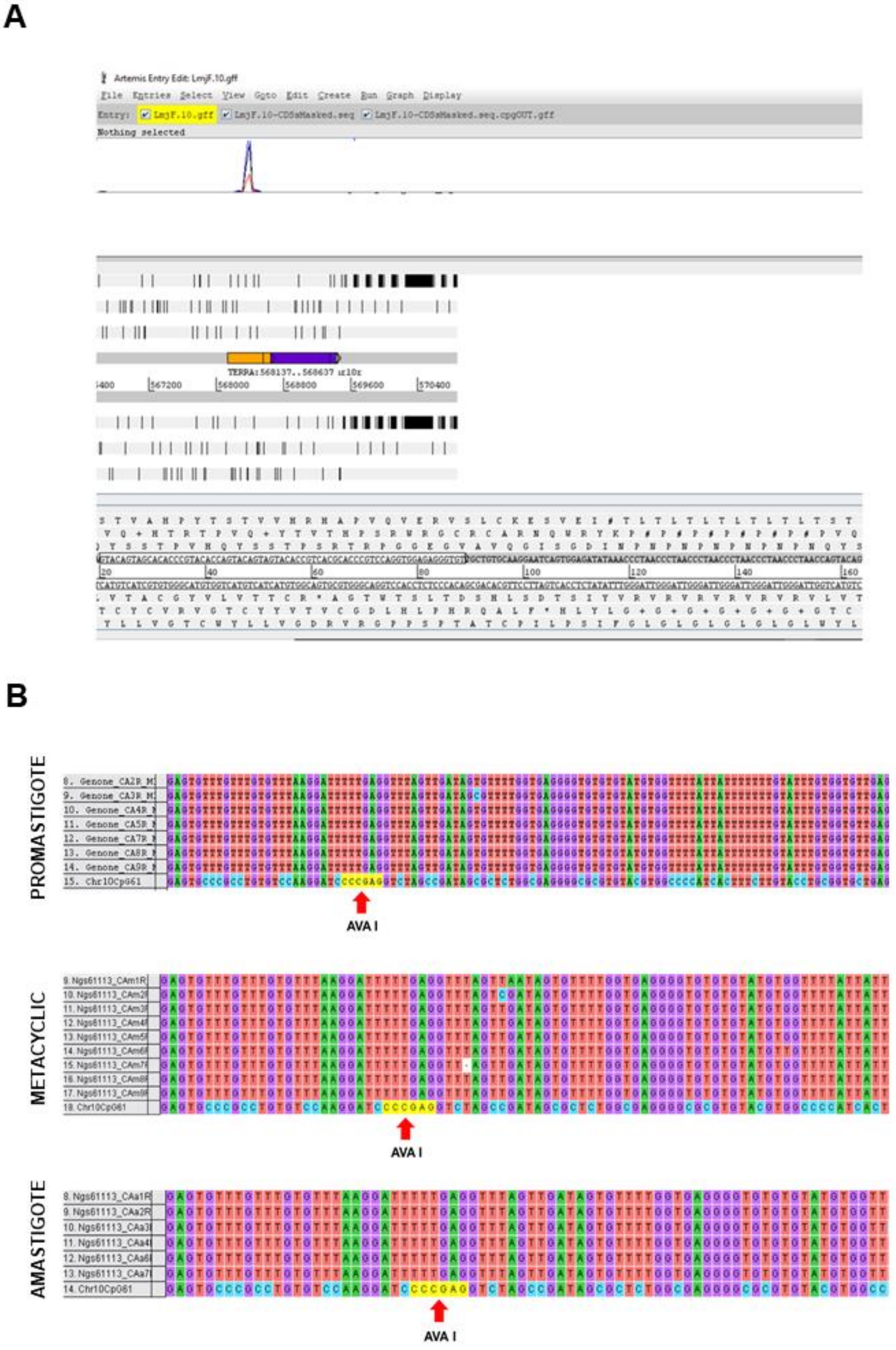
Article:

