Infection, Genetics and Evolution 50 (2017) 25–27

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Short communication

Phylogenomics of *Trypanosoma cruzi*: Few evidence of TcI/TcII mosaicism in TcIII challenges the hypothesis of an ancient TcI/TcII hybridization

Nicolás Tomasini *, Patricio Diosque

Instituto de Patología Experimental, Facultad de Ciencias de la Salud, Universidad Nacional de Salta – CONICET, Av. Bolivia 5150, 4400, Salta, Salta, Argentina

ARTICLE INFO

Article history:
Received 9 November 2016
Received in revised form 8 February 2017
Accepted 9 February 2017
Available online 10 February 2017

ABSTRACT

Phylogenetic relationships among major lineages of *Trypanosoma cruzi* are still debatable. Particularly, it is controversial the origin of two main lineages: TcIII and TcIV. Some authors proposed that these lineages have been the result of an ancient hybridization between TcI and TcII, and this was one of the most accepted evolutionary models in the scientific community for several years. In the present paper we analyse several genomes of *T. cruzi* in order to examine if there is evidence supporting that TcIII is an ancient TcI/TcII hybrid. Our results show that TcIII is mainly related to TcI and not to TcII and there is few evidence of mosaicism for TcIII. Our results challenge the hypothesis of the ancient TcI/TcII hybridization.

Trypanosoma cruzi, the aethiological agent of Chagas disease, infects several millions of people around the world. Several major Discrete Typing Units (DTUs) of the parasite have been described and clearly defined (Zingales et al., 2009). Such DTUs have been named as TcI to TcVI (and recently a seventh candidate DTU found in bats has been described) (Lima et al., 2015; Zingales et al., 2009). However, phylogenetic relationships among these DTUs are still controversial. Since several years, it is widely agreed that TcV and TcVI are relatively recent TcII/TcIII hybrids (Brisse et al., 1998; Brisse et al., 2003; Flores-Lopez and Machado, 2011; Lewis et al., 2011; Tomasini and Diosque, 2015; Sturm and Campbell, 2010; Sturm et al., 2003; Westenberger et al., 2005). However, it has also been proposed that TcIII and TcIV are the result of an ancient hybridization between TcI and TcII (Sturm and Campbell, 2010; Sturm et al., 2003; Westenberger et al., 2005; Elias et al., 2005; Tomazi et al., 2009; Ienne et al., 2010). This hypothesis was firstly based on the observation of mixed SNPs in TcIII and TcIV, some were shared with TcII and others with TcI (Westenberger et al., 2005). Accordingly, it has been proposed that these mosaic sequences of TcI and TcII were caused by recombination and loss of heterozygosis after an ancient hybridization (Westenberger et al., 2005). Such hybridization has been challenged by other authors (Flores-Lopez and Machado, 2011; Machado and Ayala, 2001). Even, in a recent paper (Tomasini and Diosque, 2015) we made a reanalysis of the data previously analysed by Westenberger et al. (2005). We questioned that mosaic sequences were only apparent. In this regard, the inclusion of an outgroup in the analysis showed that SNPs shared by TcIII/TcIV and TcII were also shared by the outgroup suggesting they are ancestral features. Such ancestral SNPs (also called plesiomorphies) are not an evidence of common ancestry between TcII and TcIII/TcIV (Tomasini and Diosque, 2015). However, controversy is still installed in the scientific community (Barnabe et al., 2016; Breniere et al., 2016). Other authors also proposed a third model with three ancestral DTUs (TcI, TcII, TcIII) although they did not clarify the phylogenetic relationships among such DTUs (de Freitas et al., 2006). In the present paper, we provide evidence from a phylogenomic analysis in order to show that TcIII is closely related to TcI. In addition, there is weak evidence of mosaicism in TcIII challenging the hypothesis of an ancient hybridization between TcI and TcII.

