Gene 581 (2016) 107–116

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research Paper

Mitochondrial genome characterization of *Tecia solanivora* (Lepidoptera: Gelechiidae) and its phylogenetic relationship with other lepidopteran insects

Viviana Ramírez-Ríos [a], Nicolás D. Franco-Sierra [a], Javier Correa Alvarez [a], Clara I. Saldamando-Benjumea [b,*], Diego F. Villanueva-Mejía [a]

[a] Departamento de Ciencias Biológicas, Escuela de Ciencias, Universidad EAFIT, Medellín, Antioquia, Colombia
[b] Escuela de Biociencias, Facultad de Ciencias, Universidad Nacional de Colombia, sede Medellín, Antioquía, Colombia

ARTICLE INFO

Article history:
Received 28 July 2015
Received in revised form 13 January 2016
Accepted 15 January 2016
Available online 21 January 2016

Keywords:
Mitogenome
T. solanivora
Lepidoptera
Genes
Phylogeny

ABSTRACT

The complete mitogenome of the potato tuber moth *Tecia solanivora* (Lepidoptera: Gelechiidae) was sequenced, annotated, characterized and compared with 140 species of the order Lepidoptera. The circular genome is 15,251 bp, containing 37 genes (13 protein-coding genes (PCGs), two rRNA genes, 22 tRNA genes and an A + T-rich region). The gene arrangement was identical to other lepidopteran mitogenomes but different from the ancestral arrangement found in most insects for the *tRNA-Met* gene (A + T-region, *tRNA-I*, *tRNA-Q*, *tRNA-M*). The mitogenome of *T. solanivora* is highly A + T-biased (78.2%) and exhibits negative AT- and GC-skews. All PCGs are initiated by canonical ATN start codons, except for Cytochrome Oxidase subunit 1 (*COI*), which is initiated by CGA. Most PCGs have a complete typical stop codon (TAA). Only *NAD1* has a TAG stop codon and the *COII* and *NAD5* genes have an incomplete stop codon consisting of just a T. The A + T-rich region is 332 bp long and contains common features found in lepidopteran mitogenomes, including the 'ATAGA' motif, a 17 bp poly (T) stretch and a $(AT)_8$ element preceded by the 'ATTTA' motif. Other tandem repeats like $(TAA)_4$ and $(TAT)_7$ were found, as well as $(T)_6$ and $(A)_{10}$ mononucleotide repeat elements. Finally, this mitogenome has 20 intergenic spacer regions. The phylogenetic relationship of *T. solanivora* with 28 other lepidopteran families (12 superfamilies) showed that taxonomic classification by morphological features coincides with the inferred phylogeny. Thus, the Gelechiidae family represents a monophyletic group, suggesting that *T. solanivora* and *Pectinophora gossypiella* have a recent common ancestor.

1. Introduction

In insects, the mitochondrial genome is a circular double-stranded molecule typically between 14,000 and 20,000 bp. It contains 13 PCGs, two rRNAs, 22 tRNAs and the A + T-rich region (also annotated as D-loop or control region of vertebrate mtDNAs, with apparently homologous function (Wolstenholme, 1992; Stewart and Beckenbach, 2009)), which are organized and oriented in different ways (Boore, 1999). The mitogenome has been widely used for phylogeny studies, phylogeography, population genetics and molecular diagnostics. It has also been used to identify novel genes and molecular markers (Cameron and Whiting, 2008), because of its small size, low recombination rate, relatively rapid evolutionary rate and multiple copies per cell (Avise, 1994), and almost strict maternal inheritance (Meusel and Moritz, 1993; Pitnick and Karr, 1998). Consequently, mitogenome sequences are rapidly increasing, with about 500 insect species currently sequenced (Cameron, 2014).