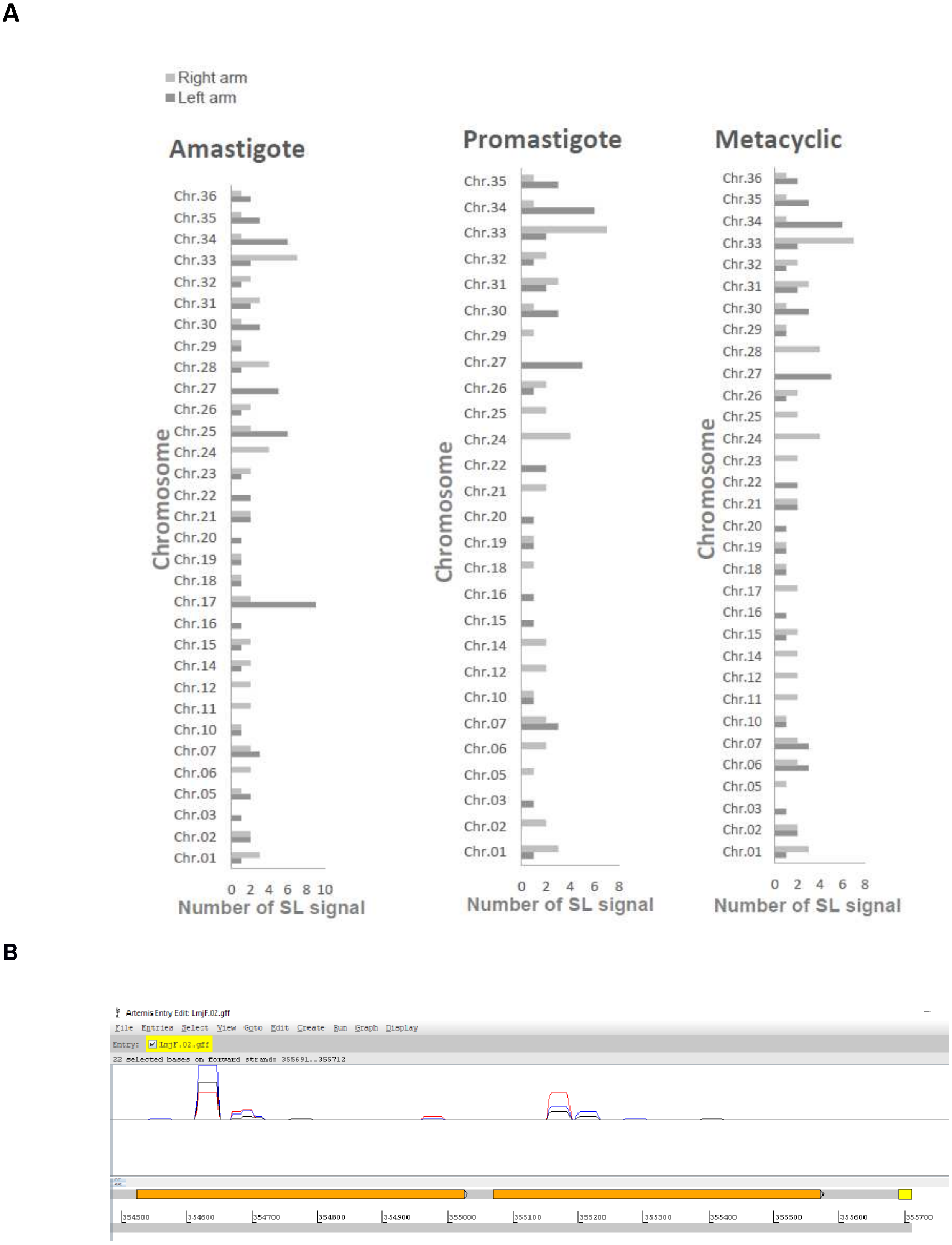
Morea, EGO, Vasconcelos, EJR orcid.org/0000-0001-5130-6622, Alves, CDS et al. (5 more authors) (2021) Exploring TERRA during Leishmania major developmental cycle and continuous in vitro passages. International Journal of Biological Macromolecules, 174. pp. 573-586. ISSN 0141-8130

