Leishmania infantum* (Chromosome 27, accession AM502245) and Lb: *Leishmania brazilensis* (Chromosome 27, GenBank accession AM494964). G2, TS1 and TS2 sequences are from *Leptomonas collosoma* (GenBank accession AF331656).

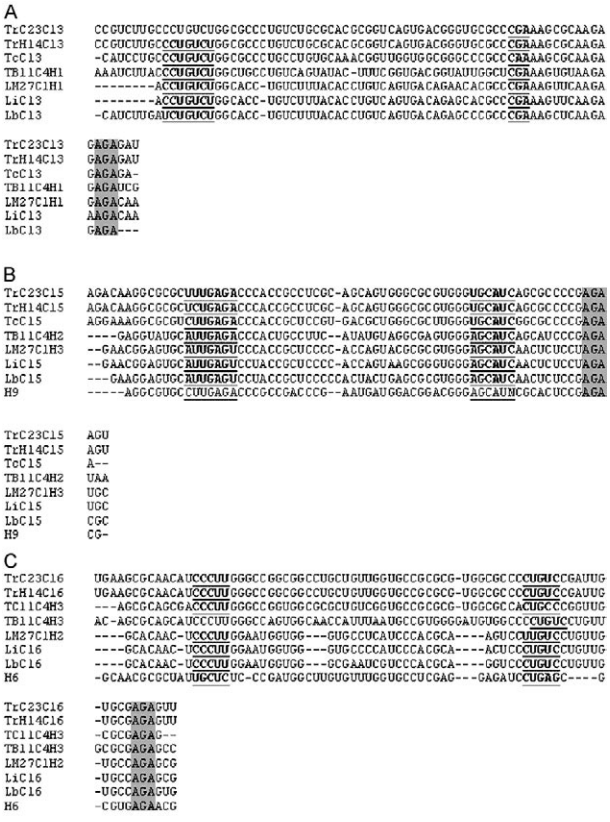

Fig. 2: multiple sequence alignment by CLUSTAL W of H/ACA Cl-small nucleolar RNAs (snoRNAs) from different trypanosomatid species. A: snoRNA Cl3; B: snoRNA Cl5; C: sno RNA Cl6. The AGA boxes are shaded, the rRNA complementarity regions are underlined and the gaps are shown as short lines. The species designation are as follows: TrC23: *Trypanosoma rangeli* C23 (GenBank accession AY028385); TrH14 *T. rangeli* H14 (GenBank accession EF100612); TC11: *Trypanosoma cruzi* (GeneDB contig 4406, Tc00.1047053487475.40); TB11: *Trypanosoma brucei* (GeneDB Tb11_snoRNA_0003, Tb11_snoRNA_0001 and Tb11_snoRNA_0006); LM27: *Leishmania major* (GeneDB LmjF27.snoRNA.0012, LmjF27.snoRNA.0003 and LmjF27.snoRNA.0011); Li: *Leishmania infantum* (Chromosome 27, accession AM502245); Lb: *Leishmania brazilensis* (Chromosome 27, GenBank accession AM494964). H6 and H9 sequences are from *Leptomonas collosoma* (Liang et al. 2004).

single hairpins ending in the AGA motif (Fig.2). Amplified sequence from the Cl gene cluster in the *T. rangeli* H14 strain (corresponding to nts 163-782 of the C23 strain) revealed 96.5% identity between the sequences of both strain, exhibiting two transversions, five transitions and two insertion-deletions (Table I).

Comparing the T. rangeli Cl gene cluster sequence with its homologue in T. cruzi - Comparative analysis of the *T. rangeli* C23 sequence with the *T. cruzi* genome revealed 81-88% identity between Cl1 and TC11C4C1, Cl2 and TC11C4C2, Cl4 and TC11C4C3 and Cl6 and TC11C4H3 (Table II). Since the current *T. cruzi* genome assembly and annotation version is fragmented and redundant, all contigs containing the Cl gene cluster were aligned with *T. rangeli* sequences. The analysis revealed

TABLE I

Comparison of Cl gene cluster sequences between KP1(+) H14 and KP1(-) C23 strains of *Trypanosoma rangeli*

