Whole genome sequencing of the Chagas disease vector *Rhodnius ecuadoriensis*

A Bacigalupo[2]; P Schwabl[2]; D C Hernández[4]; K Brunker[2]; J D Ramírez[4];
L E Hernandez-Castro[3]; B Cheaib[2]; A G Villacís[5]; M J Grijalva[1]; K R Elmer[2]; M S Llewellyn[2]

[1]Ohio University, United States; [2]Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, UK;
[3]The Roslin Institute, University of Edinburgh, UK; [4]Facultad de Ciencias Naturales, Universidad del Rosario, Colombia;
[5]Centro de Investigación para la Salud en América Latina, Pontificia Universidad Católica del Ecuador, Ecuador.

Introduction

Chagas disease or American trypanosomiasis is a chronic illness caused by the protozoan *Trypanosoma cruzi* (Trypanosomatida: Trypanosomatidae). *T. cruzi* is transmitted to mammals by hematophagous insect vectors (Fig. 1) of the Subfamily Triatominae (Hemiptera: Reduviidae), comprising more than 150 species.

Remarkably, only three genomes of triatomine species are available. More genomes will enable the study of the biology, ecology and evolution of individual species, and the Subfamily in general. The goal of this study is to provide new genomic information about *Rhodnius ecuadoriensis*, the main vector of Chagas disease in Ecuador and Northern Peru. This species maintains domestic, peridomestic and sylvatic populations, presenting high *T. cruzi* infection rates.

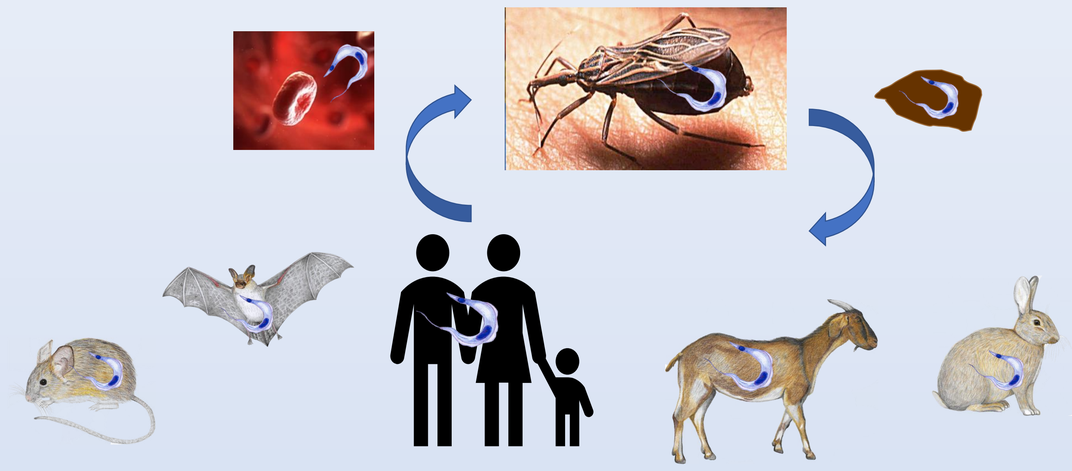

Figure 1. The vectorial transmission cycle of *Trypanosoma cruzi* between triatomine insects and mammals: Triatomines feed on infected mammal blood and multiply the parasite in their gut. When feeding again, they release infective forms in their dejections. The parasite enters through mucous membranas or wounds to a susceptible mammal. Alimentary route by ingesting infected triatomines is also possible.

Methodology

DNA from homogenized head, thorax and legs of one *R. ecuadoriensis* was extracted using a commercial kit and its quality was measured by spectroscopy; its concentration was evaluated by fluorescence and its integrity by automated electrophoresis. *R. ecuadoriensis*' DNA was sequenced using two approaches: Oxford Nanopore in a MinION flowcell to obtain long reads, and Illumina HiSeq for short reads. The hybrid assembler MaSuRCA 3.3.9 was used to incorporate both read types in the assembly; a second assembly was produced by MaSuRCA, with the same short reads, but including filtered Nanopore reads by read length, setting a minimum of 200 bp, and a quality score of 90 with the Filtlong 0.2.0 software. The draft genomes were evaluated by an intrinsic quality assessment, describing the proportion of the genome included in the assembly. The selected genome assembly was improved using SSPACE Standard 3.0 for extending contigs and scaffolding (Fig. 2). Later, the completeness of the resulting draft assembly was quantified by the accurate presence of single-copy orthologous from Insecta (insecta_odb10) and Hemiptera (hemiptera_odb10) lineage databases using benchmarking universal single-copy ortholog genes BUSCO 4.0.6 and compared to those of available triatomine genomes.