https://doi.org/10.1016/j.ijbiomac.2021.01.192

Supplementary Material

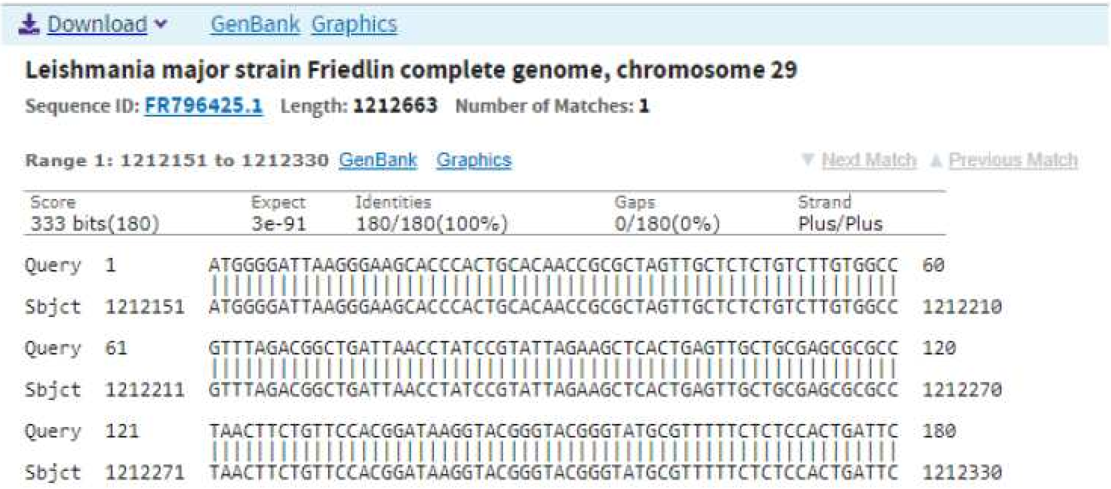


Suppl. Fig. 1. *In sílico* identification and sequencing of the subtelomeric putative CpG island at *L. major* Chr10R. A) The terminus of Chr10R is represented in the Artemis genome browser. At the top of the figure, picks indicate the position of SL signals in amastigotes (black), promastigote (red), and metacyclic (blue). The putative telomeric transcript identified in the subtelomeric region of Chr10R is represented as an orange block. A white arrow at the orange block tip indicates that telomeric transcription originates at the SL signals localized at the subtelomeric region towards the telomeric repeat (black squares). The putative CpG island location is marked in purple. The numbers indicate the position of nt at Chr10R. B) Genomic DNA from promastigotes, metacyclics, and amastigotes were sodium bisulfite-treated to search for m^{5}C. Clones containing the putative CpG island were obtained, automated sequence, and aligned with the ClustalW multiple sequences alignment tool through the MEGA software. Nucleotides in each sequenced clone are marked in colors: Guanine (purple), Adenine (green), Thymine (red), and Cytosine (blue). The position of an *Ava*I restriction site within the putative CpG island is depicted in yellow. At the bottom of each panel is the partial sequence of the putative CpG island.



22 selected bases on forward strand: 355691..355712

Suppl. Figure 2. The number of subtelomeric SL signals in each chromosome end termini present in all three *L. major* life stages. A) Subtelomeric SL signals at the R arms (light gray bars) and the L arms (gray bars). Amastigote, promastigote, and metacyclic. These data were collected from each of the three independent SL-RNA-Seq libraries using Artemis. B) The terminus of Chr2R is represented in the Artemis genome browser. At the top of the figure, picks indicate the position of SL signals in amastigotes (black), promastigote (red), and metacyclic (blue). Putative telomeric transcripts (orange blocks) were identified in the subtelomeric region of Chr2R. White arrows at the tip of the orange blocks indicate that telomeric transcription originates at the SL signals localized at the subtelomeric region towards the telomeric repeat (yellow square). The numbers indicate the position of nt at Chr2R.



