We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

4,200
Open access books available

116,000
International authors and editors

125M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
Chapter 6

Mitochondrial Genomes of Lepidopteran Insects Considered Crop Pests

Viviana Ramírez-Ríos, Javier Correa Alvarez and Diego Villanueva-Mejia

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.71158

© 2016 The Author(s). Licensee InTech. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

In this chapter, the complete mitochondrial genome of Guatemalan potato moth, *Tecia solanivora* (Povolny, 1973) (Lepidoptera: Gelechiidae) is presented as a model to understand how to characterize and study a mitogenome in insects. It was sequenced, analyzed, and compared with other lepidopteran insects. *T. solanivora* mitogenome is a circular double-stranded molecule, typically found in insects and containing 37 genes, all them well described over the other lepidopteran mitogenomes sequenced. Interestingly, in this mitogenome was found a gene arrangement in the tRNA-Met gene different from the ancestral arrangement, but commonly present in insect mitogenomes. Other important characteristics are the high A + T-biased and negative AT- and GC-skews contents, but also unusual canonical start codons in 12 protein-coding genes and an incomplete stop codon in the cytochrome oxidase subunit II gene consisting of just a Thymine. Another common feature shared with lepidopteran mitogenomes is the A + T-rich region. It is characterized by having 325 bp, the ‘ATAGA’ motif, a 17 bp poly (T) stretch and a (AT) element preceded by the ‘ATTTA’ motif. Likewise, this mitogenome has 21 intergenic spacer regions. In addition, an update about other recent mitogenomes research done mainly over lepidopteran insects considered crop pests is presented. On the other hand, a novel development based on induced mutations by CRISPR-Cas9 in the mitogenomes seeking applicable capability for pest control is shown. The utility of this study is to improve scientific databases and support future studies of population genetic in lepidopteran.

Keywords: mitogenomes, mitochondrial genome, crop pests, lepidopteran, insects

1. Introduction

Crop loss is a function of one or more biotic factors, each of which may be contributing to a reduction in yield, whereas yield loss is the reduction in yield caused by a single pathogen or a
Figure 1. Growing in the number of publications with the words “Mitogenome AND Lepidoptera” at the title, abstract or keywords of scientific articles. Records were subtracted from literature databases; Scopus, Web of Science, and ScienceDirect.
approximately 500 mitogenomes of insects have been determined and subsequently deposited in GenBank [6]. Surprisingly, one of the most recent reports shows that only 140 complete Lepidoptera mitogenomes (28 families from 12 superfamilies) have been sequenced and deposited in genomes databases [7], which contrasts with the number of described species in this order, as previously mentioned. In this perspective, it has been the growing research efforts of scientists around the world seeking to expand the knowledge barrier of one mitochondria of Lepidoptera. Figure 1 shows continuous growth in the number of scientific publications in this field, showing records subtracted from literature databases: Scopus, Web of Science, and ScienceDirect.

In the present chapter, the complete mitogenome of *Tecia solanivora* is presented as a model to understand how to study a mitochondrial genome in insects. Additionally, it presented a review about mitogenomes research done mainly over lepidopteran insects considered crop pests to provide insights into aspects like genome structure and organization, nucleotide composition, codon usage, molecular functions, interactions among genes, and notable noncoding sequences included in the A + T-rich region. The utility of this study is to improve databases and support the determination of lepidopteran population genetic studies in the future.

### 2. Mitochondrial genome in insects

In insects, the mitochondrial genome is a circular double-stranded molecule typically between 14,000 and 20,000 bp. It contains 13 PCGs, 2 rRNAs, 22 tRNAs, and a control region (also known as the A + T-rich region), which are organized and oriented in different ways [8]. This genome has been widely used for phylogeny studies, phylogeography, population genetics, and molecular diagnostics. It has also been used to identify novel genes relevant for future studies [9], because of its small size, maternal inheritance, low recombination rate, relatively rapid evolutionary frequency, and multiple copies per cell [10]. Consequently, mitogenome sequences are rapidly evolving with about 500 insect species currently sequenced [6].

### 3. Characterization of insect mitogenomes

The complete mitogenome of *Tecia solanivora* is presented as a model to understand how to study and characterize a mitochondrial genome in insects. Additionally, it presented a review about mitogenomes research done mainly over lepidopteran insects considered crop pests. The argument for presenting *T. solanivora* as a model is because this lepidopteran insect represents the most damaging potato (*Solanum tuberosum*) pest in both Central and South America and Spain [11, 12]. *T. solanivora* was reported first time in Central America in 1956, affecting potato crops (*S. tuberosum*), which resulted in a direct effect on the economy. Even though this pest has a reduced mobility, it has invaded several countries in Central and South America as well as the Canary Islands in Spain where potato is grown [13]. It is important to mention that *T. solanivora* has been producing damage in both field crops and stored potato tubers, causing economic losses ranged from 50 to 100% [14]. The economic impact of the pest in countries of
the Andean area is much more serious than in Central America, mainly because potato is an important family staple and its production is very intensive. Therefore, T. solanivora is considered the most damaging crop insect pest in such countries [15]. Nevertheless, as an insect belonging to Lepidoptera order, the study carried out on the characterization of T. solanivora mitogenome could be applied to other studies on insects considered pest for agriculture.

