Article

The Complete Nucleotide Sequence and Gene Organization of the Mitochondrial Genome of Triatoma boliviana (Hemiptera, Reduviidae, Triatominae) and Phylogenetic Comparisons

Sebastián Pita 1, Pablo Mora 2, Mirko Rojas-Cortez 3,4, Teresa Palomeque 2, Pedro Lorite 2,* and Francisco Panzera 1, *

1 Evolutionary Genetic Section, Faculty of Science, University of the Republic, Iguá 4225, Montevideo 11400, Uruguay; spita@fcien.edu.uy
2 Department of Experimental Biology, Genetics, University of Jaén, Paraje las Lagunillas sn., 23071 Jaén, Spain; pmora@ujaen.es (P.M.); tpalome@ujaen.es (T.P.)
3 ISGlobal, Barcelona Institute for Global Health, Carrer Rosselló 132, 08036 Barcelona, Spain; mirko.rojas@isglobal.org
4 Fundación Salud Naturaleza Integral (SANIT), Pasaje Fidelia de Sánchez N° 0344, Cochabamba, Bolivia
* Correspondence: plorite@ujaen.es (P.L.); fcpanzera@gmail.com (F.P.)

Abstract: The complete mitogenome of Triatoma boliviana Martínez, Chávez, Sossa, Aranda, Vargas and Vidaurre, 2007 was assembled using next generation sequencing data. The 16,719 bp long genome contains 13 protein coding genes, 22 transfer RNAs, two ribosomal RNAs, and a control region. This mitogenome showed similar nucleotide composition, gene order and orientation than other triatomines. Molecular phylogenetic analyses based on available complete mitogenomes from Reduviidae supported that Triatominae is a monophyletic group and that T. boliviana is basal to the two main Triatomini clades: North and South American. In addition, the analysis of a fragment of the 16S mitochondrial gene among Triatomini species, including species of the dispar lineage, supports the inclusion of T. boliviana in this group.

Keywords: complete mitogenome; kissing bugs; phylogeny; Triatoma boliviana

1. Introduction

Kissing bugs are hemipteran insects of the subfamily Triatominae, which includes more than 150 blood-sucking species grouped into 5 tribes and 18 genera, most of them with medical relevance as vectors of the protozoan parasite Trypanosoma cruzi, an etiological agent of Chagas disease [1,2]. Although the most recent molecular phylogenetic studies support that Triatominae is a monophyletic group, this issue has been controversial and a possible polyphyletic or even paraphyletic origin has been discussed [3,4]. In Triatominae, the most speciose and diverse tribe is Triatomini, with more than 110 species in 9 genera. The species of this tribe has been classified into lineages (also named clades or groups), complexes and subcomplexes, according to morphological, geographical, ecological, and, more recently, molecular data [1–7]. Currently, three major lineages or clades are recognized within the Triatomini tribe: South American, North American and the dispar lineage [3,4]. The latter includes only five Triatoma species (T. dispar, T. boliviana, T. carrioni, T. nigromaculata and T. venosa), distributed from Costa Rica, Panama, Colombia, Venezuela, Peru, and Ecuador, and reaching its southernmost limit in the Andean regions of Bolivia (T. boliviana) [3]. With the exception of the sylvatic T. boliviana [8], the other dispar species are involved in domestic–peridomestic T. cruzi transmission to humans [9]. These species have been grouped according to their morphological similarities, since the genetic information available for them is very limited [10]. What is more, there are no molecular data at all supporting the inclusion of T. boliviana within this group, nor to differentiate it from other related species. In all phylogenetic analyses, using both nuclear and mitochondrial
sequences, the dispar lineage was recovered as a sister group to all remaining Triatomini, being a clade that split early during the diversification of the Triatomini tribe [3,4,7,10,11]. For this reason, the dispar species clade is always in a basal position in the evolutionary trees of the tribe.

The present paper provides, for the first time, molecular information about *T. boliviana* through the description of its mitogenome. Phylogenetic analyses have been performed using the *T. boliviana* mitogenome and all available mitogenomes from Triatominae, including species from North and South American lineages. A fragment of the 16S gene has also been analyzed. The sequence of this region is available not only for species of the North and South American lineages but also for several species of the dispar lineage. The obtained results support the inclusion of *T. boliviana* in the dispar lineage.