Leishmania major strain Friedlin complete genome, chromosome 29

Sequence ID: FR796425.1 Length: 1212663 Number of Matches: 1

Range 1: 1212151 to 1212330 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
333 bits(180)	3e-91	180/180(100%)	0/180(0%)	Plus/Plus

```
Query  1        ATGGGGATTAAGGGAAGCACCCACTGCACAACCGCGCTAGTTGCTCTCTGTCTTGTGGCC  60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1212151  ATGGGGATTAAGGGAAGCACCCACTGCACAACCGCGCTAGTTGCTCTCTGTCTTGTGGCC  1212210

Query  61       GTTTAGACGGCTGATTAACCTATCCGTATTAGAAGCTCACTGAGTTGCTGCGAGCGCGCC  120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1212211  GTTTAGACGGCTGATTAACCTATCCGTATTAGAAGCTCACTGAGTTGCTGCGAGCGCGCC  1212270

Query  121      TAACTTCTGTTCCACGGATAAGGTACGGGTACGGGTATGCGTTTTTCTCTCCACTGATTC  180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1212271  TAACTTCTGTTCCACGGATAAGGTACGGGTACGGGTATGCGTTTTTCTCTCCACTGATTC  1212330
```

Suppl. Figure 3. Nucleotide sequence alignment of TERRA Chr29R. cDNA obtained from *L. major* promastigotes was used to amplify TERRA Chr29R following cloning into a PCR 4-TOPO vector. BLASTn was used for sequence alignment using the NCBI public database and confirmed that the TERRA transcript was originated from the end terminus of *L. major* Chr29R.

Suppl. Table 1. Primers used for Northern blot, RT-PCR and cloning

Primers	Sequence
SubChro.09l-F	5´ TCTCTTGCCAGCCTCTCTTC 3´
SubChro.09l-R	5´ GTTTTTATTGCCGAGGCTTG 3´
SubChro.20l-F	5´ GCGCCGTGTATTTTCAGTCT 3´
SubChro.20l-R	5´ ACTTCGCCCATCATATCAGC 3
SubChro.04l-F	5´ CACTCTCGACGCCGACCTA 3´
SubChro.04l-R	5´ GCTTTGCCAGCTGCTTGTG 3´
SubChro.10r-F	5´ CACACCGTGAACGCAAGGAAAC 3´
SubChro.10r-R	5´ TCCGTGGTGGTGCGTTCT 3´
SubChro. 29r-F	5´ ATGGGGATTAAGGGAAGCAC 3´
SubChro. 29r-R	5´ GAATCAGTGGAGAGAAAAACGCATA-3´
OligodT	5´ TTTTTTTTTTTTTTTTTTTTTTTTT-3´
DIG-TelC	5´ CCCTAACCCTAACCCTAACCCTAACCCTAA 3´
DIG-TelG	5´TTAGGGTTAGGGTTAGGGTTAGGGTTAGGG 3´
RPN8 F	5´ ATGAACCGCCGCAAGCT3´
RPN8 R	5´ GGCGCGACGACGATCTTTGATT3´

Suppl. Table 2. Normalized read counts obtained from the chromosome end termini in each of the three independent SL-RNA-Seq libraries in RPKM (Reads Per Kilobase per Million mapped reads)