3.1. Genome sequencing and assembling

T. solanivora larvae were collected from field- or storage-infested potato tubers from Colombia. All biological samples were preserved in 70% ethanol and stored at −70°C until DNA extraction. Whole T. solanivora genomic DNA was extracted using the protocol described by [13]. Each sample was analyzed by electrophoresis and the DNA concentration was quantified using a NanoDrop 2000 (Thermo Scientific, Wilmington, DE, USA). The whole genome was sequenced by Illumina Hiseq 2000 system at Chapel Hill High-Throughput Sequencing Facility in the University of North Carolina. The sequencing system generated 100 bp paired-end reads with a 342 bp insert size, these reads were checked and filtered using a homemade quality criterion script and finally the paired-end reads were assembled using de novo assembler VELVET 1.2.10 [16] with an optimized k-mer parameter of 99. The mtDNA was identified through a comparison between Tecia solanivora scaffolds and the mtDNA sequences reported in the NCBI GenBank, resulting in the identification of a single mtDNA contig with 200× coverage. To verify the topology of the mtDNA, the reads that mapped to the borders of the contig were located at an expected distance from their respective pairs. Any discrepancies that occurred, especially in the homopolymer regions, were manually edited.

3.2. Gene annotation and compositional analysis

To predict the protein-coding genes (PCGs), rRNA genes and tRNA genes from T. solanivora mtDNA, their sequence was submitted to the automatic annotator of mitochondrial genes online [dual organellar genome annotation (DOGMA), http://dogma.ccbb.utexas.edu] [17]. To determine homology between T. solanivora genes and other previously sequenced Lepidoptera species was used NCBI BLAST program and results manually curated. Then, the PCGs nucleotide composition, genome, and codon position were determined and the PCGs were translated into putative proteins for calculating of the relative synonymous codon usage (RSCU) using the invertebrate mitochondrial genetic code in MEGA version 5.2.2 [18]. The frequencies of A, T, G, and C were used to calculate the composition skew according to the AT- and GC- skew formulas. The intergenic and overlap sequences were pulled out manually from genome using SeqBuilder from the DNAStar package (DNAStar Inc., Madison, Wisconsin, USA).

3.3. Genome structure, organization, and base composition obtained

The T. solanivora mitogenome obtained was a closed circular 15,251-bp molecule (GenBank accession number KT326187) (Figure 2). It contains the typical set of 37 genes (13 PCGs, 22 tRNAs, and two rRNAs) and a large, 325-bp noncoding region (control region). A total of 24 genes were transcribed on the majority-coding strand (H-strand), while the rest were transcribed on the minority-coding strand (L-strand) (Table 1).
When we compared with other reported Lepidoptera family mitogenomes, it found an identical gene order and orientation of the mitochondrial genes of this species to other lepidopteran moths, including *Tryporyza incertulas* [19], *Corcyra cephalonica* [20], *Adoxophyes honmai* [21], *Apocheima cinerarius* [22], *Amata emma* [23], *Attacus atlas* [24], *Bombyx mori* [25], *Caligula boisduvalii* [26], *Chilo auricilius* [19], *Diaphania pyloalis* [27], *Manduca sexta* [9], *Ostrinia nubilalis*, *Ostrinia furnacalis* [28], *Samia Cynthia ricini* [29], and *Sasakia funebris* [30], among others.