### 2. Materials and Methods

A sample from a *T. boliviana* population was collected during 2008 in the locality of Vilaque (Chuma community), Ildefonso de las Muñecas Province, La Paz Department in Bolivia (15°39′ S and 68°52′ W at 2570 m above sea level). The male specimen selected for DNA extraction was conserved in absolute ethanol at −20 °C until DNA extraction.

DNA extraction was performed using leg muscle tissue with the Monarch® Genomic DNA Purification Kit (New England Biolabs® Inc., Ipswich, MA, USA) following the manufacturer’s instructions.

Low coverage sequencing was performed using the DNBseq™ sequencing platform at BGI, Hong Kong, which yielded 1.23 Gbp of PE150 reads. The mitogenome was assembled *de novo* employing NOVOPlasty v4.3.1 [12]. NOVOPlasty is a program, written in Perl, which uses whole genome sequencing (WGS) data to assemble organelle genomes. The assembly has to be initiated by a sequence that acts as a seed, and, afterwards, the assembly is extended bidirectionally. This seed sequence is a query to retrieve reads with similarity to the NGS data set [12]. In addition, an optional “bait” reference mitochondrial genome can be added to the analysis. In our analysis, K-mer length was set to 23, and the seed used was the *cytb* fragment from *Panstrongylus rufotuberculatus* from Bolivia (GenBank acc. number MZ64716). Two different runs where carried out, with and without a mitogenome as reference, using, in the latter case, the mitogenome from non-Andean *T. infestans* (GenBank acc. number KY640305) [13].

The annotation of the *T. boliviana* mitogenome was conducted using the web based services MITOS [14] (available in http://mitos.bioinf.uni-leipzig.de, accessed on 31 March 2022). The annotations of protein coding genes (PCGs) were refined by checking manually for consistent start/stop codons, open reading frames and by comparison with other *Triatoma* mitogenomes using Geneious 10.1.3 (Biomatters Ltd., Auckland, New Zealand). To determine the stop codons’ positions, we followed the procedure used in the annotation of the *Triatoma dimidiata* mitogenome [15], i.e., if the coding sequence end is within the 5′ end of transfer RNA (tRNA) sequences, then the 5′ end of the tRNA serves as the termination signal [16,17]. Consequently, incomplete T or TA stop codons are generated. The poly(A) tail generated by polyadenylation of the 3′-end of the transcript completes the TAA stop codon [18]. The annotations of rRNA genes were extended until adjacent tRNAs [18]. The 5′ end of the srRNA gene that is not flanked by a tRNA was determined by comparison with other mitogenomes. The base composition and codon usage of PCGs were analyzed with MEGA X [19]. Sequence similarities were also evaluated using MEGA X [19].

Currently, there is little information concerning mitochondrial sequences of the dispar species group: *T. dispar* (col), *T. venosa* (16S); *T carrioni* (12S, 16S, colI) and *Triatoma* sp. (16S). This last sequence comes from a museum specimen collected in Ecuador (Sucumbios) by Lent and Wygodzinsky [1] and sequenced by Hwang and Weirauch [11] and also used in Justi et al. [7]. Unfortunately, other recently reported sequences from the dispar group have not been submitted to GenBank yet [10], so we could not use them in our analyses. Despite limited sequence information, molecular phylogenies (both with nuclear and mitochondrial fragments) place the dispar species as the first clade to diverge during
the cladogenesis process of the tribe Triatomini [3,4,7,10,11]. This distinctive trait could help us to resolve whether *T. boliviana* belongs, or not, to the dispar lineage. Complete mitogenomes were compared with all available Reduviidae species recovered from a BLAST search on GenBank using *T. boliviana* as a query. Search results included mitogenome sequences from 22 triatomine species (20 of Triatomini and 2 of Rhodniini tribes) and 30 Reduviidae species from 9 subfamilies: Centrocnemidinae, Ectrichodiinae, Epiroderinae, Harpactorinae, Holoptilinae, Reduviinae, Peiratinae, Stenopodainae and Tribelocephalinae. Complete sequences were edited to start from the same nucleotide base, taking *T. infestans* mitogenome as a reference, and were aligned using MAFFT v7.453 software [20]. The alignment was later automatically trimmed using TrimAl v1.2 software [21] with the option “gappyout”. The final dataset employed to build the tree was 14,327 bp long and bears the PCGs, all tRNAs, the rDNA genes and the G + C rich region of the control region. The regions trimmed corresponded to spacers and the majority of the control region, which, in several species, was absent. The phylogenetic relationships were reconstructed using the maximum likelihood method, implemented in the R packages ape_5.4-1 [22] and phangorn_2.5.5 [23], following the author’s vignettes [24]. For tree visualization, ggtree R package [25] was employed.

