Copy number variations of genomic and transcriptomic motifs of Mucin and MASP superfamilies in different Trypanosoma cruzi strains

Anderson Coqueiro dos Santos[1], Gabriela Flavia Rodrigues Luiz[1], Najib M. El-Sayed[2], Santuza Maria Ribeiro Teixeira[1], João Luís Reis Cunha[1], Daniella Bartholomeu[1],

1 Institute of Biological Sciences, UFMG
2 Department of Parasite Genomics, Institute for Genomic Research

Abstract

Trypanosoma cruzi is the causative agent of Chagas disease, an illness that afflicts about 7 million people worldwide. Due to its extensive genetic variability, T. cruzi taxa is divided into six discrete typing units (DTUs), named TcI to TcVI. The first T. cruzi genome was sequenced in 2005, allowing the identification of hundreds of genes encoding polymorphic surface proteins from trans-sialidase (TcS), MASP and mucin (TcMUC) superfamilies. These genes are enrolled in cellular adhesion and invasion, and immune evasion processes, highlighting their important role in host parasite interactions. The high number of copies and variability of these gene families hinders the assignment of reads to specific genes. Members of these families share short motifs whose occurrence and abundance can be used to estimate the variability of these gene families among T. cruzi strains. In the present work, we aim to compare the motif amplifications in the multigene families MASP, TcMUC and TcS, derived from representatives of T.cruzi TcI, TcII and TcVI DTUs, and compare their copy number with gene expression levels. We used genomic reads from two clones derived from TcVI CL strain (CL Brener, CL-14), Y strain (TcII) and 3 representatives from DTU TcI; as well as transcriptomic reads from the same clones/strains in the amastigote, epimastigote and trypomastigote stages. These genomic and transcriptome reads were mapped in CL Brener genome, and only reads that mapped in the TcS, TcMUC and MASP genes were recovered. Kmers of 30 nucleotides were generated from these reads and their coverage was estimated. To remove redundancy, similar kmers were clustered and the number of motifs, for each library, was normalized to allow comparisons among them. For TcMUC superfamily, the analysis showed a greater number of motifs in the virulent CL Brener clone, compared to a reduced number of motifs in the non-virulent CL-14 clone. We did not detect large differences within the three clones derived from Y strain and little variance in TcI genomic motifs. When we compared the different developmental stages, a greater concordance was found between the amastigote 60 and 96 hrs from CL-14, whereas in CL Brener the pattern observed in amastigote 96 hrs and trypomastigote stage was quite similar. Similar results were observed for MASPs. We are currently performing this analysis with TcS family and comparing the motifs with higher counts in genomic and transcriptomic analysis to try to correlate copy number of the identified motifs and their expression levels.

Funding: CAPES, CNPq and FAPEMIG