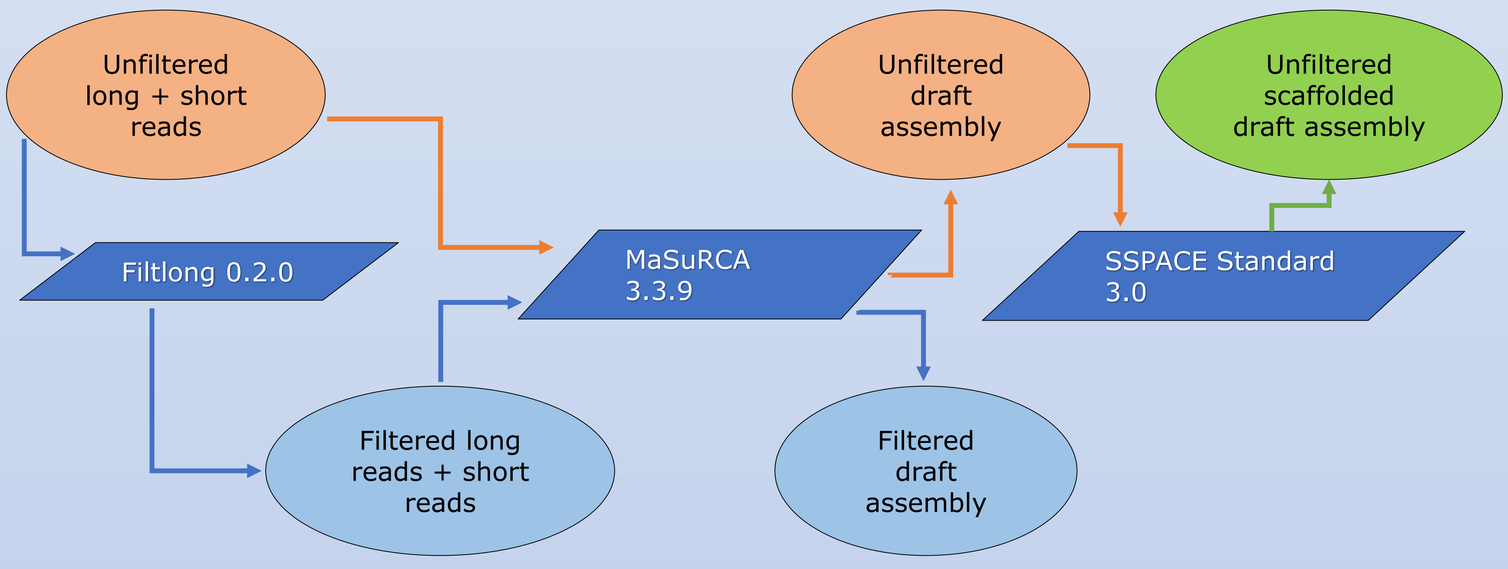

Figure 2. Schematic representation of the workflow to obtain the draft assembly.

Table 1. Comparison of assembly statistics of the filtered and unfiltered draft assemblies of *R. ecuadoriensis*.

Assembly statistics	Filtered	Unfiltered
Sum (bp)	536,322,282	559,324,901
Total number of scaffolds	7,110	5,866
Average scaffold size	75,432.1	95,350.3
N50	178,860	295,667
E-size	258,454	470,586
Estimated genome size	481,399,794	481,399,794

Results

The selected *R. ecuadoriensis* sample presented 1.81 of the 260/280 ratio; its concentration was 29.8 ng/µl, and the electrophenogram showed DNA fragment distribution from 1,850 to 22,342 bp, with an average size of 8,912 bp.

A total of 47.1 million Illumina paired reads of 150 bp each and 1,506,368 Oxford Nanopore reads were obtained from the same individual. GC content of short reads was 34%, and 33% in long reads.

The unfiltered draft assembly of *R. ecuadoriensis* produced a higher quality assembly than the one with filtered long reads, taking into account total number of scaffolds, average scaffold size, N50 and E-size (Table 1).