Comparisons of genetic distances using a mitochondrial 16S ribosomal fragment (548 bp) were carried out using a dataset including the 22 triatomine species plus all dispar species found in NCBI (three sequences) plus *T. boliviana* here analyzed, and some extra species belonging to the South American Triatomini lineage. This gene was the one with the highest number of available sequences in GenBank, considering the different dispar species (*T. carrioni*, *T. venosa* and *Triatoma* sp.). All selected sequences were larger than 500 bp. Sequences were aligned and trimmed following the same protocol as for the complete mitogenome dataset, employing MAFFT and TrimAl (final alignment with 548 bp). For the distance calculation, R package ape_5.4-1 [22] was employed. The plot was drawn using the R package ggplot2 [26].

3. Results and Discussion

3.1. Mitogenome of Triatoma boliviana

The complete *T. boliviana* mitogenome is 16,719 bp in length (GenBank accession number OM830311). The size of the other Triatominae mitogenomes ranged from 15,699 to 17,323 bp, with an average size of 16,305 bp [27]. The *T. boliviana* mitogenome contains the regions common in animal mitogenomes, i.e., 13 PCGs, 22 transfer RNAs, 2 ribosomal RNAs and the control region, with 23 genes located in the H strand and 13 genes in the L strand (Figure 1, Supplementary Table S1). The gene order and orientation were the same as previously described in other triatomines [13,15,26,28,29]. Gene overlaps were found at 11 gene junctions. The longest (14 bp) was found between ATPase8/ATPase6. Other gene overlap between PCGs was found for nd4/nd4L. Three canonical start codons, ATN, GTG and TTG, have been found in the invertebrate PCGs [30] but the 13 PCGs of the *T. boliviana* mitogenome use the regions common in animal mitogenomes, i.e., 13 PCGs, 22 transfer RNAs, 2 ribosomal RNAs and the control region, with 23 genes located in the H strand and 13 genes in the L strand (Figure 1, Supplementary Table S1). The gene order and orientation were the same as previously described in other triatomines [13,15,26,28,29]. Gene overlaps were found at 11 gene junctions. The longest (14 bp) was found between ATPase8/ATPase6. Other gene overlap between PCGs was found for nd4/nd4L. Three canonical start codons, ATN, GTG and TTG, have been found in the invertebrate PCGs [30] but the 13 PCGs of the *T. boliviana* mitogenome use ATN. ATG star codon was found in seven genes (*col*, ATPase6, collIII, nd4, nd4L, nd6 and cytB), whereas ATT is the start codon for nd2 and nd5, ATA for coll, nd3 and nd1, and ATC for ATPase8. Six out of the thirteen PCGs have the TAA stop codon (nd2, atp8, atp6, nd6, cytB and nd1) and one of them the TAG stop codon (nd4L). Six PCGs have incomplete TA and T stop codons (*col*, coll, collIII, nd3, nd5 and nd4). Incomplete stop codons are usual in metazoan mitogenomes.