Mutation	Position H14/C23	Change H14/C23	Affected region		
			TR	NTR	snoRNA
Transversion	176/338	C/A	x		Cl4
Insertion/deletion	236/398	▲/C		x	Cl4-Cl5
Transition	261/424	C/T	x		Cl5
Transition	474/637	C/T		x	Cl6-Cl1
Transversion	488/651	A/C	x		Cl1
Transition	497/660	G/A	x		Cl1
Transition	523/686	A/G	x		Cl1
Transition	604/753	C/T		x	Cl1-3' end
Insertion- deletion	581/743	△/▲		x	Cl1-3' end

NTR: non-transcribed region; snoRNAs: small nucleolar RNAs; TR: transcribed region; ▲: deletion; △: sequence CCCCCCCTCTTTTT.

the presence of Cl3 and Cl5 homologues in *T. cruzi* (84.9% and 81.2% identity, respectively).

It was also found that the gene order within the Cl cluster was conserved in both trypanosomes. In fact, transcribed regions were highly conserved between both species whereas non-transcribed regions differed in both size and sequence (Table III). The genomic sequence contained 103 contigs that include Cl-snoRNAs (*T. cruzi* GeneDB, version 4.0). Some of them, like contig 7066, carried other genes such as those coding for phosphatidylinositol kinase (Tc00.1047053506719.10), which corresponds to orthologues Tb11.47.0002 and LmjF27.0890 in *T. brucei* and *L. major*, respectively.

Comparing the T. rangeli Cl gene cluster sequence with its homologue in other trypanosomatids - Homologues for all Cl genes were detected in *T. brucei*, *L. major, L. infantum* and *L. braziliensis,* having 61.6-85.4% identity with *T. brucei* and 58.4-79.2% with the *Leishmania* species (Table II). All Cl-snoRNAs of *L. collosoma* were found to share 62.5-78.8% identity with those *T. rangeli*, except for Cl3 (Table II). Multiple alignments of all trypanosomatid snoRNAs revealed conservation of both sequence and position for the characteristic motifs of each snoRNA family (Figs 1, 2). Table IV shows the rRNA modifications carried out by Cl-snoRNA homologues from trypanosomatids, yeasts, plants and humans. Cl gene cluster organisation and genome location was studied in the trypanosomatid genome projects. This cluster in *T. brucei* was repeated five times, consisting of 32 snoRNA genes that maintained the same *T. rangeli* snoRNA gene order and were located on the antisense strand of chromosome 11 in position 456973-461605 (*T. brucei* GeneDB, version 4.0).

This cluster was flanked upstream by a gene encoding the isovaleryl-CoA dehydrogenase protein (Tb11.55.0026) and downstream by two ESAG genes (Tb11.55.0027 and Tb11.55.0028). It is worth noting that a phosphatidylinosi-

TABLE II

Comparison of Cl-small nucleolar RNAs (snoRNAs) from *Trypanosoma rangeli* C23 KP1(-) strain with their homologue sequences in different trypanosomatids