Figure 2. Map of the mitochondrial genome of *T. solanivora*. Protein-coding genes (names with underline) coded on the majority strand arrows going in clockwise direction, while the rest going counterclockwise. The tRNA genes are designated by tRNA-amino acid codes. The tRNAs two and they are located next to tRNA-val and the A + T-rich region (control region) is indicated by a square.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Direction</th>
<th>Position (bp)</th>
<th>Length (bp)</th>
<th>Anticodon</th>
<th>Start codon</th>
<th>Stop codon</th>
</tr>
</thead>
<tbody>
<tr>
<td>tRNA-Met</td>
<td>Forward</td>
<td>1–68</td>
<td>68</td>
<td>CAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tRNA-Ile</td>
<td>Forward</td>
<td>70–134</td>
<td>65</td>
<td>GAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tRNA-Gln</td>
<td>Reverse</td>
<td>136–204</td>
<td>69</td>
<td>TTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAD2</td>
<td>Forward</td>
<td>259–1269</td>
<td>1011</td>
<td>ATT</td>
<td>TAA</td>
<td></td>
</tr>
<tr>
<td>tRNA-Trp</td>
<td>Forward</td>
<td>1268–1336</td>
<td>69</td>
<td>TCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tRNA-Cys</td>
<td>Reverse</td>
<td>1329–1394</td>
<td>66</td>
<td>GCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tRNA-Tyr</td>
<td>Reverse</td>
<td>1406–1471</td>
<td>66</td>
<td>GTA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COI</td>
<td>Forward</td>
<td>1475–3010</td>
<td>1536</td>
<td>CGA</td>
<td>TAA</td>
<td></td>
</tr>
<tr>
<td>tRNA-Leu (UUR)</td>
<td>Forward</td>
<td>3006–3073</td>
<td>68</td>
<td>TAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COI</td>
<td>Forward</td>
<td>3074–3754</td>
<td>681</td>
<td>ATG</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>tRNA-Lys</td>
<td>Forward</td>
<td>3756–3826</td>
<td>71</td>
<td>CTT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tRNA-Asp</td>
<td>Forward</td>
<td>3837–3904</td>
<td>68</td>
<td>GTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATP8</td>
<td>Forward</td>
<td>3905–4072</td>
<td>168</td>
<td>ATT</td>
<td>TAA</td>
<td></td>
</tr>
<tr>
<td>ATP6</td>
<td>Forward</td>
<td>4066–4743</td>
<td>678</td>
<td>ATG</td>
<td>TAA</td>
<td></td>
</tr>
<tr>
<td>COIII</td>
<td>Forward</td>
<td>4743–5531</td>
<td>789</td>
<td>ATG</td>
<td>TAA</td>
<td></td>
</tr>
<tr>
<td>tRNA-Gly</td>
<td>Forward</td>
<td>5534–5600</td>
<td>67</td>
<td>TCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAD3</td>
<td>Forward</td>
<td>5601–5954</td>
<td>354</td>
<td>ATT</td>
<td>TAA</td>
<td></td>
</tr>
<tr>
<td>tRNA-Ala</td>
<td>Forward</td>
<td>5964–6030</td>
<td>67</td>
<td>TGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tRNA-Arg</td>
<td>Forward</td>
<td>6030–6095</td>
<td>66</td>
<td>TCG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tRNA-Asn</td>
<td>Forward</td>
<td>6101–6166</td>
<td>66</td>
<td>GTT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tRNA-Asn (AGN)</td>
<td>Forward</td>
<td>6181–6246</td>
<td>66</td>
<td>GCT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tRNA-Glu</td>
<td>Forward</td>
<td>6247–6315</td>
<td>69</td>
<td>TTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tRNA-Pho</td>
<td>Reverse</td>
<td>6314–6380</td>
<td>67</td>
<td>GAA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAD5</td>
<td>Reverse</td>
<td>6364–8097</td>
<td>1734</td>
<td>ATT</td>
<td>TAA</td>
<td></td>
</tr>
<tr>
<td>tRNA-His</td>
<td>Reverse</td>
<td>8113–8178</td>
<td>66</td>
<td>GTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAD4</td>
<td>Reverse</td>
<td>8183–9523</td>
<td>1341</td>
<td>ATG</td>
<td>TAA</td>
<td></td>
</tr>
<tr>
<td>NAD4L</td>
<td>Reverse</td>
<td>9523–9816</td>
<td>294</td>
<td>ATG</td>
<td>TAA</td>
<td></td>
</tr>
<tr>
<td>tRNA-Thr</td>
<td>Forward</td>
<td>9819–9883</td>
<td>65</td>
<td>TGT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tRNA-Pro</td>
<td>Reverse</td>
<td>9884–9949</td>
<td>66</td>
<td>TGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAD6</td>
<td>Forward</td>
<td>9952–10,479</td>
<td>525</td>
<td>ATA</td>
<td>TAA</td>
<td></td>
</tr>
<tr>
<td>Cytb</td>
<td>Forward</td>
<td>10,497–11,642</td>
<td>1146</td>
<td>ATA</td>
<td>TAA</td>
<td></td>
</tr>
<tr>
<td>tRNA-Ser</td>
<td>Forward</td>
<td>11,646–11,712</td>
<td>67</td>
<td>TGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NAD1</td>
<td>Reverse</td>
<td>11,730–12,665</td>
<td>936</td>
<td>ATA</td>
<td>TAA</td>
<td>TAG</td>
</tr>
<tr>
<td>tRNA-Leu</td>
<td>Reverse</td>
<td>12,669–12,736</td>
<td>68</td>
<td>TAG</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The typical lepidopteran arrangement of the tRNAs (tRNA-Met, tRNA-Ile, tRNA-Gln) was observed in the *T. solanivora* mitogenome but differs from the order found in ancient insects (Figure 3). In that sense, it was determined that *T. solanivora* presents several differences from the ancestral organization of the tRNA-Met region (A + T-rich region, tRNA-Ile, tRNA-Gln, tRNA-Met) [20], which is also found in the mitogenomes of *Aedes aegypti* (Diptera) [31] and *Acrida cinerea* (Orthoptera) [32]. In the case of *T. solanivora*, the order is: A + T region, tRNA-Met, tRNA-Ile, tRNA-Gln. Additionally, in the *T. solanivora* mitogenome, the tRNA-Lys gene is found after the COII gene, contrary to *A. cinerea*, where they are found in the reverse order. In addition, the NAD3 gene was located before the tRNA-Ala gene in *T. solanivora*, whereas in *A. aegypti*, the gene located in this region is tRNA-Arg.

