Deep sequencing to detect diversity of *Trypanosoma cruzi* infection in patients co-infected with HIV and Chagas disease

Natalie M Bowman, University of North Carolina at Chapel Hill School of Medicine, Department of Medicine, Division of Infectious Diseases, Chapel Hill, NC 27599, United States

Sujata Balasubramanian, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC 27599, United States

Robert H Gilman, Johns Hopkins Bloomberg School of Public Health, Department of International Health, Baltimore, MD 21205, United States

Christian Parobek, University of North Carolina at Chapel Hill School of Medicine, NC 27599, United States

Maritza Calderon, Universidad Peruana Cayetano Heredia, Lima, Peru

Andreea Waltmann, University of North Carolina at Chapel Hill, Institute of Global Health and Infectious Diseases, Chapel Hill, NC 27599, United States

Louisa A. Messenger, London School of Hygiene and Tropical Medicine, Department of Disease Control, Keppel Street, London, WC1E 7HT, United Kingdom

Leny Sanchez, Universidad Peruana Cayetano Heredia, Lima, Peru

Caryn Bern, University of California-San Francisco School of Medicine, Department of Epidemiology and Biostatistics, San Francisco, CA 94158, United States

Jonathan J Juliano, University of North Carolina at Chapel Hill School of Medicine, Department of Medicine, Division of Infectious Diseases, Chapel Hill, NC 27599, United States

Working Group on Chagas Disease in Bolivia and Peru: Daniel Clark, Jorge Flores, Roni Colanzi, Jeong Choi, Gerson Galdos, Mauricio Dorn, Omar Gandarilla, Enzo Fortuny, Anne Palumbo, Lisbeth Ferrufino, Monica Pajuelo, Melissa Reimer, Sandra Mendoza Guerrero.

Corresponding author's contact information:

Natalie Bowman

nbowman@med.unc.edu

CB#7036

Medical Biomolecular Research Building, room 2341

111 Mason Farm Road

Chapel Hill, NC 27599-7036

919-843-4385

Summary: Amplicon deep sequencing can be used to characterize multiplicity of infection and genotype *Trypanosoma cruzi* infections in patient co-infected with human immunodeficiency virus.

ABSTRACT

Chagas disease, caused by *Trypanosoma cruzi*, can reactivate and cause severe acute disease in immunocompromised patients such as those infected with HIV. We conducted amplicon deep sequencing of a 327-base pair fragment of the *tcscd5* gene using an Ion Torrent PGM directly from clinical samples from HIV patients with high parasitemia. We describe the within host diversity, both characterizing the discrete typing unit (DTUs) of the infections and confirming the presence of multi-strain infections, directly from clinical samples. This method can rapidly provide information on the genetic diversity of *T. cruzi* infection, which can have direct impacts on clinical disease.

KEYWORDS: *Trypanosoma cruzi*, Chagas disease, HIV, genetic diversity, discrete typing unit, multiclonality, amplicon deep sequencing, *tcsc5d*

INTRODUCTION

Trypanosoma cruzi, the cause of Chagas Disease, infects 5-6 million people worldwide, with more than 38,000 new infections occurring yearly.[1] *T. cruzi* is diploid, with a clonal population structure. However, the species is genetically diverse, with seven discrete typing units (DTUs) that infect humans.[2] Individual DTUs have distinct, but often overlapping, geographic distributions, with TcI predominating in human infections north of the Amazon and TcII, TcV, and TcVI being most common in the Southern Cone of South America, including Bolivia.[3] TcIII and TcIV are primarily sylvatic DTUs found mainly in the Amazon forest.[3] DTUs isolated from human infections usually reflect common local strains, confounding attempts to definitively correlate DTU with clinical outcomes, but some studies have found associations between DTU and gastrointestinal symptoms, progression to cardiomyopathy, and acute orally transmitted disease, as summarized in Messenger, *et al*.[2]

There is a paucity of data, however, about DTU-related clinical differences in HIV coinfections, though multiple DTUs have been reported in HIV-infected patients.[2] In patients infected with HIV, reactivation of chronic *T. cruzi* infection can cause severe neurological syndromes with mortality as high as 80%.[4] Bolivia has the highest prevalence of Chagas disease in the world at about 6%[1] and exceeding 50% in some villages in the Bolivian Chaco.[5] In a study of acutely ill HIV-infected patients in Santa Cruz, Bolivia, we found that 25% were also seropositive for *T. cruzi* (manuscript in preparation).