	Trypanosomatid homologue snoRNAs			
T. rangeli snoRNAs	Name	Specie	Length nt	Identity %
Cl2	TC11Cs4C2	*T. cruzi*	87	88.8
(88 nt)	TB11Cs4C2	*T. brucei*	97	85.4
	LM27Cs1C2	*L. major*	84	72.3
	LiCl2	*L. infantum*	84	72.3
	LbCl2	*L. braziliensis*	84	68.3
	TS1	*L. collosoma*	84	78.8
Cl3	TrH14Cl3	*T. rangeli* H14	75	100
(75 nt)	TcCl3	*T. cruzi*	73	84.9
	TB11Cs4H1	*T. brucei*	73	72.9
	LM27Cs1H1	*L. major*	66	77.3
	LiCl3	*L. infantum*	66	77.3
	LbCl3	*L. braziliensis*	70	79.2
Cl4	TrC23Cl4	*T. rangeli* H14	99	99
(99 nt)	TC11Cs4C3	*T. cruzi*	90	83.8
	TB11Cs4C3	*T. brucei*	101	61.6
	LM27Cs1C3	*L. major*	87	71.3
	LiCl4	*L. infantum*	85	71.3
	LbCl4	*L. braziliensis*	91	72.2
	TS2	*L. collosoma*	90	72.3
Cl5	TrH14Cl5	*T. rangeli* H14	71	98.6
(71 nt)	TcCl5	*T. cruzi*	69	81.2
	TB11Cs4H2	*T. brucei*	66	76.2
	LM27Cs1H3	*L. major*	69	74.2
	LiCl5	*L. infantum*	69	71.2
	LbCl5	*L. braziliensis*	69	74.6
	H9	*L. collosoma*	64	73.8
Cl6	TrH14Cl6	*T. rangeli* H14	77	100
(77 nt)	TC11Cs4H3	*T. cruzi*	72	85.7
	TB11Cs4H3	*T. brucei*	78	74.3
	LM27Cs1H2	*L. major*	66	64.8
	LiCl6	*L. infantum*	66	58.4
	LbCl6	*L. brazilensis*	66	66.7
	H6	*L. collosoma*	70	62.5
Cl1	TrH14Cl1	*T. rangeli* H14	86	96.5
(86 nt)	TC11Cs4C1	*T. cruzi*	97	81
	TB11Cs4C1	*T. brucei*	90	74.4
	LM27Cs1C1	*L. major*	88	70.6
	LiCl1	*L. infantum*	87	71.6
	LbCl1	*L. brazilensis*	86	74.6
	G2	*L. collosoma*	84	78.8

tol kinase gene (Tb11.47.0002) was located proximally on the sense strand (position 423851-428194).

The Cl gene cluster in *L. major* is located on both strands of chromosome 27. Two clusters of Cl2, Cl4, Cl5 and Cl1 are located in the sense strand at positions 369619- 370773, preceded upstream by a phosphatidylinositol kinase (LmjF27.0890) and a hypothetical conserved protein (LmjF27.0900) and followed downstream by another hypothetical conserved protein (LmjF27.0910). Fifty three snoRNAs are organised in the antisense strand into two arrays (Cl1, Cl3 and Cl6) and (Cl1, Cl2, Cl4 and Cl5), which are intercalated and repeat five and nine times, respectively. They are located at positions 375179-383161, flanked upstream by a hypothetical conserved protein (LmjF27.0920) and the isovaleryl-CoA dehydrogenase protein (LmjF27.0930).

Cl gene cluster snoRNAs have not been annotated yet in the *L. infantum* and *L. braziliensis* genome projects. However, BLASTN analysis has revealed their presence on chromosome 27 from both species. Forty six snoRNAs, organised similarly to those from the antisense strand of *L. major*, were located in the antisense strand in *L. infantum* at position 310106-316427 (*L. infantum* GeneDB, version 3.0).

They were flanked upstream by the isovaleryl-CoA dehydrogenase protein (LinJ27_V3.0790) and a hypothetical conserved protein (LinJ27_V3.0780) and followed downstream by another hypothetical conserved protein (LinJ27_V3.0770). A phosphatidylinositol kinase gene (LinJ27_V3.0750) was located proximally on the sense strand (position 295857-301364).

In the case of *L. braziliensis*, there were 22 snoRNAs organised into two arrays (Cl2, Cl4, Cl5 and Cl1) and (Cl3, Cl6 and a truncated Cl1), which are intercalated and repeat four and two times, respectively. This cluster was located in the sense strand in position 380881-384367 (*L. braziliensis* GeneDB, version 2.0) flanked upstream by a phosphatidylinositol kinase protein (LbrM27_V2.0970) and a hypothetical conserved protein (LbrM27_V2.0980) and followed downstream by another hypothetical conserved protein (LbrM27_V2.0990). An isovaleryl-CoA dehydrogenase protein (LbrM27_V2.1010) was located proximally in the antisense strand (positions 393327-394559). Interestingly, two snoRNAs (Cl2 and Cl1) were found 4276 nt downstream of the Cl gene cluster in position 388642-388849. It is worth noting that the copy number of each Cl-snoRNA varied within and among species, being more abundant in *L. major* than in the other species (Table V).

