Available online at www.sciencedirect.com

SCIENCE DIRECT®

Gene 346 (2005) 153–161

GENE
SECTION
EVOLUTIONARY GENOMICS

www.elsevier.com/locate/gene

Telomere and subtelomere of *Trypanosoma cruzi* chromosomes are enriched in (pseudo)genes of retrotransposon hot spot and trans-sialidase-like gene families: the origins of *T. cruzi* telomeres☆

Dong Kim[a,1], Miguel Angel Chiurillo[b,c,1], Najib El-Sayed[d], Kristin Jones[d], Márcia R.M. Santos[a], Patricio E. Porcile[a], Bjorn Andersson[e], Peter Myler[f], José Franco da Silveira[a], José Luis Ramírez[c,g,*]

[a]*Departamento de Microbiologia, Imunologia e Parasitologia, Escola Paulista de Medicina, UNIFESP, Rua Botucatu, 862, CEP 04023-062, S. Paulo, Brazil*
[b]*Decanato de Medicina, Universidad Centro Occidental Lisandro Alvarado, Barquisimeto, Venezuela*
[c]*Centro de Biotecnologia, Instituto de Estudios Avanzados (IDEA), carretera Nacional Hoyo de la Puerta, Caracas 1080, Caracas, Venezuela*
[d]*Parasite Genomics, Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA*
[e]*Center for Genomics and Bioinformatics, Karolinska Institutet, SE-171 77 Stockholm, Sweden*
[f]*Seattle Biomedical Research Institute, 307 Westlake Av. N, Suite 500, Seattle, WA 98109, USA*
[g]*Laboratorio de Genetica Molecular, Instituto de Biologia Experimental Universidad Central de Venezuela, calle Suapure, Colinas de Bello Monte, Caracas 1041-A, Venezuela*

Received 16 May 2004; received in revised form 6 September 2004; accepted 14 October 2004
Available online 28 January 2005
Received by F.G. Alvarez-Valin

Abstract

Here, we sequenced two large telomeric regions obtained from the pathogen protozoan *Trypanosoma cruzi*. These sequences, together with in silico assembled contigs, allowed us to establish the general features of telomeres and subtelomeres of this parasite. Our findings can be summarized as follows: We confirmed the presence of two types of telomeric ends; subtelomeric regions appeared to be enriched in (pseudo)genes of RHS (retrotransposon hot spot), TS (trans-sialidase)-like proteins, and putative surface protein DGF-1 (dispersed gene family-1). Sequence analysis of the *ts*-like genes located at the telomeres suggested that *T. cruzi* chromosomal ends could have been the site for generation of new *gp85* variants, an important adhesin molecule involved in the invasion of mammalian cells by *T. cruzi*. Finally, a mechanism for generation of *T. cruzi* telomere by chromosome breakage and telomere healing is proposed.

Keywords: Pathogen protozoan; Chromosome end organization; Surface protein genes; Retrotransposon Hot Spot multigene family; Telomere healing

Abbreviations: aa, amino acid(s); ASP-2, amastigote surface protein 2; BAC, bacterial artificial chromosome; *dgf-1*, dispersed gene family-1; bp, base pair(s); EST, expressed sequence tag; GPI, glycosylphosphatidylinositol; GP85, surface glycoprotein of 85 kDa; kbp, kilobase pair(s); LTR, long terminal repeat; nt, nucleotide; ORF, open reading frame; RHS, retrotransposon hot spot; SIRE, short interspersed repetitive element; SAS, SIRE-associated sequence; Tctel, *Trypanosoma cruzi* telomeric sequence; TS, trans-sialidase; UTR, untranslated region; VATc, *Trypanosoma cruzi* telomeric sequence cloned by vector-adaptor strategy; VIPER, vestigial interposed retroelement.

☆ Sequences in this work were deposited in GenBank® with accession numbers: BAC D6C AY551440 and C6 AY552588.

* Corresponding author. Centro de Biotecnologia, Instituto de Estudios Avanzados (IDEA), carretera Nacional Hoyo de la Puerta, Caracas 1080, Caracas, Venezuela. Tel.: +58 212 962 1605; fax: +58 212 962 1120.
E-mail address: jramirez@reacciun.ve (J.L. Ramírez).

[1] These authors contributed equally to this work.

0378-1119/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2004.10.014

1. Introduction

Trypanosoma cruzi is a protozoan parasite causing Chagas disease, an incurable and debilitating illness affecting 16–18 millions of people in the Latin American region (WHO, 2002). Besides its medical importance, trypanosomatids are evolutionarily interesting since they belong to one of the earliest groups of mitochondria-containing eukaryotes, have a highly plastic genome, and have unusual gene organization (Zingales et al., 1997; Myler et al., 1999; Anderson et al., 1998). Several parasites such as *Plasmodium* and *Trypanosoma brucei* have developed sophisticated evasion mechanisms to adapt to the hostile environment posed by the host, such as exposing variable surface antigens to escape the immune system. Genes coding for surface antigens in these organisms are located at subtelomeric regions, and it has been speculated that this preferred location facilitates gene switching and expression, and the generation of new variants (Cano, 2001; Barry et al., 2003).