The nucleotide composition determined in the entire *T. solanivora* mitogenome was A: 38.6, T: 39.6, C: 13.3, and G: 8.4% (Table 2). This nucleotide composition shows that highly A + T-biased (78.2%) with a similar proportion of adenine (A) and thymine (T) compared with the reported ranges found in other Lepidoptera mitogenomes. In the same way, *T. solanivora* mitogenome exhibits negative AT-skew (−0.013) and GC-skew (−0.226) values (Table 3). However, the most of Lepidoptera family members have shown higher percentages of A than T, such as: *O. nubilalis* (A: 41.3, T: 38.8%), *O. furnacalis* (A: 41.46, T: 38.92%), *B. mori* (A: 43.06, T: 38.30%) [33], *Phthonandria atrilineata* (A: 40.78, T: 40.24%) [34], *Ochrogaster lunifer* (A: 40.09, T: 37.75%) [35], *Chinese Bombyx mandarina* (A: 43.11, T: 38.48%) [36], and *A. atlas* (A: 39.8, T: 39.5%), among others. Likewise, the cytosine (C) content in the *T. solanivora* mitogenome was greater than guanine (G), which is similar to the percentages identified in other recently discovered Lepidoptera mitogenomes, except for *Antheraea yamamai* (G: 10.71, C: 10.35%) [37], *Eriogyna pyretorum* (G: 10.61, C: 7.45), and *Artogeia melete* (G: 11.33, C: 8.65%) [38].

### 3.4. Protein-coding genes (PCGs)

The protein-coding genes (PCGs) encompassed 11,191 bp of the entire assembled sequence (73.38%) and exhibited an A + T content of 76.4%. Nine of the 13 PCGs are coded on the majority strand (ATP6, ATP8, COI, COII, COIII, Cytb, NAD2, NAD3, and NAD6), while the rest (NAD1, NAD4, NAD4L, and NAD5) are coded on the minority strand. For the protein-coding genes, the A + T content was calculated for the three-codon positions, and they showed few differences from other Lepidoptera mitogenomes. *T. solanivora* showed negative AT- and GC-skew values in the second and third positions of each codon, indicating a greater inclination for the nitrogen bases, T and C, while the first position showed a slightly positive value.
Figure 3. Gene arrangement of the *T. solanivora* mitogenome. Protein-coding genes are marked by light gray, ribosomal RNA genes by light green, control region by red and tRNA genes are designated by the single letter amino acid code (white). Brown box and horizontal line represent gene clusters that changed positions.

Table 2. Nucleotide composition of *T. solanivora* mitogenome.