DISCUSSION

Non-protein coding RNAs (ncRNA) play critical roles in different processes affecting protein synthesis. snoRNAs (a type of ncRNA) are involved in RNA modifications leading to correct RNA folding and RNA-RNA and RNA-protein interactions (Zemann et al. 2006). It is known that snoRNAs from trypanosomatids are involved not only in rRNA processing and modifications but also in snRNA modifications thereby affecting protein synthesis and *trans*-splicing (Barth et al. 2005, 2008). In this study, the Cl gene cluster encoding several snoRNAs was characterised in *T. rangeli* and compared to those from *T. cruzi*, *T. brucei*, *L. major*, *L. infantum*, *L. braziliensis* and *L. collosoma*.

In accordance with their evolutionary origin, trypanosomatid Cl-snoRNAs present particular features shared with those from euglenids: H/ACA snoRNAs consist of a single hairpin (Russell et al. 2004), boxes C' and D' from C/D snoRNAs can be easily distin-

TABLE III

Intergenic regions comparison from Cl gene cluster from C23 KP1(-) *Trypanosoma rangeli* strain with its homologue sequences of trypanosomes

	Intergenic regions									
	Cl2-Cl3		Cl3-Cl4		Cl4-Cl5		Cl5-Cl6		Cl6-Cl1	
Trypanosome	bp	%	bp	%	bp	%	bp	%	bp	%
T. rangeli C23	61	-	39	-	30	-	12	-	71	-
T. rangeli H14	NA	NA	39	100	29	96.7	12	100	71	98.6
T. cruzi	70	54.9	58	51.7	58	43.1	19	36.8	106	48.1
T. brucei	109	27.5	70	22.9	92	16.3	50	14	37	25.4

bp: length of the intergenic region; NA: not available; %: percentage of identity.

TABLE IV

rRNA modifications guided by Cl-small nucleolar RNAs (snoRNAs) homologues

Trypanosoma brucei		*Leishmania major*	Yeasts		Plant		Humans	
snoRNA	Site	Site	snoRNA	Site	snoRNA	Site	snoRNA	Site
Cl1	Am1326 LSU 3′ - PTC on helix 91	Am1371			AtsnoR18	Am2924		
	Am1338 LSU 3′ - PTC on helix 90	Am1383	SnR71	Am2943	AtU29	Am2936	U29	Am4493
Cl2	Um1080 LSU 5′ on helix 27	Um847						
	Am1091 LSU 5′ on helix 37	Am858	SnR39/59	Am805	AtU51	Am814	U51/U32a	Am1511
Cl3	Ψ1357 LSU 3′ - PTC on helix 93	Ψ1402			AtsnoR53	Um2400		
Cl4	Um611 LSU 3′ - PTC on helix 74	Gm654						
Cl5	Ψ1907 SSU on helix 34	Ψ1841						
	Ψ61 SSU on helices 18-19							
Cl6	Ψ566 LSU 3′ on helix 72	Ψ610						

LSU3′: large-subunit rRNA 3’ half; LSU5′: large-subunit rRNA 5’ half; m: methylation; PTC: peptydil-transferase active site; SSU: small-subunit rRNA; Ψ: pseudouridylation. Adapted from Liang et al. (2005, 2007).

guished in spite of exhibiting some variations (Russell et al. 2006) and functional isoforms allow the presence of some plasticity in the 5’ half of transcribed regions (Liang et al. 2005, Russell et al. 2006).

Consistent with previous reports showing that snoRNAs in trypanosomatids are encoded by clusters of arrayed tandem genes, the Cl gene cluster repeat contains six snoRNA genes exhibiting an intercalated array of C/D and H/ACA snoRNAs.

Given the fact that the parasite requires large amounts of mature rRNA molecules, the arrangement of several alternating snoRNA genes might provide a solution to a lack of transcriptional controls in these parasites.

