Draft long-read assembly and annotation of the Chagas disease vector *Rhodnius prolixus*

Antonella Bacigalupo[1*], Carolina Hernández[2], Nathalia Ballesteros[3], Marina Muñoz[3], Juan David Ramírez[3], Kathryn R. Elmer[1], Martin S. Llewellyn[1]

*a.bacigalupo.1@research.gla.ac.uk

[1]School of Biodiversity, One Health and Veterinary Medicine, University of Glasgow, Scotland, United Kingdom
[2]Centro de Tecnología en Salud (CETESA), Innovaseq SAS, Bogotá, Colombia
[3]Centro de Investigaciones en Microbiología and Biotecnología-UR (CIMBIUR), Facultad de Ciencias Naturales, Universidad del Rosario, Bogotá, Colombia

Introduction

Chagas disease is a chronic infection with the protozoan parasite *Trypanosoma cruzi* (Trypanosomatida: Trypanosomatidae), which is transmitted by insect vectors of the subfamily Triatominae (Hemiptera: Reduviidae) [Fig. 1A].

American trypanosomiasis continues to be neglected, even among the 'neglected tropical diseases', with no reduction in associated disability-adjusted life years (DALYs) during recent years. In this context, seven years after the publication of the first whole genome of a triatomine species, only two genomes from other triatomine species have been released. Our aim is to enable a resurgence of Chagas disease vector research via sequencing, assembly, and annotation of both known and new triatomine vector genomes of *Rhodnius prolixus*, *Belminus herreri*, *Mepraia spinolai*, *Psammolestes arthuri*, *Rhodnius brethesi* and *Rhodnius ecuadoriensis*. Among these species, *R. prolixus* is considered a primary vector of *T. cruzi* throughout Central and South America [Fig. 1B].

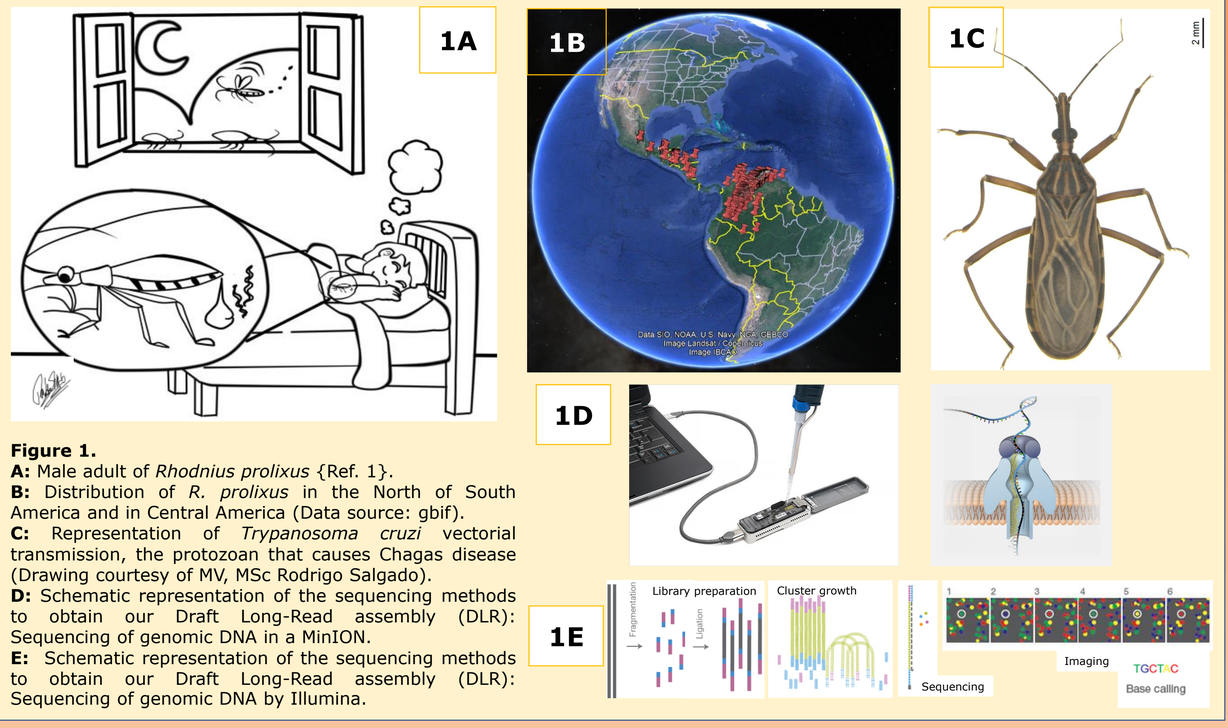

Figure 1.
A: Male adult of *Rhodnius prolixus* {Ref. 1}.
B: Distribution of *R. prolixus* in the North of South America and in Central America (Data source: gbif).
C: Representation of *Trypanosoma cruzi* vectorial transmission, the protozoan that causes Chagas disease (Drawing courtesy of MV, MSc Rodrigo Salgado).
D: Schematic representation of the sequencing methods to obtain our Draft Long-Read assembly (DLR): Sequencing of genomic DNA in a MinION.
E: Schematic representation of the sequencing methods to obtain our Draft Long-Read assembly (DLR): Sequencing of genomic DNA by Illumina.