Gelechiidae family comprises close to 4700 species and more than 760 have been described. Around 258 species are pest of crops or stored food products, including well-known species such as pink bollworm (*Pectinophora gossypiella*), tomato leafminer (*Tuta absoluta*), between other (Zhang, 1994). This family has a worldwide distribution and is among one of the most diverse Lepidoptera fauna in many regions and habitats (Karsholt et al., 2013). Members of this family are small to medium sized, often gray or brown and larvae have a wide range of feeding strategies (Karsholt et al., 2013). *Tecia solanivora* (Lepidoptera: Gelechiidae) represents the most damaging potato (*Solanum tuberosum*) pest in both Central and South America, and Spain (Pulliandre et al., 2008; Torres-Leguizamón et al., 2011). The larvae of this insect attack potato tubers, causing losses between 50 and 100% of potato crops (Zeddam et al., 2008). Most of the studies in *T. solanivora* have focused on improving pest management. However, little is known about its evolutionary biology. Torres-Leguizamón et al. (2011) established the migratory pattern of the species from sequences of the mitochondrial gene *Cytb*, and argued that

Abbreviations: A, adenosine; G, guanosine; C, cytidine; T, thymidine; Nt, nucleotide; rRNA, ribosomal RNA; tRNA, transfer RNA; mtDNA, mitochondrial DNA.

* Corresponding author.
E-mail address: cisaldam@unal.edu.co (C.I. Saldamando-Benjumea).

http://dx.doi.org/10.1016/j.gene.2016.01.031
0378-1119/

the center of origin of this insect was Guatemala. The most recent study based on the population genetics of this moth was performed by Villanueva-Mejía et al. (2015), in Colombia, where the authors employed nine microsatellites in 120 individuals collected in four departments of this country (Antioquia (North), Boyacá (Center), Nariño (South) and Norte de Santander (East)). They found that this species is genetically structured, as populations from eastern Colombia (Norte de Santander) are differentiated from the rest. This differentiation has occurred because Norte de Santander represents the first location in this country invaded by this pest and also because the movement of potato tubers from this region to the rest of Colombia is not typical, as this crop is produced mainly for domestic consumption (Villanueva-Mejía et al., 2015). In contrast, potatoes produced in Boyacá are transferred all over the country; genetically homogenizing the pest populations by transferring infested potatoes all over the country.

Recently, Karsholt et al. (2013), reexamined the higher-level phylogeny of this family based on DNA sequence data for one mitochondrial gene (*cytochrome c oxidase subunit I*) and seven nuclear genes (*Elongation Factor-1α*, *wingless*, *Ribosomal protein S5*, *Isocitrate dehydrogenase*, *Cytosolic malate dehydrogenase*, *Glyceraldehyde-3-phosphate dehydrogenase* and *Carbamoylphosphate synthase domain protein*). They concluded that this family displays a wide array of life history strategies but diversity patterns only correlated below the subfamily level suggesting multiple origins of its members or reversals of traits such as internal/external feeding and leaf mining.

In this study we characterized the complete mitogenome of *T. solanivora* and we compared it with other Lepidoptera mitogenomes, including several aspects like: genome structure and organization, nucleotide composition, codon usage, molecular functions, interactions among genes, and notable non-coding sequences included in the A + T-rich region. Phylogenetic analyses were carried out with 140 complete Lepidoptera mitogenomes (28 families from 12 superfamilies) to elucidate the intra- and interordinal molecular relationships, employing maximum likelihood and Bayesian Inference. The utility of this study is to improve the databases of this important potato pest, as this represents the first study of the complete mitogenome of this species.

2. Material and methods

2.1. Insect collection and DNA extraction

Larvae of *T. solanivora* were collected from field or storage infested potato tubers from the Nariño department in Colombia during 2011 and 2012. Samples were preserved in 70% ethanol and stored at −70 °C until DNA extraction. Whole genomic DNA of *T. solanivora* was extracted using the protocol described by Villanueva-Mejía et al. (2015). Each sample was analyzed by electrophoresis and the DNA concentration was quantified using a NanoDrop 2000 (Thermo Scientific, Wilmington, DE, USA).

2.2. Genome sequencing and assembling

Total DNA was sequenced on an Illumina Hiseq 2000 system at the University of North Carolina at Chapel Hill High-Throughput Sequencing