The Cl gene cluster order, size and sequence is highly conserved between the KP1(+) and KP1(-) strains from *T. rangeli*, having few mutations that affect both the transcribed and non-transcribed regions. The biological significance of this finding needs to be addressed by

TABLE V

Copy number of Cl-small nucleolar RNAs (snoRNAs) from trypanosomatids

	Copy number (number of polymorphic or truncated sequences - identity %)					
Trypanosomatid	Cl1	Cl2	Cl3	Cl4	Cl5	Cl6
T. brucei	5 *	5 *	5 (3-98)	6 (1-100)[a]	6 (1-98)	5 *
L. major	17 (1-98)	12 (4-98)	5 *	11 *	11 (1-98)	5 *
L. infantum	13 (3-98)	10 *	3 *	9 *	9 (3-98)	4 (1-100)[b]
L. braziliensis	7 (3-90)[a]	5 (1-97)	2 *	4 *	4 *	2 *

a: truncated versions with the first half of the molecule; *b*: truncated version missing the firsts 20 nt. Asterisks means absence of polymorphic or truncated-sequences.

analysing more strains. Nevertheless, it is reasonable to assume that the observed changes, especially transversion, could affect the spatial structure and function of the RNA molecule.

We observed that transcribed regions and gene order of the Cl gene cluster were conserved among the three species of trypanosomes studied. However, Cl-snoRNA and intergenic spacer sequence identity was higher with *T. cruzi* homologues than with those of *T. brucei*; this finding agrees with previous reports revealing a closer *T. rangeli* and *T. cruzi* phylogenetic relationship than that between *T. rangeli* and *T. brucei* (Maia da Silva et al. 2004, Diez et al. 2005, Cuervo et al. 2006).

Orthologues from *T. cruzi* phosphatidylinositol kinase (Tc00.1047053506719.10), Tb11.47.0002 and LmjF27.0890, surround the Cl-snoRNA array in *T. brucei* and *L. major*. Consequently, it is possible that the *T. cruzi* Cl gene cluster is located in the gene synteny region similar to what has been reported for chromosomes 11 and 27 of *T. brucei* and *L. major*, respectively (El-Sayed et al. 2005).

Interestingly, Cl gene cluster members of leishmanias were found in different tandem arrays from those observed in trypanosomes. Taking into account the efficient expression of trypanosomatid snoRNAs, one could hypothesize that expression is influenced by extended stems formed in the extragenic flanking sequences of adjacent snoRNA molecules (Liang et al. 2007); these array order differences could have important biological consequences.

T. rangeli intergenic regions of the Cl gene cluster range from 12-93 nt. This range is in accordance with the minimum 10 nt distance needed for the proper processing of each snoRNA (Xu et al. 2001, Liang et al. 2004). Indeed, Cl gene cluster C/D box snoRNA expression has been described in *L. collosoma* (Xu et al. 2001) and *T. brucei* (Barth et al. 2008). Although there is no specific expression data from Cl-H/ACA snoRNAs, other B2 cluster H/ACA snoRNAs have been detected in *L. collosoma* by Northern blot or primer extension analysis (Xu et al. 2001, Liang et al. 2004) suggesting the expression of these snoRNAs as well. An intense signal corresponding to approximately 90 nt has been shown in Northern blot assays in *T. rangeli* using the total Cl cluster as a probe, which seems to include all Cl-snoRNAs (Morales et al. 2002).

In spite of having the same copy number, C/D snoRNAs in *T. brucei* have different levels of expression, with Cl1 showing the greatest. However, snoRNAs Cl2 and Cl3 exhibited greater target methylation as compared to Cl1 (Barth et al. 2008). These results showed that other factors aside from copy number and expression level were influencing the modification guiding process, such as the presence of secondary structure at the modification site. The Cl-snoRNA copy numbers from leishmanias differ among each snoRNA according to its array. In contrast to C/D snoRNAs Cl1 and Cl2, H/ACA snoRNAs Cl3 and Cl6 are less represented in the genome. This dosage effect might be a compensatory mechanism driven by a need for Cl-snoRNA differential expression. Likewise, in archaea and plants some Cl-snoRNAs can function as double guides, guiding two modifications at proximal (Cl1 and Cl2) and distal (Cl5) sites on the same rRNA molecule (Omer et al. 2000, Brown et al. 2003).