The *T. boliviana* mitogenome A + T content for the H-strand was 74.1%. This bias towards A and T nucleotides is a typical feature of invertebrate mitogenomes. The codon usage of the PCGs also indicates the preference for the use of A + T codons (Supplementary Table S2). There is an asymmetric use of codons for the same amino acids, as can be observed by the relatively synonymous codon usage values (RSCU is the number of times a codon appears in a gene in relation with the number of expected occurrences under equal codon usage). For synonymous codons, higher RSCU values were observed for NNT or NNA codons for all amino acids (Supplementary Table S2). For
example, for phenylalanine, the codon TTT is used 282 times while the codon TTC is used only 49 times. In addition, the most frequently used codons are A + T rich. In *T. boliviana*, the four most frequently used codons were TTA (Leu, 9.41%), ATT (Ile, 9.40%), TTT (Phe, 7.62%), and ATA (Met, 6.48%). The preference for the use of A + T codons is a common trend for heteropteran species [31], including triatomines species [13,28]. The length of the tRNAs ranged from 62 to 72 bp (Supplementary Figure S1). All tRNAs could fold into the typical secondary structure, except for the tRNA-Ser (AGN), which lacks a stable stem–loop structure in the DHU arm.

In addition to the control region, eight other noncoding intergenic spacers (IGS) were identified. The IGS between the tRNA-Ser (UCN) and *nd1* genes is the largest one, with 767 bp. The other IGS are shorter, ranging from 1 to 14 bp. The IGS between the tRNA-Ser (UCN) and *nd1* genes seems to be very variable among *Triatoma* species. For example, in *T. dimidiata* (314 bp), this region showed a complex organization with a 135 bp sequence tandemly repeated twice and an unidentified open reading frame (ORF) [15]. This region in *T. infestans* is shorter (94 bp) and lacks both the tandem repeats and the ORF [13]. In *T. boliviana*, this region is longer (757 bp) and contains a 152 bp sequence tandemly repeated four times and an incomplete copy of this sequence that is 129 bp in length (Supplementary Figure S2). A search in the sequence database did not find any sequence with significant similarity with this 152 bp sequence.

The control region is the most variable region in length and organization between Reduviidae species [28,32]. The control region of the *T. boliviana* mitogenome is 1824 bp in length and, as well as other triatomines, contains four different parts. The first region (445 bp), the G + C rich region, is located after the *lrRNA* gene and ends in a sequence of 13 guanines.

**Figure 1.** Graphical map of the mitogenome of *Triatoma boliviana* plotted with Geneious 10.1.3. The transcriptional direction is indicated with arrows. PCGs are depicted in blue, tRNAs in orange, rRNAs in red and noncoding intergenic spacers in black.
This region seems to be conserved among triatomine species (Supplementary Figure S3). Comparisons of the G + C rich region of *T. boliviana* with all available homologous sequences (included in Figure 2) showed that the sequence similarities with North and South Triatomini species range from 69.1 to 74.3%, whereas with the two Rhodniini species range from 63.2 to 63.8%. The following 237 bp constitute the A + T rich region. Further downstream in the *T. boliviana* mitogenome is the repeat region. In all analyzed Triatoma mitogenomes, this region is formed by one or two sequences organized in tandem, with a high similarity between the different copies of the repeated sequence [28]. In *T. boliviana*, this region is less organized, with a tandem array of a 201 bp sequence (Supplementary Figure S4). However, only the first and third copies of the array are complete. The second one is internally deleted and the fourth copy is incomplete and presents only the first 68 bp of the 201 bp sequence. The fourth part of the control region, with 87 bp, contains two inverted repeats of 10 bp with the potential to form a stem and loop structure with a perfect match (Supplementary Figure S5). This structure seems to be conserved in mitogenomes and has been related with the replication mechanism [33].

Figure 2. Maximum likelihood phylogenetic tree obtained from the complete mitogenomes alignment. All heteropteran species retrieved from a BLAST search on NCBI were included. Triatominae is shown as a clade, depicted in the blue box. *Triatoma boliviana* is in red. Black circles denote bootstraps support above 95.
3.2. Phylogenetic Analyses

A maximum likelihood tree using all GenBank retrieved mitogenomes from Reduviidae recovered Triatominae as monophyletic (Figure 2), in concordance with the most recent molecular phylogenies [4]. Similar to other reports previously mentioned, Triatomini and Rhodniini differentiation is the most ancestral splitting within the subfamily. Within Triatomini, *T. boliviana* is basal to the two monophyletic lineages: North and South American Triatomini.