Next generation sequencing (NGS) is a powerful tool to understand genetics and epidemiology of infections that has been rarely utilized to study the epidemiology or pathogenesis of Chagas disease. In this study, we demonstrate the utility of amplicon deep sequencing to characterize the diversity of chronic *T. cruzi* infection in immunocompromised persons. In contrast to chronically infected immunocompetent persons, HIV/*T. cruzi*-

coinfected persons frequently exhibit very high levels of parasitemia due to immunosuppression, which provides sufficient genetic material for sequencing without first culturing the parasite. Culture of *T. cruzi* can select for certain strains more adapted to growth *in vitro*, which can lead to underestimation of parasite diversity and mismeasurement of haplotype frequencies.[6] Sequencing directly from clinical samples can provide a more accurate representation of the diversity and relative frequencies of parasite populations in human infection. Amplicon deep sequencing from clinical samples is sensitive enough to detect mixed genotype infections and avoids biases introduced by culture. We targeted the *tcsc5d* gene, which putatively encodes a C-5 sterol desaturase (lathosterol/episterol oxidase) on chromosome 22, because it is a single-copy nuclear gene that contains single nucleotide polymorphisms (SNPs) without significant insertions or deletions, allowing us to estimate multiplicity of infection. Additionally, it can be used to distinguish between different DTUs.[7] Here we show that deep sequencing of *T. cruzi* directly from clinical samples is possible with sufficient levels of parasitemia.

METHODS

Blood samples were obtained from HIV-positive subjects from Cochabamba (HIV167) and Santa Cruz (HIV70, HIV76, HIV141), Bolivia. *T. cruzi* infection was confirmed by at least two serological tests and by quantitative polymerase chain reaction (qPCR) as described below.[8] Human blood samples were also examined by microscopy. DNA from stock strains (maintained by the Miles laboratory, London School of Tropical Medicine and Hygiene) of each DTU maintained in culture were analyzed in parallel for reference (Supplemental table) with the exception of the reference strain from DTU IV, which was degraded and did not amplify; however, TcIV is primarily a sylvatic strain that very rarely infects humans.

DNA was extracted from either clotted blood (all samples) or whole blood with EDTA and guanidine (HIV167) using a Qiacube automated system with Qiagen reagents (Qiagen N.V.,

Hilden, Germany). Presence of *T. cruzi* was detected using quantitative PCR with the Cruzi1 and Cruzi 2 primers and Cruzi3 probe using FastStart Universal Probe Master (Rox) reagents from Roche Diagnostics according to published methods.[8] Samples were considered positive if Ct was less than the cutoff determined using a standard curve.

A fragment of the *tcsc5d* gene (GenBank: HQ586986 (CL-Brener) and JN050564–JN050587; https://tritrypdb.org/tritrypdb/app/record/gene/TcCLB.507853.10) was amplified using hemi-nested PCR with outer primers forward 5'-GGACGTGGCGTTTGATTTAT and reverse 5'-TCCCATCTTCTTCGTTGACT-3' using published conditions.[7] The second reaction was performed using inner forward primer 5'-CCTTGTGATGGATTGGTCA-3' with the same reverse primer to obtain a 326 bp long fragment (See Supplemental Table S1 for reaction conditions). A 10-base barcode tag was included at the 5' end of the inner forward primer so amplicons could be pooled during library preparation. Hemi-nested PCR was performed using Phusion Taq (NEB, Ipswich, MA) for the outer PCR and Roche Diagnostics FastStart High Fidelity enzyme blend (Hoffman-La Roche, Basel, Switzerland) to generate the inner PCR fragment. Amplicon libraries were generated from pooled amplicons using the Ion Plus Fragment Library Kit and Ion Xpress™ Barcode Adapters 1-16 kit following manufacturer instructions. Sequencing was performed on the Ion Torrent platform using a 318-size chip by the UNC High Throughput Sequencing Facility. Samples were amplified and sequenced in duplicate.

