

# Supplementary Material

# Telomeric repeat-binding factor homologues in *Entamoeba histolytica*: new clues for telomeric sequences search

Francisco Javier Rendón-Gandarilla<sup>1,+</sup>, Víctor Álvarez-Hernández<sup>1</sup>, Elizabeth J. Castañeda-Ortiz<sup>1</sup>, Helios Cárdenas-Hernández<sup>1</sup>, Rosa Elena Cárdenas-Guerra<sup>1</sup>, Jesús Valdés<sup>2</sup>, Abigail Betanzos<sup>3,4</sup>, Bibiana Chávez-Munguía<sup>4</sup>, Anel Lagunes-Guillen<sup>4</sup>, Esther Orozco<sup>4</sup>, Lilia López-Canovas<sup>1</sup> and Elisa Azuara-Liceaga<sup>1\*</sup>.

\* Correspondence: Corresponding Author: elisa.azuara@uacm.edu.mx

## **Supplementary Data**

#### **1** Supplementary Tables

Table S1. Paired primers used for assessing the transcript levels of *ehtrf-like* genes in *E. histolytica* 

Gene	Accession	Sequences of paired primers	Amplicon			
	(5'-3')					
Ehtrf-1	EHI_001090	F: ACCCTCAGTTTGCAACCCTT	235			
		R: AGGACTTTGTTCCATTTTTACTTC				
Ehtrf-2	EHI_001110	F: AAAAGAGCATTAGCAACGGTTAT R: TTTCTGCATTTGGATCTGAACAA	240			
Ehtrf-3	EHI_148140	F: GCGCGTTCTCAAATACAACC	145			
		R: GTTCAGAGGCTGCAGGAAGT				
Internal control						
40s rps2	EHI_020280	F: ATTCGGAAATAGAAGAGGAGG	105			

R: ACTAATCTT CCAAGCTTGGT

## Table S2. Telomeric probes designed for EMSA assays.

DNA	Oligonucleotides	Characteristics	Reference
probe			
HsTel	Forward 5'-GCTCTAGGGTTAGGGTTAGGGTTAGGGTTA-3' Reverse 5'-AATCCCAATCCCAATCCCAATCCCATCTAC-3'	Human telomeric sequence	Hwang et al., 2001
EhTel	Forward 5'-GCTCTATTAGTATTTTAGTATTTTAGTATTT-3' Reverse 5'-AATCATAAAATCATAAAATCATAAAATCATAACGAGATCTAC- 3'	The EhTEL contains one STR sequence present in the NK2 tRNA array	Tawai et al., 2008
mut-TEL	Forward 5'-GCTCTAGATGTGGGTAGCTAGCTCTGACTCAGG-3' Reverse 5'-ACCCGATCGATCGAGACTGAGTCCCGAGATCTAC-3'	Mutated telomeric sequence	Hwang et al., 2001
non-Rel	Forward 5 <sup>-</sup> -GCTCTAGATGTCGAGTCTGTGTGTGTACATTACGTGTG-3 <sup>-</sup> Reverse 5 <sup>-</sup> -AGCTGAGACACACATGTAAATGCACACCGAGATCTAC-3 <sup>-</sup>	Non-related sequence	Hwang et al., 2001

		~				5	1		1	
	Full-length (% identity)				MYB DBD (% identity)					
	EhTRF- like I	EhTRF- like II	EhTRF- like III	HsTFR1	HsTFR2	EhTRF- like I	EhTRF- like II	EhTRF- like III	HsTFR1	HsTFR2
EhTRF- like I	-	35.89	33.91	13.86	14.1	-	93.44	89.13	36.66	33.33
EhTRF- like II		-	63.94	13.89	13.15		-	80.32	36.66	33.33
EhTRF- like III			-	14.35	12.6			-	36.66	31.66
HsTFR1				-	22.8				_	51.66
HsTFR2					-					-

Table S3. Identity matrix of EhTRF-like of E. histolytica and H. sapiens TRF proteins

Percent of identity based on the full-length protein or the MYB DBD sequence pairwise sequence alignment.