Currently, there are sequenced genomes for three DTUs of *T. cruzi*. Unfortunately, neither TcIII nor TcIV representative genomes are yet available. However, the first sequenced genome for *T. cruzi* belongs to a TcVI strain (CL-Brener) (El-Sayed et al., 2005). Because TcVI is a hybrid between TcII and TcIII, we actually have available information at genomic level about TcIII, although mixed with TcII. In addition, the CL-Brener genome project included the sequencing of the Esmeraldo strain (TcII) which helped the researchers to classify a high percentage of the CL-Brener genome as Esmeraldo-like (TcII) or Non-Esmeraldo-like (TcIII) (El-Sayed et al., 2005). Despite around 50% of the annotated genome correspond to repetitive regions that cannot be classified as Esmeraldo or Non-Esmeraldo, most of the core genome (around 30 Mb) was classified. Even the genome of CL-Brener was divided into two different genomes and it was organized in contigs and scaffolds at the chromosome level for each parental (Weatherly et al., 2009). Consequently, although no genome of a TcIII strain has been yet sequenced, the Non-Esmeraldo-like is a relatively good representative of the core genome of a TcIII strain. In addition, the recently sequenced genome of *T. cruzi marinkellei* (Franzen et al., 2012), a different subspecies of *T. cruzi* which is used as outgroup in many phylogenetic analyses, is essential to correctly conclude about phylogenetic relationships between TcI,

* Corresponding author.
E-mail address: nicotomasini@yahoo.com.ar (N. Tomasini).

http://dx.doi.org/10.1016/j.meegid.2017.02.010
1567-1348/

TcII and TcIII and to analyse if Non-Esmeraldo-like (TcIII) genome is a mosaic of TcI and TcII sequences (Tomasini and Diosque, 2015).

We downloaded from TriTrypDB (http://tritrypdb.org) and analysed complete genomes from CL-Brener Esmeraldo-like (herein CLBRTcII), CL-Brener Non-Esmeraldo-like (herein CLBRTcIII), *T. cruzi* Sylvio X10 (TcI) and *T. cruzi marinkellei* strain B7 (outgroup). The genomes were aligned using the software MAUVE (Darling et al., 2004) with default parameters. A total of 1,334,419 SNPs (excluding sites that implied a gap or an N) resulted from the alignment of the four genomes. We analysed parsimony-informative sites. In this case, parsimony-informative corresponded to sites where two genome share the same base and the other two share a different base (e.g. outgroup = A, TcI = A, $CLBR_{TcII}$ = G, $CLBR_{TcIII}$ = G, note that this SNP supports $CLBR_{TcII}$/$CLBR_{TcIII}$ cluster). Particularly, 96,621 SNPs were parsimony-informative about phylogenetic relationships among the genomes. In order to analyse such phylogenetic relationships, genomes were fragmented in 500 bp segments and each one was classified according the support (number of SNPs) to the alternatives phylogenies. We used a sign test to evaluate the statistical significance of the number of SNPs supporting one of the alternative topologies. We finally analysed such fragments that supported one of the phylogenies according the *p* value on the sign test. From a total of 4407 fragments with $p < 0.05$, 98.5% clustered TcI with $CLBR_{TcIII}$ and just 1.5% clustered $CLBR_{TcII}$ with $CLBR_{TcIII}$ (Fig. 1). However, many fragments may have low support caused by mosaicism in such fragments. Alternatively, we scanned the entire dataset of parsimony-informative SNPs for runs of 4 consecutive polymorphic sites supporting the same topology. We observed that most of the runs of 4 SNPs (21,560 = 94.6%) favoured the clustering of TcI with $CLBR_{TcIII}$ and just 1164 favoured the alternative topology. However, considering the size of the dataset, many runs of 4 SNPs may be also caused by random homoplasy. Consequently, we implemented the RDP algorithm (Martin et al., 2005) to scan the genome in order to detect recombination breakpoints in a more statistical way. The RDP algorithm scans the genome SNP by SNP (and using a window of 5 SNPs in this case) looking for regions where sequence distance is shorter between $CLBR_{TcII}$ and $CLBR_{TcIII}$ than TcI and $CLBR_{TcIII}$. Just 41 regions in the genome were statistically significant breakpoints ($p < 0.05$, 706 SNPs corresponding to 0.7% of the total of parsimony-informative SNPs). Consequently, these results show that most of the $CLBR_{TcIII}$ genome cluster with TcI and not with $CLBR_{TcII}$. These results challenge the hypothesis that TcIII genome is a mosaic of TcII and TcI sequences as previously proposed (Sturm and Campbell, 2010; Westenberger et al., 2005).