		SL Signal Localization (nt)	**Amastigote**	**Promastigote**	**Metacyclic**
			AH045	AH043	AH044
Chr1	L	514..1014	16,87	7,843	37,916
	R	267996..268496	3,233	14,833	6,896
Chr2	L	660..1160	8,234	0,941	6,641
	R	355070..355570	2,209	8,47	4,284
Chr3	L	39..539	134,957	43,919	157,232
	R	382633..383133	0,402	0	0
Chr4	L	1935..2435	0,803	0	0
	R	471243..471747	1,004	0	0
Chr5	L	2425..2923	2,008	0	0,643
	R	464743..465243	40,367	12,548	65,763
Chr6	L	207..707	0,201	0,314	1,499
	R	515564..516064	97,603	45,488	111,39
Chr7	L	786..1286	13,054	3,451	23,778
	R	594822..595322	19,681	4,078	29,775
Chr8	L	3073..3573	0	0	0,214
	R	574378..574878	0	0	0
Chr9	L	1746..2246	0	0	0,214
	R	572009..572509	0	0	0
Chr10	L	1116..1616	7,431	2,196	15,637
	R	568137..568637	30,526	16,627	41,986
Chr11	L	957..1457	0,402	0	0
	R	581910..582410	1,205	0,941	1,714
Chr12	L	198..698	0,402	0	0
	R	673787..674287	3,414	1,569	5,355
Chr13	L	645..1145	0,201	0	0,428
	R	652654..653154	0,201	0	0
Chr14	L	870..1370	5,623	0,314	0,643
	R	620778..621278	21,288	9,411	33,845
Chr15	L	357..857	18,878	6,588	22,492
	R	627492..627992	1,205	0	1,285
Chr16	L	400..900	127,326	32,312	241,631
	R	714012..714512	0,201	0,314	0,857
Chr17	L	903..1403	1,205	0,314	0,643
	R	682964..683464	4,418	0,941	4,07
Chr18	L	735..1235	1,607	0,314	3,642
	R	738632..739132	13,054	4,078	26,991
Chr19	L	699..1299	44,699	11,245	34,821
	R	701479..701979	47,797	18,509	67,477
Chr20	L	1479..1979	79,93	41,096	90,826
	R	740629..741129	0	0	0
Chr21	L	1374..1874	6,627	0,627	7,283
	R	771448..771948	3,816	2,51	4,713
Chr22	L	1317..1817	31,731	14,744	36,202
	R	715815..716315	0,602	0	1,071
Chr23	L	757..1257	2,209	0	0
	R	770300..770800	4,017	0	4,284

Chr24	L	1920..2420	0,402	0	0
	R	840073..840573	3,816	1,255	3,427
Chr25	L	1818..2318	1,205	0,627	0,857
	R	910268..910768	24,702	3,765	32,989
Chr26	L	720..1220	59,044	16,627	71,761
	R	1089941..1090441	5,222	1,255	16,923
Chr27	L	1176..1676	3,012	1,255	3,427
	R	1129388..1129888	0,602	0,627	0,428
Chr28	L	453..953	2,41	0,627	0,214
	R	1158246..1158746	4,217	0,627	2,571
Chr29	L	288..792	1,793	0	1,275
	R	1210488..1210988	89,369	22,273	124,457
Chr30	L	1473..1973	9,238	5,961	26,134
	R	1397822..1398322	14,661	8,784	24,42
Chr31	L	1557..2057	7,431	3,137	12,21
	R	1483014..1483514	22,292	17,568	42,2
Chr32	L	1575..2075	7,029	2,823	12,424
	R	1603890..1604390	40,166	12,548	67,691
Chr33	L	486..986	26,71	9,725	51,197
	R	1582297..1582797	4,418	1,882	4,498
Chr34	L	390..890	5,422	1,255	7,497
	R	1864900..1865400	21,087	6,902	29,99
Chr35	L	306..806	7,431	7,215	13,281
	R	2089541..2090041	8,033	3,451	16,066
Chr36	L	1260..1760	7,632	0,941	8,997
	R	2679877..2680377	8,033	0,627	10,496

Suppl. Table 3. Comparison among the normalized read counts obtained from each chromosome end terminus versus data obtained from the three independent SL-RNA-Seq libraries in RPKM (Reads Per Kilobase per Million mapped reads)

Name	Total	RPKM
AMA, META, PRO	0	
AMA, PRO	1	Chr35L_7
AMA, META	6	Chr21L_7; Chr27L_3; Chr17R_4; Chr15R_1; Chr23R_4; Chr33R_4
META, PRO	0	

Suppl. Table 4. Structural features found at the chromosome ends of *L. major* Amastigotes (R and L arms). Presence (+) or absence (-) of SL and CSB.