## 2. Supplementary Figures

HsTRF1 HsTRF2 EhTRF-like I EhTRF-like II EhTRF-like III	0 MAEDVSSAAPSPRGCADGRDADPTEEQMAETERNDEEQFECQELLECQVQVGAPEEEEEEEEDAGLVAEAAAGMML MAAGAGTAGPASGPGVVRDPAASQPRKRPGREGGEGARRSDTMAGGGGSSDGSGRAAGRRASRSSGRARRGRHEPGLGGPAERGAGEARLEEAV-NRWVL MNNPQFATLQRIVSQEPLHPTQNIPIEVKMEQSPMIVSTDLETLCRKKTL MNCIEDKCIVAHHIDQLKQNRVDDVNVFDN	79 99 50 36 53
HsTRF1 HsTRF2 EhTRF-like I EhTRF-like II EhTRF-like III	DFLCLSICRAFRDGRSED DFLCLSICRAFRDGRSED PROTRDINGALLVRPLGKEHTVSRLL-RVMOCLSRIEAQFENDERITPLSAINVLEMIKTE EDICEVIMDQDIECGLFNQLIQ-NEYINPDELSQNYKKKTLKYALQIINGIDPNCPQVESLGKISECVIQLFQEDSKSEILKPLPSLIGLTGMITLF ISICEVILDSGINSDLFQCLN-QQIFVESDIPQSLHNKALATVINFFSSIEPFNIEIEVIDFKIISNVIQIYCSDSNAEILKTVPSLIGITGMISIF LSICEYLIDSGINSDLFQQVFMESDIPQSLHNKALSIAFVIQFLCVEPPCCEVENLPKVIANVIQIYCSDSNAEILKTVPSLIGIVGMISVF	164 184 147 133 150
HsTRF1 HsTRF2 EhTRF-like I EhTRF-like II EhTRF-like III	$\begin{array}{l} HDKLHEEIQNLIKIQAIAVCMENGNFKEAEEVFERIFGDPNSHMPFKSKLLMIISQKDTFHSFFQHFSYNHMMEKIKSYVNYVLOEKSSTFLMK FTUTEAVVESSKLVKEAAVITCIKNKEFEKASKILKHMSKDPTTQKLRNLLHIIRENLAHPVIQHFSYETFQQKMLRFLESHLDD-AEPYLLT YQYVSKKVQTQAMWDWLDLLPALS-QLPYLKTSKSCSEEAIDILSETEIPDNLTNPEEFLMRITSIHQIIARYPVIATLEHIKTQLEKYHQSLNPS HLYVSKKVQTQALMHULLPASSEDPYVASSKVCLSAIEALLKTPIKEGITDSNEVLQNIQEVYILSRYPLIATLEHIKLLPYKEICPSS HMFARSQIQPSELNRELNTLLPASSEQPYVASSKVCLSAIEALLKAPPKDETTDSFKYLSTIPDVQYILSRYPLIATIDHIKQLLYQYKESSIQS$	258 280 243 229 246
HsTRF1 HsTRF2 EhTRF-like I EhTRF-like II EhTRF-like III	O O AAAKVVESKRTRTIITSQDKPSGNDVEMETEANLDTRKSVSDKQSAVTE	306 380 264 278 295
HsTRF1 HsTRF2 EhTRF-like I EhTRF-like II EhTRF-like III	SEGTVSLLRSHKNLFLSKLOHGTQQQDLNKKERRVGTPQSTKKKKESRRATESRIPVSKSQPVT SAPADGEGGSBLQPKNKRMTISRLVLEEDSQSTEPSAGLNSSQEARSAPPSKPTVLNQPLPGEKNPKVPKGKNNSSNGVEEKETWVEEDELFQVQAAPDE DNKIDDL	371 480 295 350 370
HsTRF1 HsTRF2 EhTRF-like I EhTRF-like II EhTRF-like III	PEKHRARKROAWLWEEDKNLRSGVRKYGEGNWSKILLHYKENNRISVMLKDRWRTMKKLKLISSDSED	439 542 395 439 458
HsTRF1 HsTRF2 EhTRF-like I EhTRF-like II EhTRF-like III	439 542 GFKQLDLSQ 404 NY 441 N 459	

Figure S1. Alignment of complete amino acid sequences of TRF proteins from *H. sapiens* and *E. histolytica*. In the alignment, conserved hydrophobic residues involved in the dimer interphase formation such as leucines, isoleucines, or phenylalanines are shown with open circles and identical amino acids are highlighted in colors. The position of three  $\alpha$ -helices within the MYB DBD and conserved tryptophans or aromatic residues marked with black arrowheads are also showed.



Figure S2. EhTRF-like proteins and their putative TRFH domain. Secondary structure alignment of complete sequence of HsTRF1 and EhTRF-like proteins performed in PSIPRED. E: beta sheet, H: helix and C: random coil. The nine and three  $\alpha$ -helices of TRFH and MYB DBD domains, respectively are shown.