Material and Methods

As first step, we extracted the DNA and sequenced one *R. prolixus* individual [Fig. 1C] by long (Oxford Nanopore Technologies) [Fig. 1D] and short (Illumina) reads [Fig. 1E], assembled the long reads using Flye 2.8.3-b1695, polished the obtained genome assembly with Racon 1.5.0, and posteriorly we scaffolded and annotated it with the software RagTag 2.1.0 and GeMoMa 1.9.0, using as reference the previous assemblies and annotations for *R. prolixus* (GCA_000181055.3; RproC3.5) and the tropicopolitan triatomine species *Triatoma rubrofasciata* (Triatoma_chr_assembly; Triatoma_chr_genome) {Refs. 2-4}. Visualization of the genomes was performed using assembly-stats 17.02.

Results

Our new assembly has 1,270 scaffolds (Figs. 2A, 3B & 3F) and a N50 of 1,466,963 (Fig. 3D) that includes 18,279 genes coding for 18,328 mRNAs with 93,193 CDSs. The benchmarking universal single copy ortholog (BUSCO) gene completeness of our draft assembly reached 98.4% of the hemiptera_odb10 (Fig. 2B) and 97.8% of the insecta_odb10 databases, respectively, which are higher than the 96.6% and 95.1% of the current *R. prolixus* reference, and the 98.2% and 97.7% of the *T. rubrofasciata* chromosomal assembly, respectively {Ref. 5}.

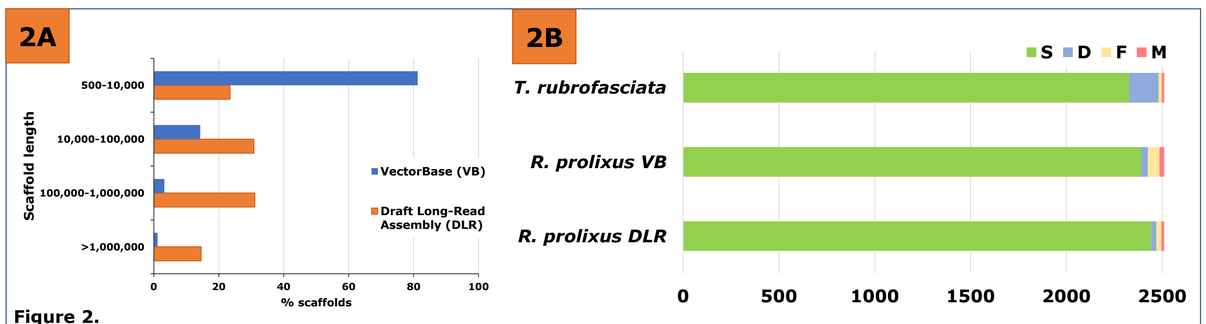

Figure 2.
A: Comparison of the percentage of scaffold lengths composition between the VB (blue) and DLR (Orange) assemblies.
B: Graphical representation of BUSCO completeness among the *T. rubrofasciata*, *R. prolixus* VectorBase (VB) reference and our Draft Long-Read assembly (**DLR**). S: single copy; D: duplicated; F: fragmented; M: missing BUSCOs.

Discussion

We will complement these encouraging results by combining this reference-based annotation with ab initio and transcriptome evidence, and we will follow this procedure for the other triatomine species.
These genomes will allow for new research on the genomic basis for adaptation in these vectors.