According to our analyses using complete mitogenomes, the closest reduviid subfamily to Triatominae would be Stenopodainae species: *Oncocephalus breviscutum* and *Canthesancus helluo*. Previous phylogenetic comparisons using morphological characters and molecular sequences showed a close relationship between Triatominae and the Stenopodainae and Reduviinae (*Zelurus clade*) subfamilies [3,4,7,11,34–36]. Unfortunately, no *Zelurus* mitogenome is available to be included in the present analysis. Hence, it was impossible for us to test its relationship with the Triatominae species.

The 16S rDNA gene pairwise genetic distances, measured using the corrected Kimura 2-parameters model, showed more than 7% divergence among *T. boliviana* and the other three analyzed species of the dispar lineage. Similarly, this distance was similar among all the dispar species compared between each other (Table 1; Figure 3), supporting the species status of *T. boliviana*. The genetic distances were greater when species from the dispar lineage were compared with other lineages within Triatomini (~14% and 13% with South and North American Triatomini, respectively) or Rhodniini (~17%) tribes (Figure 3; Table 1).
Table 1. Summary of calculated genetic distances (corrected Kimura-2p) of the 16S ribosomal DNA fragment (548 bp) among *Triatoma boliviana* and other triatomine species.

<table>
<thead>
<tr>
<th></th>
<th><em>T. boliviana</em></th>
<th><em>Triatoma sp.</em> JQ897849</th>
<th><em>T. venosa</em> JQ897850</th>
<th>South Am. Triatomini</th>
<th>North Am. Triatomini</th>
<th>Rhodniini Tribe</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>T. boliviana</em></td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Triatoma sp.</em> JQ897849</td>
<td>7.549</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>T. venosa</em> JQ897850</td>
<td>7.862</td>
<td>7.185</td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Cytogenetic data also support the inclusion of *T. boliviana* within the dispar lineage. All studied dispar species, including *T. venosa*, *T. carrioni*, *T. dispar* and *T. boliviana*, showed that the euchromatic X chromosome is slightly larger than the heterochromatic Y chromosome [37–39]. This feature is distinctive of this group and was not observed in the remaining 58 species of *Triatoma* studied so far [39].

In conclusion, the molecular analyses presented here, together with the chromosomal evidence already published, support the inclusion of *T. boliviana* within the dispar lineage.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/arthropoda1010002/s1, Figure S1: Secondary structure of tRNAs in *Triatoma boliviana* mitogenome; Figure S2: Sequence of the intergenic spacer located between tRNA-Ser (UCN) and nd1 genes; Figure S3: Alignment of the G + C rich regions of the Triatominae species included in Figure 2; Figure S4: Alignment of the repeat sequences located in the control region; Figure S5: Stem–loop structure on last part of the control region; Table S1: Annotation of the complete mitogenome of *Triatoma boliviana*; Table S2: Codon usage of *Triatoma boliviana* mitogenome protein coding genes.

**Author Contributions:** Conceptualization, S.P. and F.P.; formal analysis, S.P., P.M. and P.L.; resources, M.R.-C.; writing—original draft preparation, S.P., P.L. and F.P.; writing—review and editing, S.P., P.M., M.R.-C., T.P., P.L. and F.P.; supervision, P.L. and F.P.; funding acquisition, S.P., P.L., T.P. and F.P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by “Comisión Sectorial de Investigación Científica” (CSIC-Udelar, project grant no. 160), “Ministerio de Educación y Cultura, Dirección Nacional de Innovación, Ciencia y Tecnología, Fondo Vaz Ferreira” (II/FVF/2019/054), “Programa de Desarrollo de las Ciencias Básicas” (FEDECIIBA), “Agencia Nacional de Investigación e Innovación” (ANII) from Uruguay; the Spanish Junta de Andalucía through the “Programa Operativo FEDER Andalucía 2014–2020” and the Program of Academic Mobility of AUIP (Ibero-American University Postgraduate Association: AUIP) from Spain.

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** The *T. boliviana* mitogenome sequence was submitted to NCBI (Acc. number OM830311).

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**


34. Weirauch, C. Cladistic analysis of Reduviidae (Heteroptera: Cimicomorpha) based on morphological characters. Syst. Entomol. 2008, 33, 229–274. [CrossRef]


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.