It was especially interesting that all Cl-snoRNA sequences were highly conserved amongst all trypanosomatids, ranging from 61.6-88.8% identity for C/D and from 58.4-85.7% for H/ACA snoRNAs. This finding suggests it is important for these parasites to maintain the rRNA modifications performed by Cl-snoRNAs.

Methylation and pseudouridylation mapping carried out by Cl-snoRNA homologues in *T. brucei* and *L. major* has revealed that four modifications (Am1326, Am1338, Um611 and Ψ1357) are located within the LSU rRNA peptidyl transferase active site. Although blocking individual rRNA modifications has not had any effect, King et al. (2003) have shown that depleting multiple modifications in the LSU reaction centre region have had synergistic, negative effects on growth.

It has been reported recently that position U611 within the LSU is 66.2% hypermethylated in bloodstream versus procyclic forms; this may help the parasite to adapt to a higher vertebrate host temperature (Barth et al. 2008).

It is of the upmost importance that homologues that carry out Am1338 (Cl1) and Am1091 (Cl2) LSU modifications have been found in yeast and even humans, implying an importance of these modifications for rRNA structure and function. On the other hand, other Cl-snoRNA modifications such as Um1080 (Cl2), Um611 (Cl4), Ψ 1907, Ψ 61 (Cl5), and Ψ566 (Cl6) seem to be trypanosomatid-specific.

The finding that trypanosomatid Cl-snoRNAs share important traits such as sequence, function and specificity renders this cluster a good target candidate for medically important interventions of these parasites. For instance, a therapeutic gene expression silencing approach could be addressed. In fact, Liang et al. (2003) have reported the silencing of *T. brucei* TBC4 C/D snoRNA. On the other hand, species-specific differences in the non-transcribed regions of Cl-snoRNAs could be useful for

developing PCR-based diagnostic tools. Indeed, Morales et al. (2002) have developed a PCR test specific for *T. rangeli* detection, which does not amplify the DNA of any *T. cruzi* groups (Pavia et al. 2007).

ACKNOWLEDGEMENT

To Dr. RS Nicholls and M Montilla, from the Parasitology Laboratory, Instituto Nacional de Salud, for kindly providing the *Trypanosoma rangeli* H14 strain.

REFERENCES

Acosta I, Romanha AJ, Cosenza H, Krettli AU 1991. Trypanosomatids isolates from Honduras: differentiation between *Trypanosoma cruzi* and *T. rangeli*. *Am J Trop Med Hyg 44:* 676-683.

Bachellerie JP, Michot B, Nicoloso M, Balakin A, Ni J, Fournier MJ 1995. Antisense snoRNAs: a family of nucleolar RNAs with long complementarities to rRNA. *Trends Biochem Sci 20:* 261-264.

Balakin AG, Smith L, Fournier MJ 1996. The RNA world of the nucleolus: two major families of small nucleolar RNAs defined by different box element with related functions. *Cell 86:* 823-834.

Barth S, Hury A, Liang XH, Michaeli S 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.

Barth S, Shalem B, Hury A, Tkacz ID, Liang XH, Uliel S, Myslyuk I, Doniger T, Salmon-Divon M, Unger R, Michaeli S 2008. Elucidating the role of C/D snoRNA in rRNA processing and modification in *Trypanosoma brucei*. *Eukaryot Cell 7*: 86-101.

Brown JW, Echeverria M, Qu LH 2003. Plant snoRNAs: functional evolution and new modes of gene expression. *Trends Plant Sci 8*: 42-49.

Cuervo C, López MC, Puerta C 2006. The *Trypanosoma rangeli* histone H2A gene sequence serves as a differential marker for KP1 strains. *Infect Genet Evol 6*: 401-409.

D'Alessandro A, Saravia N 1999. *Trypanosoma rangeli*. In HM Gilles, *Protozoal Diseases*, Oxford University Press Inc, New York, p. 398-412.

Decatur WA, Fournier MJ 2003. RNA-guided nucleotide modification of ribosomal and other RNAs. *J Biol Chem 278*: 695-698.