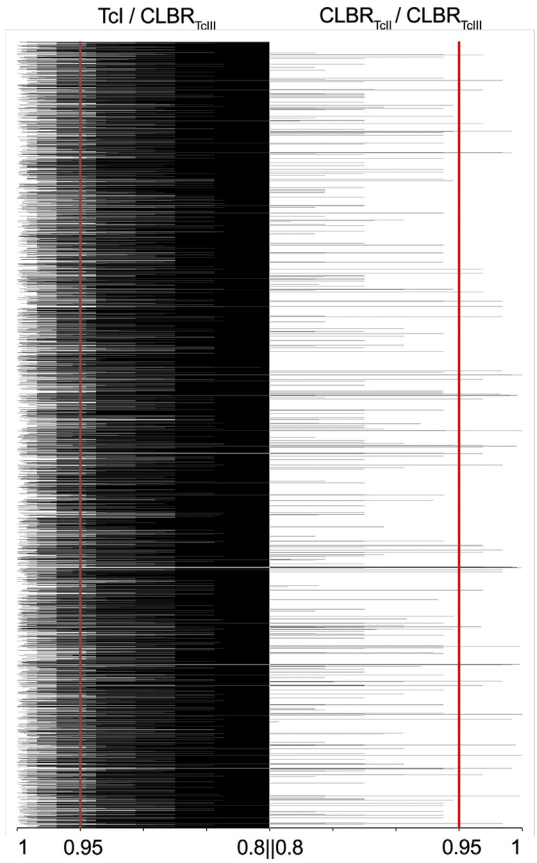

Fig. 1. Schematic representation of the genome of CL-Brener strain supporting alternative clusters. The genome was divided in 500 bp fragments and only informative fragments about phylogenetic relationships were kept. Lines shown at the left side represent fragments that supported TcI/$CLBR_{TcIII}$ cluster whereas lines at the right side represent fragments that supported $CLBR_{TcII}$/$CLBR_{TcIII}$ cluster. Length of the lines indicates [1 – *p* value] support to the cluster according the sign test. Red lines identifies support higher than 0.95. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Despite that our results suggest that most of the $CLBR_{TcIII}$ genome cannot be called a "mosaic", the few regions that support the $CLBR_{TcII}$/$CLBR_{TcIII}$ may suggest some introgression of TcII into the ancestral TcIII genome. However, such low percentage of sequences may be also explained by recent gene conversion in the CL-Brener genome or simply assembly errors. Around 50% of such regions showed identical sequences of $CLBR_{TcII}$ and $CLBR_{TcIII}$ which is better explained by recent gene conversion in CL-Brener strain. However, although introgression sounds a less likely scenario than gene conversion, it cannot be discarded. In this regard, a meiosis-like mechanism plus gamete fusion as mode of reproduction cannot be discarded for *T. cruzi* (Lewis et al., 2009). If this is the case, successive backcrosses of the TcI/TcII hybrid with the parental TcI may have reduced the proportion of TcII genome in TcIII leading to a few introgressed TcII sequences. A number of *n* backcrosses of the hybrid with TcI would reduce the TcII content in an average proportion of 0.5^n. Consequently, it will be required at least 7 backcrosses to obtain <1% of TcII sequences in the $CLBR_{TcIII}$ genome. Alternatively, inbreeding in the TcI/TcII hybrid may have also contributed to reduce the TcII sequences. Genome sequencing of a TcIII strain will help to determine if such few TcII sequences into the $CLBR_{TcIII}$ genome are caused by introgression or gene conversion.

Concluding, this is the first phylogenomic analysis of *T. cruzi* lineages showing that TcIII is mainly related to TcI and not to TcII as we previously proposed (Tomasini and Diosque, 2015) and in agreement with others (Flores-Lopez and Machado, 2011; Machado and Ayala, 2001). Our study shows there is very few evidence of mosaics in the genome of TcIII contrary to the proposed by (Westenberger et al., 2005). The genome sequencing of any reference TcIII strain will finally help to elucidate if some minor degree of introgression may have occurred in the origin of the TcIII.