<table>
<thead>
<tr>
<th>Species</th>
<th>Length (bp)</th>
<th>A%</th>
<th>G%</th>
<th>T%</th>
<th>C%</th>
<th>A + T%</th>
<th>G + C%</th>
<th>AT-skew</th>
<th>GC-skew</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>T. solanivora</em></td>
<td>15,251</td>
<td>38.6</td>
<td>8.4</td>
<td>39.6</td>
<td>13.3</td>
<td>78.2</td>
<td>21.7</td>
<td>-0.013</td>
<td>-0.226</td>
</tr>
<tr>
<td><em>A. selene</em></td>
<td>15,236</td>
<td>38.54</td>
<td>8.05</td>
<td>40.37</td>
<td>13.03</td>
<td>78.91</td>
<td>21.08</td>
<td>-0.023</td>
<td>-0.236</td>
</tr>
<tr>
<td><em>C. raphaelis</em></td>
<td>15,314</td>
<td>39.37</td>
<td>7.30</td>
<td>43.29</td>
<td>10.04</td>
<td>82.66</td>
<td>17.34</td>
<td>-0.047</td>
<td>-0.158</td>
</tr>
<tr>
<td><em>E. pyretorum</em></td>
<td>15,327</td>
<td>39.17</td>
<td>7.63</td>
<td>41.65</td>
<td>11.55</td>
<td>80.82</td>
<td>19.18</td>
<td>-0.031</td>
<td>-0.204</td>
</tr>
<tr>
<td><em>A. yamamai</em></td>
<td>15,338</td>
<td>39.26</td>
<td>7.69</td>
<td>41.04</td>
<td>12.02</td>
<td>80.30</td>
<td>19.71</td>
<td>-0.022</td>
<td>-0.220</td>
</tr>
<tr>
<td><em>C. boisduvalii</em></td>
<td>15,360</td>
<td>39.34</td>
<td>7.58</td>
<td>41.28</td>
<td>11.79</td>
<td>80.62</td>
<td>19.37</td>
<td>-0.024</td>
<td>-0.217</td>
</tr>
<tr>
<td><em>S. cynthia ricini</em></td>
<td>15,384</td>
<td>39.65</td>
<td>7.81</td>
<td>40.13</td>
<td>12.41</td>
<td>79.78</td>
<td>20.22</td>
<td>-0.006</td>
<td>-0.227</td>
</tr>
<tr>
<td><em>M. sexta</em></td>
<td>15,516</td>
<td>40.67</td>
<td>7.46</td>
<td>41.11</td>
<td>10.76</td>
<td>81.78</td>
<td>18.22</td>
<td>-0.005</td>
<td>-0.181</td>
</tr>
<tr>
<td><em>A. perryi</em></td>
<td>15,566</td>
<td>39.22</td>
<td>7.77</td>
<td>40.94</td>
<td>12.07</td>
<td>80.16</td>
<td>19.84</td>
<td>-0.021</td>
<td>-0.217</td>
</tr>
<tr>
<td><em>A. honmii</em></td>
<td>15,680</td>
<td>40.15</td>
<td>7.88</td>
<td>40.24</td>
<td>11.73</td>
<td>80.39</td>
<td>19.61</td>
<td>-0.001</td>
<td>-0.196</td>
</tr>
</tbody>
</table>

Table 3. Comparison of nucleotide composition and skewness between *T. solanivora* and other lepidopteran mitogenomes.
for the GC-skew indicating a greater bias for G than C (Table 3). In general, the codons of *T. solanivora* mitogenome present high A + T content for the first position (71.9%). Similar values were observed in *A. emma* (73.1%), *Antheraea pernyi* (72.9%) [39], *C. boisduvalii* (73.8%), *B. mandarina* (75.0%) [33], and *M. sexta* (74.8%), but in the second position was found a lower A + T percentage (69.5%).

Twelve PCGs were identified in the *T. solanivora* mitogenome with the typical ATN initiation codons (isoleucine and methionine), except for the COI gene that is initiated by CGA initiation codon (arginine). The typical ATN codon represents a putative codon commonly observed in the order Lepidoptera [9, 40], and is thus considered a synapomorphy of this group of insects [41]. The methionine start codon, ATG, was used by five of the 13 PCGs, and ATA was used to initiate protein synthesis in the *NAD1, NAD6,* and *cytb* genes. In contrast, an atypical isoleucine codon (ATT) was used to initiate protein synthesis in the *ATP8, NAD2, NAD3,* and *NAD5* genes. Arginine (CGA) was used for the COI gene for which a nucleotide sequence of four or six base pairs has been proposed, much like TTAG in *Maruca vitrata* [42], to serve in a nonstandard initiation process located immediately upstream from the putative arginine CGA start codon of COI. In *T. solanivora*, this tetrnucleotide sequence consisted of TTGG. The high A + T percentages in insect mitogenomes result in high probabilities of finding a noncoding triplet or a coding triplet within the tRNA-Tyr gene, a result that could potentially produce generalized annotation errors for the gene COI [43]. Previous studies have discussed the possibility that translation initiation of this gene involves an unusual sequence of four to six nucleotides (ATAA, TTAA, GTAA, ATTA, or ATTTAA) located immediately before the coding primer of the COI gene. This sequence apparently functions as the translation initiator in the majority of insects from the family Diptera, including *Drosophila yakuba* [44]. However, in *T. solanivora*, the sequence TTGG was found immediately before putative initiation codon CGA.

For the stop codon genes, we found the TAA codon in 11 of the PCGs, coinciding with the mitogenomes of other Lepidoptera, including *T. incertulas, S. funebris, S. cynthia,* and *A. emma* of the family Hesperiidae. In the *NAD1* gene, the TAG stop codon was found and similar results were obtained in five species of the family Hesperiidae [41], while the *COII* gene used a single T as an incomplete stop codon, which is commonly found in the majority of Lepidoptera species to date [9, 41]. This truncated codon could be a representative of a recognition site for an endonuclease that splits the polycistronic pre-mRNA, where a post-transcriptional polyadenylation then occurs, resulting in a functional stop codon (TAA) [9, 19, 45].