Diez H, Thomas MC, Urueña CP, Santander SP, Cuervo CL, López MC, Puerta CJ 2005. Molecular characterization of the kinetoplastid membrane protein-11 genes from the parasite *Trypanosoma rangeli*. *Parasitology 130*: 643-651.

Dunbar DA, Chen AA, Wormsley S, Baserga SJ 2000a. The genes for small nucleolar RNAs in *Trypanosoma brucei* are organized in clusters and are transcribed as a polycistronic RNA. *Nucleic Acids Res 28:* 2855-2861.

Dunbar DA, Wormsley S, Lowe TM, Baserga SJ 2000b. Fibrallarin-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.

El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, Caler E, Renauld H, Worthey EA, Hertz-Fowler C, Ghedin E, Peacock C, Bartholomeu DC, Haas BJ, Tran AN, Wortman JR, Alsmark UC, Angiuoli S, Anupama A, Badger J, Bringaud F, Cadag E, Carlton JM, Cerqueira GC, Creasy T, Delcher AL, Djikeng A, Embley TM, Hauser C, Ivens AC, Kummerfeld SK, Pereira-Leal JB, Nilsson D, Peterson J, Salzberg SL, Shallom J, Silva JC, Sundaram J, Westenberger S, White O, Melville SE, Donelson JE, Andersson B, Stuart KD, Hall N 2005. Comparative genomics of trypanosomatid parasitic protozoa. *Science 309*: 404-409.

Ganot P, Bortolin ML, Kiss T 1997. Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell 89:* 799-809.

Guhl F, Vallejo GA 2003. *Trypanosoma* (*Herpetosoma*) *rangeli* Tejera, 1920 - an updated review. *Mem Inst Oswaldo Cruz 98:* 435-442.

King TH, Liu B, McCully RR, Fournier MJ 2003. Ribosome structure and activity are altered in cells lacking snoRNPs that form pseudouridines in the peptidyl transferase center. *Mol Cell Biol 11:* 425-435.

Kiss-Laszlo Z, Henry Y, Kiss T 1998. Sequence and structural element of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *EMBO J 17*: 797-807.

Levitan A, Xu Y, Ben-Dov C, Ben-Shlomo H, Zhang Y, Michaeli S 1998. Characterization of a novel trypanosomatid small nucleolar RNA. *Nucleic Acids Res 26:* 1775-1783.

Liang XH, Hury A, Hoze E, Uliel S, Myslyuk I, Apatoff A, Unger R, Michaeli S 2007. 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. *Eukaryot Cell 6:* 361-377.

Liang XH, Liu L, Michaeli S 2001. Identification of the first trypanosome H/ACA RNA that guides pseudouridine formation on rRNA. *J Biol Chem 276:* 40313-40318.

Liang XH, Liu Q, Michaeli S 2003. Small nucleolar RNA interference induced by antisense or double-stranded RNA in trypanosomatids. *Proc Natl Acad Sci USA 100:* 7521-7526.

Liang XH, Ochaion A, Xu YX, Liu Q, Michaeli S 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.

Liang XH, Uliel S, Hury A, Barth S, Doniger T, Unger R, Michaeli S 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.

Liang XH, Xu YX, Michaeli S 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.

Maia da Silva FM, Noyes H, Campaner M, Junqueira AC, Coura JR, Añez N, Shaw JJ, Stevens JR, Teixeira MM 2004. Phylogeny, taxonomy and grouping of *Trypanosoma rangeli* isolates from man, triatomines and sylvatic mammals from widespread geographical origin based on SSU and ITS ribosomal sequences. *Parasitology 129:* 549-561.

Maxwell ES, Fournier MJ 1995. The small nucleolar RNAs. *Annu Rev Biochem 64:* 897-934.

Morales L, Romero I, Diez H, Del Portillo P, Montilla M, Nicholls S, Puerta C 2002. Characterization of a candidate *Trypanosoma rangeli* small nucleolar RNA gene and its application in a PCR-based parasite detection. *Exp Parasitol 102:* 72-80.