Sequence extraction, clustering, and comparison of clusters from replicates were all done using SeekDeep version 2.1.1 (https://seekdeep.brown.edu/). [9] Data was first demultiplexed according to barcode and then replicate PCRs were compared to find haplotypes. The demultiplexed sequences were matched to barcodes and primers and both were removed. The quality filtering, also using Seekdeep, was done using a sliding window of 50 bases, moving by 5 bases, requiring an average quality score of 20 or above setting a minimum length of 300bp. Replicates from resulting filtered sequences were then clustered

to compared to find haplotypes. Only haplotypes that were found in both sample replicates were retained. Replicates where prevalence of any given haplotype was absolutely different by more than 15% were rejected. Samples that had fewer than 200 reads in any replicate were rejected. Sequences were aligned using MEGA version 10.0.[10] Data were visualized using SeekDeep.[9]

Ethics: This analysis was exempt from Institutional Review Board approval at UNC because only de-identified samples were used. The studies that served as sources of human samples were approved by the Institutional Review Boards at John Hopkins Bloomberg School of Public Health (Baltimore, Maryland, USA); the London School of Tropical Medicine and Hygiene (London, United Kingdom); Universidad Católica Boliviana (Santa Cruz, Bolivia); Hospital Clínico Viedma (Cochabamba, Bolivia); Instituto de Desarrollo Humano (Cochabamba, Bolivia); Colectivo de Estudios Aplicados, Desarrollo Social, Salud y Medio ambiente (Cochabamba, Bolivia); Asociación Benéfica PRISMA (Lima, Peru); and Universidad Peruana Cayetano Heredia (Lima, Peru).

RESULTS

We amplified and sequenced the target amplicon from four samples from HIV-infected patients (Table 1) as well as cultures of reference DTUs. After quality filters were applied, we identified 10 unique haplotypes (Figure 1A and Supplemental Table S2). The predominant haplotype in the four human samples was identical to the TcV reference DTU. The distribution of single nucleotide polymorphisms (SNP) across the amplicon varied, with some areas showing high levels of diversity and others not (Figure S1 and S2).

At least one (HIV76) and possibly two (HIV70) samples showed evidence of infection by at least two strains of *T. cruzi* and potentially as many as four based on the number of haplotypes identified. *T. cruzi* is a diploid organism, making it impossible to distinguish

between infection by two unique strains or a strain heterozygous at the target gene; however, the widely different haplotype frequencies in both samples, with haplotype 1 representing the majority of reads with less than 20% abundance for any other haplotype, suggest that these represent polyclonal infections rather than heterozygosity.

Haplotype sequences included several SNPs useful for genotyping that were consistent with DTU TcV,[7] the most common in Bolivian human infections. Blood from patient HIV70 was cultured in Santa Cruz, Bolivia. The culture was extracted and resulting DNA was analyzed using PCR-restriction fragment length polymorphism analysis with the restriction enzymes hsp60 and gpi (Figure S3). The PCR-RFLP pattern was most consistent with DTU TcV or VI, concordant with our sequencing results. The dominant haplotype in HIV70 is identical to TcV, as expected. Haplotype 4, the minority population found in this patient, differs by only one nucleotide from one of the haplotypes (haplotype 9) sequenced from the TcVI reference strain (Figure 1B), also a relatively frequent cause of human infections in Bolivia.[2]

DISCUSSION

Though polyclonal infections have been characterized from culture and using other modalities,[11,12] to our knowledge no one has used deep sequencing to detect polyclonal *T. cruzi* infections directly from clinical samples. Multiplicity of infection does occur in chronic Chagas disease, and here we show that it can be seen in reactivation, defined by microscopically detectable parasitemia, in HIV-infected patients. HIV/*T. cruzi*-coinfected persons often exhibit high-level parasitemia that provides the opportunity to study multiplicity of infection.[13] Here we show that deep sequencing can be done directly from clinical samples, though sensitivity is currently limited to qPCR Ct values of less than 25-28 (approximately 0.0005-0.00005 ng parasite DNA/µl) by the primers we used. Though we only present data from four patients here, we provide proof-of-concept for use of this technique. To our knowledge, this work represents the first use of this genotyping scheme

directly from human samples (without passage through culture) and the first use of deep sequencing to characterize *T. cruzi* populations in HIV-infected patients.