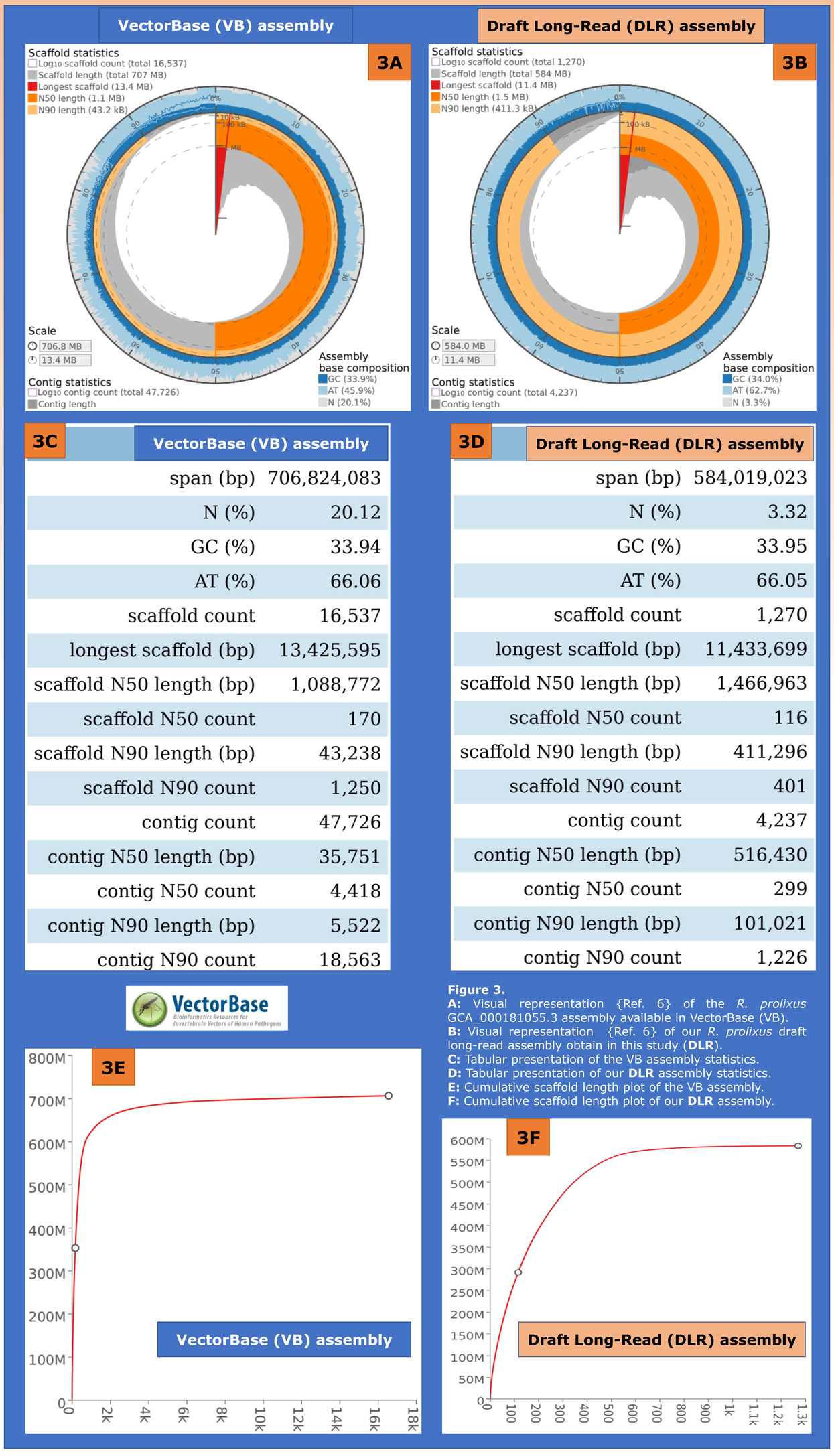

Figure 3.
A: Visual representation {Ref. 6} of the *R. prolixus* GCA_000181055.3 assembly available in VectorBase (VB).
B: Visual representation {Ref. 6} of our *R. prolixus* draft long-read assembly obtain in this study (**DLR**).
C: Tabular presentation of the VB assembly statistics.
D: Tabular presentation of our **DLR** assembly statistics.
E: Cumulative scaffold length plot of the VB assembly.
F: Cumulative scaffold length plot of our **DLR** assembly.

References

1) Rossana Falcone, Juliana Damieli Nascimento. Coleção de Triatominae FCFAR - Unesp Araraquara. https://www2.fcfar.unesp.br/#!/triatominae/subfamilia-triatominae/rhodnius/rhodnius-prolixus/ Accessed 28 July 2022.
2) Mesquita et al. 2015 Genome of *Rhodnius prolixus*, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. Proc Natl Acad Sci U S A 112(48): 14936-41. https://doi.org/10.1073/pnas.1506226112
3) Amos et al. 2021. VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center. Nucleic Acids Res gkab929. https://doi.org/10.1093/nar/gkab929
4) Liu et al. 2019. A chromosomal-level genome assembly for the insect vector for Chagas disease, *Triatoma rubrofasciata*. Gigascience 8(8): giz089. https://doi.org/10.1093/gigascience/giz089.
5) Manni et al. 2021. BUSCO Update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol 38(10): 4647–4654.
6) Challis & Grüning. 2017. Assembly statistic visualisation. https://github.com/rjchallis/assembly-stats

Funding: This work was supported by Minciencias Convenio 727 DIeI from UR Colombia; ANID - Programa Becas - Doctorado Becas Chile 2019 72200391; and Wellcome [204820/Z/16/Z].