Acknowledgements

This work has been supported by the ANPCyT budget PICT-2014-2449.

References

Barnabe, C., Mobarec, H.I., Jurado, M.R., Cortez, J.A., Breniere, S.F., 2016. Reconsideration of the seven discrete typing units within the species *Trypanosoma cruzi*, a new proposal of three reliable mitochondrial clades. Infect. Genet. Evol. 39, 176–186.

Breniere, S.F., Waleckx, E., Barnabe, C., 2016. Over six thousand *Trypanosoma cruzi* strains classified into Discrete Typing Units (DTUs): attempt at an inventory. PLoS Negl. Trop. Dis. 10, e0004792.

Brisse, S., Barnabe, C., Banuls, A.L., Sidibe, I., Noel, S., Tibayrenc, M., 1998. A phylogenetic analysis of the *Trypanosoma cruzi* genome project CL Brener reference strain by multilocus enzyme electrophoresis and multiprimer random amplified polymorphic DNA fingerprinting. Mol. Biochem. Parasitol. 92, 253–263.

Brisse, S., Henriksson, J., Barnabe, C., Douzery, E.J., Berkvens, D., Serrano, M., De Carvalho, M.R., Buck, G.A., Dujardin, J.C., Tibayrenc, M., 2003. Evidence for genetic exchange and hybridization in *Trypanosoma cruzi* based on nucleotide sequences and molecular karyotype. Infect. Genet. Evol. 2, 173–183.

Darling, A.C., Mau, B., Blattner, F.R., Perna, N.T., 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14, 1394–1403.

de Freitas, J.M., Augusto-Pinto, L., Pimenta, J.R., Bastos-Rodrigues, L., Goncalves, V.F., Teixeira, S.M., Chiari, E., Junqueira, A.C., Fernandes, O., Macedo, A.M., Machado, C.R., Pena, S.D., 2006. Ancestral genomes, sex, and the population structure of *Trypanosoma cruzi*. PLoS Pathog. 2, e24.

Elias, M.C., Vargas, N., Tomazi, L., Pedroso, A., Zingales, B., Schenkman, S., Briones, M.R., 2005. Comparative analysis of genomic sequences suggests that *Trypanosoma cruzi* CL Brener contains two sets of non-intercalated repeats of satellite DNA that correspond to *T. cruzi* I and *T. cruzi* II types. Mol. Biochem. Parasitol. 140, 221–227.

El-Sayed, N.M., Myler, P.J., Bartholomeu, D.C., Nilsson, D., Aggarwal, G., Tran, A.N., Ghedin, E., Worthey, E.A., Delcher, A.L., Blandin, G., Westenberger, S.J., Caler, E., Cerqueira, G.C., Branche, C., Haas, B., Anupama, A., Arner, E., Aslund, L., Attipoe, P., Bontempi, E., Bringaud, F., Burton, P., Cadag, E., Campbell, D.A., Carrington, M., Crabtree, J., Darban, H., da Silveira, J.F., de Jong, P., Edwards, K., Englund, P.T., Fazelina, G., Feldblyum, T., Ferella, M., Frasch, A.C., Gull, K., Horn, D., Hou, L., Huang, Y., Kindlund, E., Klingbeil, M., Kluge, S., Koo, H., Lacerda, D., Levin, M.J., Lorenzi, H., Louie, T., Machado, C.R., McCulloch, R., McKenna, A., Mizuno, Y., Mottram, J.C., Nelson, S., Ochaya, S., Osoegawa, K., Pai, G., Parsons, M., Pentony, M., Pettersson, U., Pop, M., Ramirez, J.L., Rinta, J., Robertson, L., Salzberg, S.L., Sanchez, D.O., Seyler, A., Sharma, R., Shetty, J., Simpson, A.J., Sisk, E., Tammi, M.T., Tarleton, R., Teixeira, S., Van Aken, S., Vogt, C., Ward, P.N., Wickstead, B., Wortman, J., White, O., Fraser, C.M., Stuart, K.D., Andersson, B., 2005. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. Science 309, 409–415.