One reason for limited sensitivity is the use of a single-copy nuclear gene like *tcsc5d*, which distinguishes between different copies within a single genome and multiple infection within a host, in contrast to other schemes that target multi-copy genes.[11] The *tcsc5d* gene is useful for DTU-typing,[7] though our truncated amplicon could not definitively distinguish all six compared to sequencing the full gene; however, the predominant haplotype in our clinical samples was identical to the TcV reference strain, the most common DTU in human infections in this area of Bolivia.[3] We included results from the reference strains to demonstrate the robustness of our methods. The TcII strain contained two haplotypes (7 and 8), which could reflect *T. cruzi's* diploid genome, though it could also represent PCR or sequencing artifact, as the haplotypes differed by only 2 SNPs. TcV is a hybrid of TcII and TcII, but we isolated only one haplotype. TcVI, also a TcII/TcIII hybrid, contained two haplotypes that differed by 6 SNPs.[14]When we aligned the haplotypes against reference *tcsc5d* sequences (Figure S2),[7] we identified a number of conserved heterozygous SNPs in TcV and TcVI. The levels of heterozygosity observed in this short fragment for TcV and TcVI are consistent with other *T. cruzi* housekeeping genes.[15] Though similar across this fragment, there is a SNP at position 111 which can distinguish between TcV and TcVI, although we acknowledge that gene conversion events in other TcV samples could potentially confound this.

In conclusion, this work demonstrates the feasibility of amplicon deep sequencing for *T. cruzi* directly from clinical samples, without need for passage through culture or xenodiagnosis. Culture from blood or by xenodiagnosis is laborious and inefficient, requiring up to several weeks to grow sufficient parasite quantities; furthermore, culture can select for culture-adapted strains, impose genetic bottlenecks reducing diversity in multiclonal populations, and introduce bias into sequencing results.[6] The *tcsc5d* gene is an ideal sequencing target

because it is a single copy nuclear gene, allowing more accurate determination of multiplicity of infection, and it is sufficiently variable to detect inter- and intra-DTU variation. While other genotyping algorithms compatible with deep sequencing exist, notably use of the mini-exon region,[12] they are unable to differentiate all seven DTUs. This is the first time *tcsc5d* has been used to characterize parasite genetic diversity in human infections directly from extracted blood samples using deep sequencing technology. We hope this technique can be extended to investigate intra-host parasite dynamics, parasite evolution, and transmission patterns.

FUNDING

This work was supported by the National Institute of Allergy and Infectious Diseases [grant number K23 AI113197 to NMB, K24 AI13499 to JJJ, P30 AI50410 to NMB]; the Burroughs Wellcome Fund/American Society of Tropical Medicine and Hygiene [to NMB]; and the Fogarty International Center [R25 TW009340 to NMB].

ACKNOWLEDGEMENTS

The Working Group on Chagas Disease in Bolivia and Peru includes Daniel Clark, Jorge Flores, Roni Colanzi, Jeong Choi, Gerson Galdos, Mauricio Dorn, Omar Gandarilla, Enzo Fortuny, Anne Palumbo, Lisbeth Ferrufino, Monica Pajuelo, Melissa Reimer, Sandra Mendoza Guerrero.Thank you to the Michael Miles group (London School of Tropical Medicine and Hygiene) for the reference strains. We acknowledge Steven R. Meshnick's (now deceased) contributions as well.

REFERENCES

1. World Health Organization. Chagas disease in Latin America: an epidemiological update based on 2010 estimates. Wkly Epidemiol Rec. **2015**; 90:33–44.

2. Messenger LA, Miles MA, Bern C. Between a bug and a hard place: Trypanosoma cruzi genetic diversity and the clinical outcomes of Chagas disease. Expert Rev Anti Infect Ther. **2015**; 13(8):995–1029.