The CDpT or Codons Per Thousand Codons of the *T. solanivora* mitogenome was calculated, and five amino acid families were identified. The most common families were: phenylalanine (Phe), asparagine (Asn), isoleucine (Ille), lysine (Lys), and leucine 2 (Leu 2), as shown in (Figure 4a), being in *T. solanivora*, phenylalanine (Phe) the most abundant, instead of Leucine 2 (Leu2), which dominates in other Lepidoptera mitogenomes [35]. Additionally, when the relative synonymous codon use (RSCU) was determined, we identified that *T. solanivora* presents all typical codons found in other invertebrates (Figure 4b). The codons were richer in A or T at the third position and consequently have less G or C.
3.5. Transfer RNA and ribosomal RNA genes

It was found that *T. solanivora* contains a typical set of 22 tRNAs with a high A + T bias, accounting for 80.7% of the tRNAs, slightly positive AT-skew (0.004) and a clearly positive GC-skew (0.161). These results suggest that tRNAs exhibit a higher inclination for nitrogen bases A and G than for T and C. Similar results were reported by [46] in *E. pyretorum* (AT-skew = 0.039 and GC-skew = 0.174) and [38] in *A. melete* (AT-skew = 0.034 and GC-skew = 0.142). Among the tRNA genes, 14 are coded on the H-strand and eight on the L-strand with lengths ranged

Figure 4. Codon distribution and relative synonymous codon usage (RSCU) in *T. solanivora* mitogenome. (a) Codon distribution and (b) RSCU. Codon families are provided on the X axis and the RSCU on the Y axis. This mitogenome presents all possible codon families existing in Lepidoptera.
between 65 and 71 bp, and these genes exhibited positive AT-skew (0.047) as well as in almost all lepidopteran mitogenomes. However, the rearrangement of certain tRNAs were found translocated in the *T. solanivora* mitogenome compared with out-groups, which are described in (Figure 3).

Similar to other mitochondrial sequences from insect species, there were two rRNAs in *T. solanivora* with a total length of 2099 bp and an AT content of 83.7% (Table 2). The large ribosomal gene (rRNA-Large), located between tRNA-Leu1 and tRNA-Val, has a length of 1329 bp, whereas the small gene (rRNA-Small), located between tRNA-Val and the A + T-rich region, has a length of 770 bp (Table 1). These rRNA lengths were within the range of values reported for other Lepidoptera, as their values range between 1314 bp in *Euploea mulciber* (Nymphalidae) [47] and 1330 bp in *Coreana raphaelis* (Lycaenidae) [48]. For the rRNA-Large and rRNA-Small, the length ranges from 739 bp for *Protantigius superans* (Lycaenidae) to 788 bp in (Nymphalidae) [49]. The rRNAs in *T. solanivora* have an A + T content of 83.7%, and similar values were reported for other Lepidoptera, including *C. cephalonica* (80.43%), *T. incertulas* (82.8%), and *Dichocrocis punctiferalis* (85.1%) [50].

### 3.6. Noncoding and overlapping regions

Most of the intergenic regions in this mitogenome were short (≤15 bp) and the total length of the noncoding regions in the mtDNA of *T. solanivora* was 199 bp. This region is composed by 21 intergenic spacer sequences, ranging from 1 to 54 bp and showed highly A + T-biased (88.8%) (Table 2). The intergenic spacers longer were denominated S1, S2, S3, and S4. Intergenic sequence S1 is commonly found in Lepidoptera mitogenome order between the tRNA-Gln and NAD2 genes and length ranges between 38 pb in *T. incertulas* and 88 pb in *Sasakia charonda* [50]. However, this region has not been identified in insects that belong to other orders [9].

This sequence (S1) could be considered as a mitogenome marker for Lepidoptera order, and it most likely originated from a partial NAD2 gene duplication [19]. Intergenic sequence S2 (23 bp) was found between rRNA-Large and tRNA-Val. Intergenic sequences S3 and S4 (17 bp) separate genes NAD6 and Cytb, and the tRNA-Ser2 and NAD1 genes, respectively. The latter sequence contains the “ATACTAA” motif, typically found in other lepidopterans [9, 23, 51]. This motif plays an apparent role as a recognition site for the protein implicated in mitochondrial transcription termination (mtTERM) [52]. Furthermore, this sequence has been recognized for being highly conserved, with a length ranging between 17 and 20 pb [23].

Furthermore, in the *T. solanivora* mitogenome, three principal overlap sequences were identified and were designated as OLS1, OLS2, and OLS3. OLS1 was found overlapping the tRNA-Phe and NAD5 genes. This sequence presents the greatest length, with a total of 17 bp. OLS2 was found between the tRNA-Trp and tRNA-Cys genes, with a total length of 8 bp and the 7-bp OLS3-overlapped genes ATP8 and ATP6, which are consistent with the same genes found in other lepidopterans, although they differ in length [37, 52]. In addition, an unusual overlap region (OLS1) was found between the tRNA-Phe and NAD5 genes; it is important to mention that this 17-bp region has not been reported before.
3.7. The A + T-rich region

The A + T-rich region is a noncoding region with 325 bp length located between rRNA-Small and tRNA-Met. The region contains 91.1% AT nucleotides, with negative AT- and GC-skew values (Table 2), meaning that it is biased for the nitrogen base thymine, as reported for the mitogenomes of other lepidopterans. One exception to this trend is A. honmai, which has a positive AT skew (0.028), indicating a bias for adenines [21]. The length of this region is variable in the other Lepidoptera, and it can be as long as 1270 bp, as reported in Papilio bianor (Papilionidae) [53].