	MYB DBD	
	RARKKOAWLWBDDKNIKSEVRKIE – EENWSKILLH – WK – – F-NNKTSVMIKDRWRTM	
HeTRE2		TRF1/2
PITRE2	N I TKKOKWTVELESEWYKACYOKYC - ECHWAATSKN - YP E - YNB TAWM IKDRWEITM	110 1/2
BtTRF2	NTTINGOKWTYDESEWYKAGYOKYG-EGNWAATSKN-YPE-YNRTAVMIKDRWRTM	
DrTRF2	KKYTRKMWSVOESEWEKOGVŸRYGVCHWERTRSA-FPF-AGRTAVNEKORWRTM	
TrTRF2	SGORRRMWTEAETOKLIEGVRKFGAGNWSKIRAY-YSF-NDRTNVNLKDRWRTL	
AtTRF-like 3	GRKLHRAWTISEVEKLVEGVSKYGVGKWTEIKKLSFSPYTHRTTVDLKDKWRNL	Plants
AtTRF-like 6	RRKHHRAWTLSEIAKLVEGVSKYGAGKWSEIKKHLFSSHSYRTSVDLKDKWRNL	TRF-like
AtTBP1	QRRTRPFSVTEVEALVSAVEEVGTGRWRDVKLRSFENASHRTYVDLKDKWKTL	
AtTRP4	QRRTRRPFSVTEVEALVSAVEEVGTGRWRDVKLRSFENASHRTYVDLKDKWKTL	Plants
OsTBP1	QRRIR <mark>R</mark> PFTVAEV LLVEAVEHLGTGRWRDVKFRAFENVHH <mark>RT</mark> YVDLKDKWKTL	IBP
EhTRF-like I	IKKKKRRFTEEETQNLIEGVQQFGIGHWKLILNN-FKF-DDRSCVDLKDKWRNL	
EhTRF-like II	LKKKKRRFTEEETQNLIEGVQQFGIGHWKLILNN-FKF-DDRSCVDLKDKWRNL	
En I RF-like III	LKKKKRRFSEEETQNLIEGVQQFGIGHWKSILNA-YKF-DGRSCVDLKDKWRNI	
EdB0EC67	LKKKKRRFTEEETQNLIEGVQQFGIGHWKLILSE-YEFAPGRTSVDLKDKYRNL	
EdB0EC69	IKKKKRRFTEEETONLIEGVOOFGIGHWKLILNN-FKF-DDRSCVDLKDKWRNL	
EIN_079420	QKKKKRRFSEEETONLIDGVKQYGIGHWKNILGS-YKF-DGRSCVDLKDKWRNI	Entamoeba
EIN_023650	OKKKKRRFSEETONLIAGVEOFGVGHWKSILSA-YEF-DGRSCVDLKDKWRNI	TRF-like
EMO_106010		
EMO_103990		
EMO_113000	V K K K K K K S E E E TONI I NO VE E G - V GHWK SI I NA - YI F - DGKS C V D K A K WK NI	
EnK2G4T7	S KKNARKESE BEDUNILT DUVQUEG - VGHMASTINA - VK F- DGRSCVDD KDAWAAL	
EnH4C7	$\mathbf{L}$ KKKRP FTFFFFONITE GV00FG = TOHMKITTN - FK = -FF DORS CVDTKDKWPNI	
EnG4T7	$\mathbf{L} \mathbf{K} \mathbf{K} \mathbf{K} \mathbf{R} \mathbf{R} \mathbf{F} \mathbf{T} \mathbf{E} \mathbf{E} \mathbf{E} \mathbf{T} \mathbf{O} \mathbf{N} \mathbf{I} \mathbf{T} \mathbf{E} \mathbf{G} \mathbf{V} \mathbf{O} \mathbf{C} \mathbf{G} = \mathbf{I} \mathbf{G} \mathbf{H} \mathbf{K} \mathbf{K} \mathbf{I} \mathbf{I} \mathbf{I} \mathbf{N} \mathbf{N} - \mathbf{F} \mathbf{K} - \mathbf{E} - \mathbf{D} \mathbf{R} \mathbf{S} \mathbf{C} \mathbf{V} \mathbf{D} \mathbf{K} \mathbf{K} \mathbf{K} \mathbf{K} \mathbf{N} \mathbf{N} \mathbf{I}$	
NF0021980	PROK VEWSEDEVDAULACYKREGCNWAKILAT-VKKM-ENPVRDSLSURDKYRNL	
AcL8GQH5	GRERERWIEDEVANTLAGVKKEGVGYWKHIVEH-YDE-GGRUAVDIKDKYRNL	Naegieria/Acanthamoeba
AcL8GYK3	GRRKKRFWTDDEIELLLEGVRTHGLGCWAKILSE-YEFAPGRTSVDLKDKYRNL	I RF-like
LaTRF	KVRRHKFSPOEDAAILHGVARFGOMSGSFOYILHA-YRSV-WRAGRTALHLYDHWRGA	
tbTRF	GKOKRYKFSASEDEAIIKGLARETKGOORFOOLYYA-YRSV-WHPARTVSOLYDHWRGT	other protozoa
BbTRF2	QVRSRFRWAEYEVENIIKGVQEYGVGHWSEIIDN-YRF-KNRTNVNLKDKWRQL	TRF-like
TtTRF	SARKRRWTDDEIKWLIEGVAQFGHGNWAAILAA-FPF-TGRTNVNLKDKWRTL	
TvTBP	EKRQAIPWTQEEVDAIEDGVKKYGIGHWTLVYEL-HKDIFMKNERKSSDVGDKWKNM	
EHI_009820	KPKKSKKWDKEEILKLYTCIMKYCADFGLIELMF-DGRURKQIKAKFKAE	Non TRF
EHI_074810	ERKKNHVWSTBETYQFYQALKIFGTDFEKIQKRLGDHIRRDVSVKFKRE	