Ni J, Tien AL, Fournier MJ 1997. Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. *Cell 89*: 565-573.

Omer AD, Lowe TM, Russell AG, Ebhardt H, Eddy SR, Dennis PP 2000. Homologues of small nucleolar RNAs in Archaea. *Science 288:* 517-522.

Pavia PX, Vallejo GA, Montilla M, Nicholls RS, Puerta CJ 2007. Detection of *Trypanosoma cruzi* and *Trypanosoma rangeli* infection in triatomine vectors by amplification of the histone H2A/SIRE and the sno-RNA-C11 genes. *Rev Inst Med Trop São Paulo* 49: 23-30.

Pearson WR 1990. Rapid and sensitive sequence comparison with FASTP and FASTA. *Meth Enzymol 183:* 63-98.

Roberts TG, Sturm NR, Yee BK, Yu MC, Hartshorne T, Agabian N, Campbell DA 1998. Three small nucleolar RNAs identified from the spliced leader-associated RNA locus in kinetoplastid protozoans. *Mol Cell Biol 18:* 4409-4417.

Russell AG, Murray NS, Gray MW 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.

Russell AG, Schnare MN, Gray MW 2004. Pseudouridine-guide RNAs and other Cbf5p-associated RNAs in *Euglena gracilis. RNA 10:* 1034-1046.

Samarsky DA, Fournier MJ, Singer RH, Bertrand E 1998. The snoRNA box C/D motif directs nucleolar targeting and also couples snoRNA synthesis and localization. *The EMBO J 17:* 3747-3757.

Sanger F, Nicklen S, Coulson A 1977. DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci USA 74:* 5463-5546.

Stephen A, Madden T, Schaffer A, Zhang Z, Miller W, Lipman D 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search program. *Nucleic Acids Res 25:* 3389-3402.

Thompson JD, Higgins DG, Gibson TJ 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673-4680.

Uliel S, Liang X, Unger R, Michaeli S 2004. Small nucleolar RNAs that guide modification in trypanosomatids: repertoire, targets, genome organization and unique functions. *Int J Parasitol 34:* 445-454.

Vallejo GA, Guhl F, Carranza JC, Lozano LE, Sánchez JL, Jaramillo JC, Gualtero D, Castañeda N, Silva JC, Steindel M 2002. kDNA markers define two major *Trypanosoma rangeli* lineages in Latin-America. *Acta Trop 81:* 77-82.

Vallejo GA, Guhl F, Carranza JC, Moreno J, Triana O, Grisard EC 2003. Parity between kinetoplast DNA and mini-exon gene sequences supports either clonal evolution or speciation in *Trypanosoma rangeli* strains isolated from *Rhodnius colombiensis, R. pallescens* and *R. prolixus* in Colombia. *Infect Genet Evol 3:* 39-45.

Vallejo GA, Guhl F, Carranza JC, Triana O, Pérez G, Ortiz PA, Marín DH, Villa LM, Suárez J, Sánchez IP, Pulido X, Rodríguez IB, Lozano LE, Urrea DA, Rivera FA, Cuba-Cuba C, Clavijo JA 2007. Interacción tripanosoma-vector-vertebrado y su relación con la sistemática y la epidemiología de la tripanosomiasis americana. *Biomédica 27*: 110-118.

Weinstein LB, Steitz JA 1999. Guided tours: from precursor snoRNA to functional snoRNP. *Curr Opin Cell Biol 11:* 378-384.

Xu Y, Liu L, Lopez-Estraño C, Michaeli S 2001. Expression studies on clustered trypanosomatid box C/D small nucleolar RNAs. *J Biol Chem 276:* 14289-14298.

Zemann A, Anja op de Bekke, Kiefmann M, Brosius J, Schmitz J 2006. Evolution of small nucleolar RNAs in nematodes. *Nucleic Acids Res 34:* 2676-2685.

Zuñiga C, Palau T, Penin P, Gamallo C, de Diego JA 1997. Characterization of a *Trypanosoma rangeli* strain of Colombian origin. *Mem Inst Oswaldo Cruz 92:* 523-530.