Flores-Lopez, C.A., Machado, C.A., 2011. Analyses of 32 loci clarify phylogenetic relationships among *Trypanosoma cruzi* lineages and support a single hybridization prior to human contact. PLoS Negl. Trop. Dis. 5, e1272.

Franzen, O., Talavera-Lopez, C., Ochaya, S., Butler, C.E., Messenger, L.A., Lewis, M.D., Llewellyn, M.S., Marinkelle, C.J., Tyler, K.M., Miles, M.A., Andersson, B., 2012. Comparative genomic analysis of human infective *Trypanosoma cruzi* lineages with the bat-restricted subspecies *T. cruzi marinkellei*. BMC Genomics 13, 531.

Ienne, S., Pedroso, A., Carmona, E.F.R., Briones, M.R., Zingales, B., 2010. Network genealogy of 195-bp satellite DNA supports the superimposed hybridization hypothesis of *Trypanosoma cruzi* evolutionary pattern. Infect. Genet. Evol. 10, 601–606.

Lewis, M.D., Llewellyn, M.S., Gaunt, M.W., Yeo, M., Carrasco, H.J., Miles, M.A., 2009. Flow cytometric analysis and microsatellite genotyping reveal extensive DNA content variation in *Trypanosoma cruzi* populations and expose contrasts between natural and experimental hybrids. Int. J. Parasitol. 39, 1305–1317.

Lewis, M.D., Llewellyn, M.S., Yeo, M., Acosta, N., Gaunt, M.W., Miles, M.A., 2011. Recent, independent and anthropogenic origins of *Trypanosoma cruzi* hybrids. PLoS Negl. Trop. Dis. 5, e1363.

Lima, L., Espinosa-Alvarez, O., Ortiz, P.A., Trejo-Varon, J.A., Carranza, J.C., Pinto, C.M., Serrano, M.G., Buck, G.A., Camargo, E.P., Teixeira, M.M., 2015. Genetic diversity of *Trypanosoma cruzi* in bats, and multilocus phylogenetic and phylogeographical analyses supporting Tcbat as an independent DTU (discrete typing unit). Acta Trop. 151, 166–177.

Machado, C.A., Ayala, F.J., 2001. Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of *Trypanosoma cruzi*. Proc. Natl. Acad. Sci. U. S. A. 98, 7396–7401.

Martin, D.P., Williamson, C., Posada, D., 2005. RDP2: recombination detection and analysis from sequence alignments. Bioinformatics 21, 260–262.

Sturm, N.R., Campbell, D.A., 2010. Alternative lifestyles: the population structure of *Trypanosoma cruzi*. Acta Trop. 115, 35–43.

Sturm, N.R., Vargas, N.S., Westenberger, S.J., Zingales, B., Campbell, D.A., 2003. Evidence for multiple hybrid groups in *Trypanosoma cruzi*. Int. J. Parasitol. 33, 269–279.

Tomasini, N., Diosque, P., 2015. Evolution of *Trypanosoma cruzi*: clarifying hybridisations, mitochondrial introgressions and phylogenetic relationships between major lineages. Mem. Inst. Oswaldo Cruz 110, 403–413.

Tomazi, L., Kawashita, S.Y., Pereira, P.M., Zingales, B., Briones, M.R., 2009. Haplotype distribution of five nuclear genes based on network genealogies and Bayesian inference indicates that *Trypanosoma cruzi* hybrid strains are polyphyletic. Genet. Mol. Res. 8, 458–476.

Weatherly, D.B., Boehlke, C., Tarleton, R.L., 2009. Chromosome level assembly of the hybrid *Trypanosoma cruzi* genome. BMC Genomics 10, 255.

Westenberger, S.J., Barnabe, C., Campbell, D.A., Sturm, N.R., 2005. Two hybridization events define the population structure of *Trypanosoma cruzi*. Genetics 171, 527–543.

Zingales, B., Andrade, S.G., Briones, M.R., Campbell, D.A., Chiari, E., Fernandes, O., Guhl, F., Lages-Silva, E., Macedo, A.M., Machado, C.R., Miles, M.A., Romanha, A.J., Sturm, N.R., Tibayrenc, M., Schijman, A.G., 2009. A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. Mem. Inst. Oswaldo Cruz 104, 1051–1054.