There was an overall reduction in the number of scaffolds composing the assembly after the use of SSPACE, mainly in those below 25,000 bp (Fig. 3).

When comparing the scaffolded draft assembly using the Benchmarking Universal Single-Copy Orthologs (BUSCO) of the Insecta database, *R. ecuadoriensis* presented the highest completeness (Table 2). Using the Hemiptera database, it was second, behind *R. prolixus* (Table 3).

Table 2. Comparison of BUSCO results of *R. ecuadoriensis* against the three other available triatomine genomes, using the Insecta database.

BUSCO report Insecta database	*R. ecuadoriensis* n (%)	*R. prolixus* n (%)	*T. rubrofasciata* n (%)	*T. infestans* n (%)
Complete BUSCOs	**1,318 (96.4)**	1,306 (95.5)	1,308 (95.7)	1,236 (90.4)
Complete and single copy	**1,310 (95.8)**	1,296 (94.8)	1,264 (92.5)	1210 (88.5)
Complete and duplicated	**8 (0.6)**	10 (0.7)	44 (3.2)	26 (1.9)
Fragmented BUSCOs	**15 (1.1)**	26 (1.9)	20 (1.5)	28 (2.0)
Missing BUSCOs	**34 (2.5)**	35 (2.6)	39 (2.8)	103 (7.6)
Total BUSCOs searched	**1,367 (100)**	1,367 (100)	1,367 (100)	1,367 (100)

Table 3. Comparison of BUSCO results of *R. ecuadoriensis* against the three other available triatomine genomes, using the Hemiptera database.

BUSCO report Hemiptera database	*R. ecuadoriensis* n (%)	*R. prolixus* n (%)	*T. rubrofasciata* n (%)	*T. infestans* n (%)
Complete BUSCOs	**2,435 (97.0)**	2,466 (98.2)	2,402 (95.7)	2,297 (91.6)
Complete and single copy	**2,405 (95.8)**	2,438 (97.1)	2,324 (92.6)	2,253 (89.8)
Complete and duplicated	**30 (1.2)**	28 (1.1)	78 (3.1)	44 (1.8)
Fragmented BUSCOs	**22 (0.9)**	16 (0.6)	26 (1.0)	40 (1.6)
Missing BUSCOs	**53 (2.1)**	28 (1.2)	82 (3.3)	173 (6.8)
Total BUSCOs searched	**2,510 (100)**	2,510 (100)	2,510 (100)	2,510 (100)

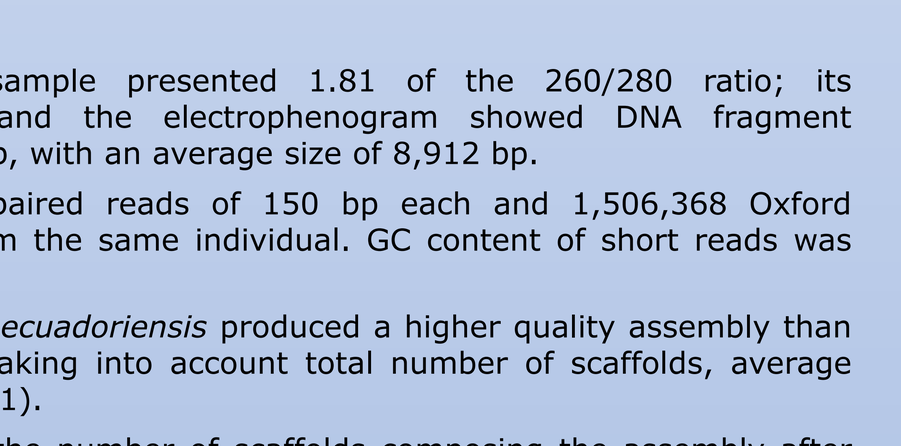

Figure 3. Comparison of scaffold length distribution in *R. ecuadoriensis* assemblies.

Discussion

Genomics of vectors can provide clues on targets to prevent and control vectorial diseases. Using a short and long read combined approach, we have produced a draft assembly of *R. ecuadoriensis* that shows comparable quality in terms of completeness to previously published triatomine genomes. Transcriptomics and genome annotation are now underway.

Funding: Agencia Nacional de Investigación y Desarrollo ANID - Programa Becas - Doctorado Becas Chile 2019 - 72200391
Drawings' credits: Instagram @ilustra.nativa