This A + T region is a conserved structure commonly found in other Lepidoptera, which includes the “ATAGA” motif followed by a 17-bp poly-T stretch, just like in T. solanivora mitogenome. This motif is immediately followed by the tRNA-Met gene [35, 41, 54] and it seems it has an important role in the replication initiation in minor strand of mtDNA in addition to gene regulation [19, 30, 41]. Furthermore, eight microsatellite regions were identified within the mitogenome of T. solanivora, referred to as (TAA)$_n$ (AT)$_n$ and (TAT)$_n$. These were the most representative microsatellites found in the species, although the mononucleotide sequences, (T)$_n$ and (A)$_n$ were also identified [55]. These represent the relevant regions of this genome for future studies. Also, in most lepidopteran mitogenomes, the (AT)$_n$ microsatellite has been previously reported. This microsatellite is preceded by the “ATTTA” motif that is commonly found in other mitogenomes [9, 41]. In S. funebris, the same (AT)$_n$ microsatellite was identified as that found in T. solanivora. Nevertheless, most lepidopteran mitogenomes report (AT)$_n$, where $n$ ranges from 7 to 12 [19, 23].

3.8. Phylogenetic relationships

To illustrate the phylogenetic relationship of T. solanivora (Lepidoptera: Gelechiidae) with other 16 Lepidoptera families, we used a concatenated set of PCGs of 72 other complete Lepidoptera mitogenomes obtained from GenBank with previous elimination of start and stop codons. The phylogenetic relationship among the eight Lepidoptera superfamilies was inferred using both Bayesian Inference and maximum likelihood methods, which produced similar topologies to previously analyzed phylogenies obtained for other lepidopterans. The results obtained with both methods produced similar and consistent topologies. Our results showed high support values for the majority of the nodes and thus the interrelationships are well-resolved within order Lepidoptera. We used Aedes aegypti (Diptera) and Acrida cinerea (Orthoptera) as out-groups, the phylogenetic trees revealed nine Lepidoptera clades. Species of the Papilionoidea, Noctuidae, Bombycidae, Geometridae, Pyralidae, Gelechiidae, Tortricidae, Yponomeutidae, and Hepialidae superfamilies cluster monophyletic groups, with strongly supported bootstrapped and posterior probabilities (100). All those results and analysis were published by authors of the present chapter in 2016 [7].

4. Other recent studies with mitogenomes of Lepidopteran considered crop pests

A search carried out on August 04, 2017 in Scopus database showed that after publishing the scientific paper made by authors from this chapter [7], scientists have published 62 other
studies on mitochondrial genomes of lepidopteran insects. Most of them were focused on understanding their composition, organization, motifs, and the inference of phylogenetic relationships between these organisms [41, 56–58]. However, recently, [59] reported besides of typifying the mitogenomes of Mesophleps albilinella and Dichomeris ustalella (Lepidoptera: Gelechiidae), the prediction of the secondary structures of the tRNAs [44, 60]. In this model, single polynucleotide chain form four or five arms to fold itself with each other, like a clover leaf, call them an acceptor arm, DHU or D arm, anti-codon arm, TψC arm, and variable arm [61], which play a role in proper folding of the tRNA into the L-shaped tertiary structure while modifications in or around the anti-codon loop contribute to the function of tRNAs in decoding [62]. Prediction of secondary structures of tRNAs has been using tRNAscan-SE 1.21, Mito/Chloroplast, invertebrate genetic code for the prediction of tRNA isotypes, and a cutoff of 1 score [63]. This method allowed to find and predict the secondary structure of 21 tRNAs in both species, except for the tRNAser (AGN), which has a truncated DHU arm, with consideration given to the anti-codons [59]. Frequently, have been reported tRNA-like structures into the A + T-rich region in Lepidoptera [37, 49, 64], but it seems to be fake tRNAs of random secondary structures, owing to the reduced sequence complexity (>90% A + T) in this noncoding region [59], suggesting that are fake tRNAs of random secondary structures, owing to the reduced sequence complexity (>90% A + T) in this noncoding region [65].

On the other hand, we must highlight the importance of studying of insect pests mitogenomes, this allows to propose hypotheses related with the evolutionary origin of the different larval stages, which causes significant damage to crops during this state, and could predict which is the most adaptable state to each type of environment as for example in Parapoynx crisonalis moth (Lepidoptera: Crambidae) [66] in which its larvae are lacking tracheal gills because they are pest in aquatic crops [67]. This analysis also can be extrapolated to the study of moths that are also plagues only in larval state but in terrestrial crops such as T. solanivora, and in this way to be able to find the most viable way to control this type of pests.