Type	L arm (L)	R arm (R)	CSB	SL
1	9; 12; 13; 24	3; 4; 22	-	-
2	1; 2; 3; 5; 7; 15; 16; 21; 27; 29; 35	1; 2; 11; 17; 18; 24; 31	-	+
3	10; 20;14;17;18;19;22;23; 25; 26; 28; 30; 31; 32; 33; 34; 36	5; 6; 7; 10; 12; 14; 15; 19; 21; 23; 25; 26; 28; 29; 30; 32; 33; 34; 35; 36	+	+
4	4; 6; 11	9; 13; 16; 20; 27	+	-

Suppl. Table 5. Structural features found at the chromosome ends of *L. major* Promastigotes (R and L arms). Presence (+) or absence (-) of SL and CSB.

Type	L arm (L)	R arm (R)	CSB	SL
1	2; 5; 9; 12; 13; 21; 24; 29;	3; 4; 11; 17; 22;	-	-
2	1; 3; 7; 15; 16; 27; 35	1; 2; 18; 24; 31	-	+
3	10; 19; 20; 22; 26; 30; 31; 32; 33; 34	5; 6; 7; 10; 12; 14; 19; 21; 25; 26; 29; 30; 32; 33; 34; 35	+	+
4	4; 6; 11; 14; 17; 18; 23; 25; 28; 36	9; 13; 15; 16; 20; 23; 27; 28; 36	+	-

Suppl. Table 6. Structural features found at the chromosome ends of *L. major* Metacyclics (R and L arms). Presence (+) or absence (-) of SL and CSB.

Type	**L arm (L)**	**R arm (R)**	**CSB**	**SL**
1	5; 9; 12; 13; 24;	3; 4; 22	-	-
2	1; 2; 3; 7; 15; 16; 21; 27; 29; 35	1; 2; 11; 17; 18; 24; 31;	-	+
3	6; 10; 18; 19; 20; 22; 26; 30; 31; 32; 33; 34; 36	5; 6; 7; 10; 12; 14; 15; 19; 21; 23; 25; 26; 28; 29; 30; 32;33; 34; 35; 36	+	+
4	4; 11; 14; 17; 23; 25; 28;	9; 13; 16; 20; 27;	+	-

Suppl. Table 7. qPCR from RNA: DNA hybrids immunoprecipitation. The table shows the observed values of *CT*. RNase H treatment was used as a control. Undetermined means that it was not possible to detect amplification.

	Chr10R	**Chr20L**	**Chr29R**
Amastigote input	13,882	15,372	16,91
Amastigote + RNase H	Undetermined	Undetermined	Undetermined
Amastigote -RNase H	27,381	27,146	30,823
Promastigote P1 input	13,676	15,282	17,825
Promastigote P1+RNase H	Undetermined	Undetermined	Undetermined
Promastigote P1 - RNase H	Undetermined	Undetermined	Undetermined
Promastigote P3 Input	14,594	15,946	17,406
Promastigote P3 + RNase H	Undetermined	Undetermined	Undetermined
Promastigote P3 – RNase H	Undetermined	Undetermined	33,126
Promastigote P6 Input	11,248	12,219	15,001
Promastigote P6 + RNase H	Undetermined	Undetermined	Undetermined
Promastigote P6 – RNase H	Undetermined	Undetermined	31,931
Metacyclic M1 Input	18,689	19,452	20,757
Metacyclic M1 + RNase H	Undetermined	Undetermined	Undetermined
Metacyclic M1 – RNase H	Undetermined	Undetermined	Undetermined
Metacyclic M3 Input	16,936	17,37	19,186
Metacyclic M3 + RNase H	Undetermined	Undetermined	Undetermined
Metacyclic M3 – RNase H	Undetermined	Undetermined	34,369
Metacyclic M6 Input	16,749	17,614	19,471
Metacyclic M6 + RNase H	Undetermined	Undetermined	Undetermined
Metacyclic M6_- RNase H	Undetermined	Undetermined	32,909