3. Brenière SF, Waleckx E, Barnabé C. Over Six Thousand Trypanosoma cruzi Strains Classified into Discrete Typing Units (DTUs): Attempt at an Inventory. PLoS Negl Trop Dis. **2016**; 10(8):e0004792.

4. Almeida EA de, Ramos Júnior AN, Correia D, Shikanai-Yasuda MA. Co-infection Trypanosoma cruzi/HIV: systematic review (1980-2010). Rev Soc Bras Med Trop. **2011**; 44(6):762–770.

5. Samuels A, Clark E, Galdos-Cardenas G, et al. Epidemiology of and impact of insecticide spraying on Chagas disease in communities in the Bolivian Chaco. PLoS Negl Trop Dis. **2013**; 7(8):e2358.

6. Devera R, Fernandes O, Coura JR. Should Trypanosoma cruzi be called “cruzi” complex? A review of the parasite diversity and the potential of selecting population after in Vitro culturing and mice infection. Mem Inst Oswaldo Cruz. Fundação Oswaldo Cruz; **2003**; 98(1):1–12.

7. Cosentino RO, Agüero F. A simple strain typing assay for Trypanosoma cruzi: discrimination of major evolutionary lineages from a single amplification product. PLoS Negl Trop Dis. **2012**; 6(7):e1777.

8. Piron M, Fisa R, Casamitjana N, et al. Development of a real-time PCR assay for Trypanosoma cruzi detection in blood samples. Acta Trop. **2007**; 103(3):195–200.

9. Hathaway NJ, Parobek CM, Juliano JJ, Bailey JA. SeekDeep: single-base resolution de novo clustering for amplicon deep sequencing. Nucleic Acids Res. **2018**; 46(4):e21–e21.

10. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. **2018**; 35(6):1547–1549.

11. Llewellyn MS, Messenger LA, Luquetti AO, et al. Deep Sequencing of the Trypanosoma cruzi GP63 Surface Proteases Reveals Diversity and Diversifying Selection among Chronic and Congenital Chagas Disease Patients. PLoS Negl Trop Dis. **2015**; 9(4):e0003458.

12. Villanueva-Lizama L, Teh-Poot C, Majeau A, Herrera C, Dumonteil E. Molecular Genotyping of Trypanosoma cruzi by Next-Generation Sequencing of the Mini-Exon Gene

Reveals Infections With Multiple Parasite Discrete Typing Units in Chagasic Patients From Yucatan, Mexico. J Infect Dis. Oxford Academic; **2019**; 219(12):1980–1988.

13. Freitas V de, Silva S da, Sartori A, et al. Real-time PCR in HIV/Trypanosoma cruzi coinfection with and without Chagas disease reactivation: association with HIV viral load and CD4 level. PLoS Negl Trop Dis. **2011**; 5(8):e1277.

14. Lewis MD, Llewellyn MS, Yeo M, Acosta N, Gaunt MW, Miles MA. Recent, Independent and Anthropogenic Origins of Trypanosoma cruzi Hybrids. PLoS Negl Trop Dis. Public Library of Science; **2011**; 5(10):e1363.

15. Yeo M, Mauricio I, Messenger L, et al. Multilocus sequence typing (MLST) for lineage assignment and high resolution diversity studies in Trypanosoma cruzi. PLoS Negl Trop Dis. **2011**; 5(6):e1049.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to report.

FUNDING STATEMENT

This work was supported by the National Institute of Allergy and Infectious Diseases [grant number K23 AI113197 to NMB, K24 AI13499 to JJJ]; the Burroughs Wellcome Fund/American Society of Tropical Medicine and Hygiene [to NMB]; the University of North Carolina Center for AIDS Research [P30 AI50410 to NMB]; and the Fogarty International Center [R25 TW009340 to NMB].

MENTION OF MEETINGS WHERE PREVIOUSLY PRESENTED

Preliminary findings from this work were presented as a poster at the October 2015 annual meeting of the American Society of Tropical Medicine and Hygiene in Philadelphia, PA.