5. Novel techniques for pest control using mtDNA

Pest species represent a major ongoing threat to global biodiversity, demanding effective management approaches are required that regulate pest numbers, while minimizing collateral damage to nontarget species. Species-specific pest controls have been developed in order to be long-lasting measures and effectives [68]. One of these methods is called the sterile insect technique (SIT), whereby sterile males are introduced into target populations, so that they could be produced continuously within the targeted populations for control, and thus reducing production of females when mating with them. However, the SIT generally requires continuous large-scale production and introduction of sterile evils to sustain population suppression [69].

At the level of maternally inherited mitochondrial DNA (mtDNA) has been identified naturally occurring mutations that cause male infertility. These mutations have little or no impact
on females, and hence are minimally or not selected against (i.e. are self-perpetuating in nature). Due to those kinds of mutations, have only been identified in some model systems such as mice and fruit flies, they are likely to be widespread in nature threatening small populations viability of endangered species. Currently, a novel variant of the SIT, is the recently proposed Trojan female technique (TFT), based on the use of naturally occurring mutations or induced by CRISPR-Cas9 (clustered, regularly interspaced, short palindromic repeats system) in the mtDNA [69]. The consortium aims to harness these mutations to develop a widely applicable capability for pest control, through the release of Trojan females carrying the mutations [68, 69].

With this technique, males that inherit these mutations will have fewer offspring than wild-type males, while females will remain normal (fertile). It is well known that mtDNA is generally maternally inherited, so this sex-bias in effects will reduce selection pressure against the TFT mutation. When females carrying the TFT mutation are released into a pest population, they could cause multi-generational population suppression. However, while promising well and scientific means to control pest populations or disease vectors, the release of genetically engineered animals raises into ethical issues and a debate is currently underway discussing safety and regulatory concerns [68, 69].

6. Conclusion

In this chapter, the complete mitochondrial genome of *T. solanivora* was presented as a model to understand how to characterize and study a mitogenome in insects. It was sequenced, analyzed, and compared with other lepidopteran insects. This mitogenome shares many features with those reported previously in Lepidoptera but exhibited several subtle differences in the codon distribution within the A + T region. The phylogenetic relationships of nine clades of the order Lepidoptera were developed using Bayesian and maximum likelihood inference, which provided well-supported results compared with other phylogenies based on both molecular and morphological traits. In addition, an update about other recent mitogenomes research done mainly over lepidopteran insects considered crop pests was made. On the other hand, it was shown a novel development based on induced mutations by CRISPR-Cas9 in the mitogenomes seeking applicable capability for pest control. The utility of all information presented in this chapter is to improve scientific databases and support the determination of lepidopteran population genetic studies in the future.

Author details

Viviana Ramírez-Ríos, Javier Correa Alvarez and Diego Villanueva-Mejia*

*Address all correspondence to: dvillanu@eafit.edu.co

Department of Biological Sciences, School of Sciences, Universidad EAFIT, Medellín, Antioquia, Colombia
References


[9] Cameron SL, Whiting MF. The complete mitochondrial genome of the tobacco hornworm, Manduca sexta, (Insecta: Lepidoptera: Sphingidae), and an examination of mitochondrial gene variability within butterflies and moths. Gene. 2008;408:112-123


[22] Liu S, Xue D, Cheng R, Han H. The complete mitogenome of Apocheima cinerarius (Lepidoptera: Geometridae: Ennominae) and comparison with that of other lepidopteran insects. Gene. 2014;547:136-144


[26] Hong MY, Lee EM, Jo YH, et al. Complete nucleotide sequence and organization of the mitogenome of the silk moth Caligula boisduvalii (Lepidoptera: Saturniidae) and comparison with other lepidopteran insects. Gene. 2008;413:49-57


[34] Yang L, Wei ZJ, Hong GY, Jiang ST, Wen L. The complete nucleotide sequence of the mitochondrial genome of *Phthonandria atrilineata* (Lepidoptera: Geometridae). Molecular Biology Reports. 2009;36:1441-1449


[40] Fenn JD, Cameron SL, Whiting M. The complete mitochondrial genome sequence of the Mormon cricket (*Anabrus simplex*: Tettigoniidae: Orthoptera) and an analysis of control region variability. Insect Molecular Biology. 2007;16:239-252


[53] Hou LX, Ying S, Yang XW, Yu Z, Li HM, Qin X. The complete mitochondrial genome of Papilio bianor (Lepidoptera: Papilionidae), and its phylogenetic position within Papilionidae. Mitochondrial DNA. Epub ahead of print 2014. DOI: 10.3109/19401736.2013.873923