Telebox VxxKDxxR

Figure S3. Multiple alignment of the MYB DBD amino-acid sequence of several TRF proteins. ClustalW alignment of the MYB DBD belonging to TRF proteins from different organisms. Identical amino acid residues are black boxed. Conserved tryptophans and telobox residues are showed in asterisks and arrowheads, respectively. HsTRF1 (Homo sapiens, NP 059523.2), MmTRF1 (Mus musculus, NP 033378.1), PtTRF1 (Pan troglodytes, XP 016815057.1), XITRF1 (Xenopus laevis, DQ146827.1), HsTRF2 (H. sapiens, NP 005643.2), PtTRF2 (P. troglodytes, XP 016785565.1), BtTRF2 (Bos taurus, NP 001341392.1), BbTRF2 (Branchiostoma belcheri, XP 019634930.1), DrTRF2 (Danio rerio, BC096857.1), TrTRF2 (Thecamonas trahens, XP 013759355.1), TtTRF (Takifugu rubripes, XP 011616883.1), AtTRF-like-1 (Arabidopsis thaliana, NP 001330208.1), AtTRF-like 3 (A. thaliana, NP 001185023.1), AtTRF-like 6 (A. thaliana, NP 974133.1), AtTBP1 (A. thaliana, XP 015625433.1), AtTRP1 (A. thaliana, NP 196886.1), OsTBP1 (Oryza sativa, XP 015625433.1), EhTRF-like I (E. histolytica, XP 649878.1), EhTRF-like II (E. histolytica, XP 649880.1), EhTRF-like III (E. histolytica, XP 657528.1), EdBOEC67 (Entamoeba dispar, XP 001735901.1), EdB0ESZ2 *(E.* dispar, XP 001741195.1), EdB0EC69 (*E*. dispar. XP 001735903.1), Ein079420 (E. invadens, XP 004184366.1), Ein023650 (E. invadens, XP 004257447.1), EnH4C7 (*E*. XP 008857362.1), EnK2H4C7 nuttalli. *(E.* nuttalli. XP 008860337.1), EnK2G4T7(E. nuttalli, XP 008860339.1), NF0021980 (Naegleria fowleri), LaTRF2 (Leishmania amazonensis, ABU53006.1), TbTRF (Trypanosoma brucei, AAX86992.1), TvTBP (Trichomonas vaginalis, XP 001314728.1), EHI 009820 (E. histolytica, XP 001913661.1), and EHI 074810 (E. histolytica, XP 649576.1).



**Figure S4. Expression and purification of rEhTRF-like I from** *E. histolytica*. A) Coomassie brilliant blue-stained SDS-PAGE on 12% gel of recombinant EhTRF-like I. NI: Not Induced; I: Induced; SF: Soluble Fraction; E1-2: elutions from the purification by nickel affinity chromatography. B) WB assay with the anti-His antibody against the recombinant protein. C) WB assay with the anti-TRF-like I antibody in total extracts (TE) from trophozoites of *E. histolytica* or the rEhTRF-like I. The letter M and lines on the left side show the position of the molecular weight markers in kilodalton (kDa).