CURRENT AFFILIATIONS

Sujata Balasubramanian: College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, Texas

Christian Parobek: Department of Obstetrics and Gynecology, Women and Infants Hospital, Brown University, Providence, Rhode Island

Leny Sanchez: Instituto Nacional de Salud, Laboratorio de VIH, Lima, Perú

Table 1: Patient Characteristics

Code	Age	Sex	CD4 count	HIV viral load	Micromethod result	Parasite load	Clinical description
HIV70	41	Male	179	83,653	Positive	556,726 (clot)	Diagnosed with HIV 1 year prior to enrollment; sensory and motor deficit below T5; hypotensive (BP 80/60); LP with protein 130, glucose 18, 2 TNC; rapidly declined and developed respiratory failure and died 22 days after enrollment
HIV76	43	Male	NA	NA	Positive	228,424 (clot)	Newly diagnosed HIV; diarrhea, pneumonia, malnutrition, GCS 12; Died 6 days after enrollment
HIV141	57	Male	NA	NA	Positive	134,254 (clot)	Newly diagnosed HIV; gastroenteritis, dehydration, oropharyngeal thrush; died 3 months after enrollment
HIV167	27	Male	15	6705	Positive	9951 (clot) 21,462 (EDTA)	Newly diagnosed HIV; pneumonia, gastroenteritis, oropharyngeal thrush, malnutrition (weight 42 kg); mild bilateral tremor; LAFB on EKG; normal MMSE

EKG: electrocardiogram; GCS: Glasgow Coma Score; HIV: human immunodeficiency virus; LAFB: left anterior fascicular block; LP: lumbar puncture; MMSE: mini mental status exam; NA: not available; TNC: total nucleated cells

Figure 1A: *tcsc5d* Haplotype distribution in human clinical samples, cultured reference strains, and a dog from rural Bolivia.

Figure 1B: Heatmap showing the number of SNPs different (blue and pink) and percent sequence identity (red and orange) between haplotypes.

Figure 1A

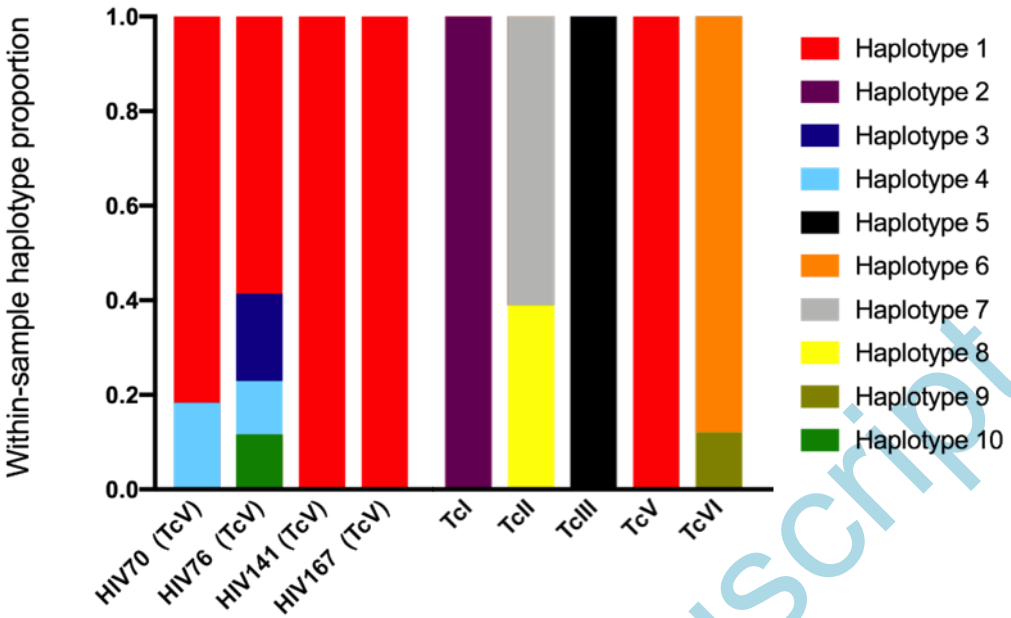

Figure 1b