

Ensamblaje del genoma de una cepa colombiana de *Trypanosoma cruzi* I (TcI) procedente del departamento del Tolima como herramienta para el análisis de variabilidad genética y evolución.

Autores:

*María Camila Hoyos Sánchez*¹; *Brayhan Dario Suarez*²; *Aura María Rodríguez*²; *Hader Sebastián Ospina Zapata*²; *Valentina Herrera Sánchez*²; *Carlos Mario Ospina*²; *Hamilton Julián Barbosa*²; *Julio Cesar Carranza Martínez*²; *Gustavo Adolfo Vallejo*²; *María Clara Echeverry*³; *Daniel A. Urrea*²; *Jorge Duitama*¹

Correo: bdsuarezr@ut.edu.co

Afiliación:

¹ TICSw: Tecnologías de Información y Construcción de Software, Universidad de los Andes.

² Laboratorio de Investigaciones en Parasitología Tropical (LIPT), Universidad del Tolima.

³ Infecciones y Salud en el Trópico, Universidad Nacional de Colombia, Sede Bogotá.

Resumen:

El agente causal de la enfermedad de Chagas (*Trypanosoma cruzi*), es un parásito digénico que afecta alrededor de 7 millones de personas y en promedio 14.000 mueren al año. *T. cruzi* tiene una amplia diversidad genética y está dividido en seis DTUs; (TcI – TcVI), TcI es la DTU más abundante en Colombia y pesar de los avances en el estudio de haplotipos de *T. cruzi* en el país, existe un gran desconocimiento acerca de la genómica de este parásito ya que sus genes se encuentran en grupos dispersos de repeticiones en tándem. La mayoría de los ensamblajes de *T. cruzi* I se han realizado con tecnologías 454, Sanger e Illumina, estos ensamblajes están altamente fragmentados debido a la complejidad del genoma.

De acuerdo con lo mencionado anteriormente, el presente trabajo tiene como objetivo el proporcionar el ensamblaje genómico y anotación de genes de alta calidad de una cepa colombiana de *T. cruzi* I para analizar la variabilidad genética dentro y entre especies de la familia Trypanosomatidae.

Se empleó una cepa *T. cruzi* caracterizada como TcI aislada de *Didelphis marsupialis* procedente de municipio de Coyaima-Tolima. Usando Gentra Puregene Kit se obtuvo ADN de alta calidad y la secuenciación se realizó mediante la tecnología Pacific Bioscience (PacBio) HiFi. Se obtuvieron ensamblajes haploides de 120 y 313 contigs con Hifiasm y Flye, respectivamente. La cepa tiene un genoma diploide de aproximadamente 90 Mb (~40 Mb haploide), y alrededor de 14,000 genes. Usando distintas herramientas (BUSCO, Canu, Hifiasm total y NGSEP ploidía 1 y 2) se reporta por primera vez un ensamblaje en el que se separa la información de las dos copias de cada cromosoma. Se realizó el mapeo de las lecturas crudas de PacBio a diferentes ensamblajes de referencia, incluidos los dos genomas haploides y el genoma intermedio ensamblado. Se hizo llamado de variantes con el software NGSEPlinux_4.1.0 donde se reporta una baja cantidad de variantes de las lecturas alineadas a los ensamblajes realizados a partir de las mismas lecturas. Se realizó también un dendrograma con genomas TcI aislados de Colombia, donde se observó 2 grupos diferentes, en un extremo se posicionaron todas las cepas de parásitos aislados de un paciente con VIH

y cardiomiopatía, en el otro extremo se posicionaron todas las cepas de parásitos aislados de un paciente chagásico agudo infectado por transmisión oral y nuestro ensamblaje. Se determinaron los genes parálogos en el genoma de *T. cruzi* (ensamblaje Flye) y se determinaron los genes ortólogos presentes en 5 diferentes especies de la familia Trypanosomatidae incluido nuestro ensamblaje de *T. cruzi* mediante OrthoFinder.

Se concluye que las nuevas tecnologías de secuenciación ayudan a mejorar la calidad del genoma de *T. cruzi* y a entender la complejidad de este parásito, se demuestra la importancia de utilizar tecnologías de lecturas largas en genomas con alta complejidad como *T. cruzi*, para entender la variabilidad dentro de la DTU TcI y este trabajo propone esta cepa como un genoma de referencia para *T. cruzi* en Colombia.