In previous works, we have described the basic elements of *T. cruzi* telomeres, and found that they were enriched in (pseudo)genes from the *ts* (trans-sialidase)-like family and sequences related to VIPER (vestigial interposed retroelement) (Chiurillo et al., 1999, 2002). Members of the *ts*-like gene family display great sequence diversity and encode many surface proteins related with cell invasion, virulence, and evasion from the host immune system (Weston et al., 1999; Frasch, 2000). We speculated that the preferred telomeric location of the *ts*-like family genes could be connected to the generation of variants via non-homologous recombination (Chiurillo et al., 1999, 2002).

Here, we sequenced and analyzed two large telomeric regions from *T. cruzi* that together with the information in *T. cruzi* Genome Project database, allowed us to draw a finer picture of the organization of telomeric and subtelomeric regions of this parasite. We also discuss the functional importance of this organization to explain the generation of genetic variability, and the origins of *T. cruzi* telomeres.

2. Materials and methods

2.1. Nucleotide (nt) sequencing of BAC and cosmid telomeric clones

The telomeric BAC (BAC D6C) and cosmid (C6) recombinants here studied were selected from *T. cruzi* (clone CL Brener) libraries constructed in pBeloBAC11 or cosmid Lawrist-7 by hybridization with an 18-mer telomeric probe $(5'\text{-CCCTAA-}3')_3$ (Chiurillo et al., 1999, 2002). Sheared DNA from the selected recombinants (1.6–2 kbp) was cloned into a modified pUC18 vector via *Bst*XI linkers. Sequences were assembled using the TIGR assembler and gaps were closed using a combination of BAC walking, directed PCR or transposon insertion. In BAC D6C end-DNA sequence readings confirmed the presence of the telomeric adaptor and *Bam*HI site used for cloning. Open reading frames (ORF) were assigned by NCBI programs: BLASTN, BLASTX, and ORF Finder.

2.2. In silico assembling of telomeric contigs

We searched TIGR *T. cruzi* WGS database, which contains 994,060 sequences and represents an approximately 21-fold coverage of *T. cruzi* haploid genome (ca. 43 Mb in clone CL Brener), for contigs containing the 189-bp telomeric junction, and then they were further extended towards the subtelomeric regions using DNASTAR MegAlign software. To validate in silico assembling, primers based on the contig sequences were used to amplify, by polymerase chain reaction, specific fragments on *T. cruzi* genomic DNA. The amplified fragments of the expected size were cloned in plasmid vector and sequenced.

3. Results

3.1. Sequence organization of T. cruzi telomeres

Assembling of shotgun sequences from recombinant BAC D6C and BLAST search analysis allowed the construction of the map presented in Fig. 1. A summary of the characteristics of the 29,248-bp contig in BAC D6C is as follows, from telomere to centromere (right to left).

(1) A telomere sequence type II (Chiurillo et al., 1999) whose basic elements are 66 copies of hexameric repeats 5′-TTAGGG-3′, followed by a 189-bp junction, a truncated sequence of 100 bp from the 5′-UTR (untranslated region) of a *gp85* gene (GenBank accession no. M64836) from group II of the *ts*-like family, the spacer between two *gp85* genes, a sequence of 530 bp without homology in GenBank database, and short interspersed repetitive element (SIRE)-associated sequence (SAS) SZ23 (Vazquez et al., 1999, 2000). In addition, we identified two sequences between nt 26,572 and 27,845 that share 84–86% of identity with the recently described L1Tc non-long terminal repeat (LTR) retrotransposon flanking sequences Seq3Tc (Olivares et al., 2000). These elements are part of (pseudo)genes of *T. cruzi rhs* family (Bringaud et al., 2002a). Translation of the subtelomeric *rhs*-related sequences showed the presence of three putative peptides of 115, 344, and 178 aa (amino acids), and the derived aa sequences shared 28–30% identity with the *T. brucei* RHS proteins, including an ATP/GTP binding motif and a putative insertion site for retroelements (Bringaud et al., 2002a) (Fig. 2). This result confirmed the presence of *T. brucei* RHS-like sequences in the subtelomeric regions of *T. cruzi* (Bringaud et al., 2002a). The block containing the telomeric unit (excepting the 189-bp junction) plus the *rhs* (pseudo)gene (Seq3Tc sequence) is 2.6 kbp long, and part of it was duplicated upstream of an *asp-2* (amastigote surface protein-2) (pseudo)gene (coordinates 18,156 to 20,611) from group II of the *ts*-like family (Low and Tarleton, 1997).