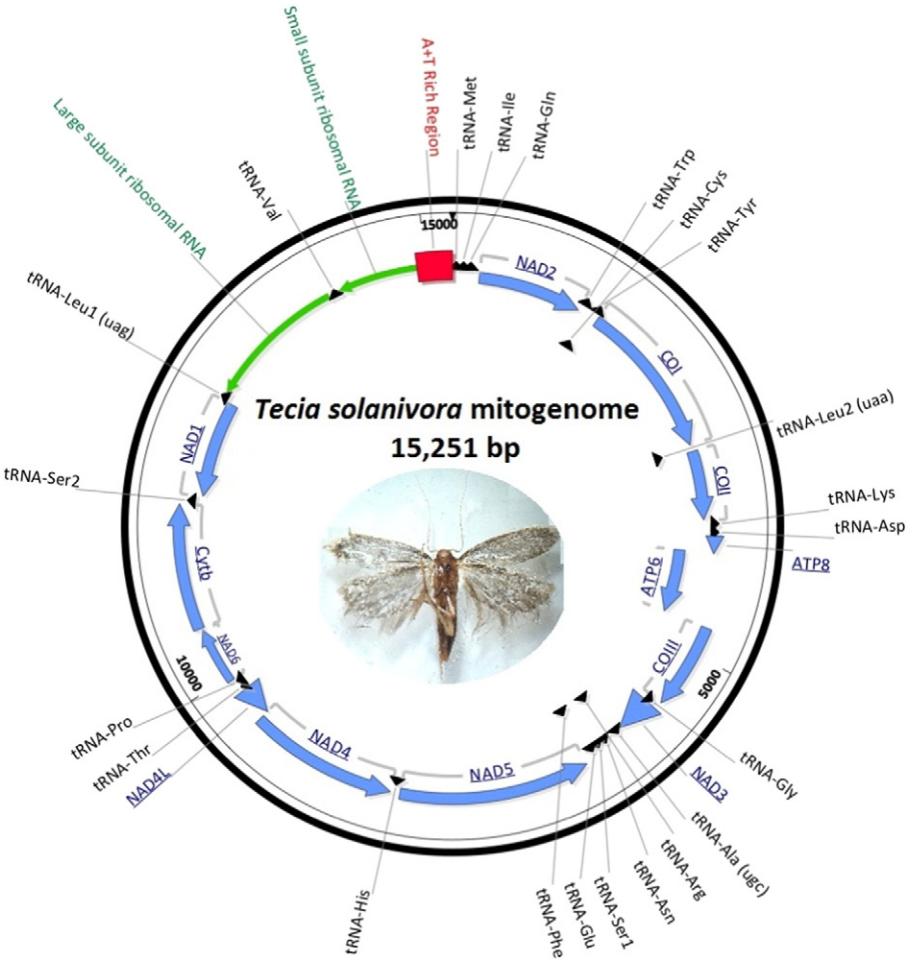

Fig. 1. Map of the mitochondrial genome of *T. solanivora*. Blue arrows represent protein-coding genes (underlined blue names) coded on the heavy strand (clockwise arrows) and light strand (counterclockwise arrows). The tRNA genes are designated by black letter tRNA-amino acid codes. The rRNAs are labeled in green, and the A + T-rich region is labeled in red.

Facility. The whole-genome shotgun strategy was used to produce 100-bp paired-end reads (342-bp insert size). The raw sequences generated from these reads were checked and filtered using a homemade quality criterion script. The paired-end reads were assembled into longer scaffolds using the de novo assembler VELVET 1.2.10 (Zerbino and Birney, 2008) with an optimized *k*-mer parameter of 99. The mtDNA was identified through a comparison between *T. solanivora* scaffolds and the mtDNA sequences reported in the NCBI GenBank, resulting in the identification of a single mtDNA contig with 200× coverage in comparison to 20–25× for nuclear contigs.

2.3. Gene annotation and compositional analysis

To predict the protein-coding genes (PCGs), rRNA genes and tRNA genes from *T. solanivora* mtDNA, the sequence was submitted to the automatic annotators of mitochondrial genes online: Dual Organellar Genome Annotation (DOGMA, http://dogma.ccbb.utexas.edu) (Wyman et al., 2004) and MITOS WebServer (Bernt et al., 2013), following indications by Cameron (2014), and curated manually. Gene sequences were manually curated using the NCBI BLAST program to determine sequence homology with other previously sequenced Lepidoptera species. The nucleotide sequences of the PCGs were translated into putative proteins based on the Invertebrate Mitochondrial Genetic Code for the determination of Relative Synonymous Codon Usage (RSCU) using MEGA version 5.2.2 (Tamura et al., 2011). The nucleotide compositions of each gene, the entire genome and each codon position of the PCGs were calculated with MEGA. Composition skew analysis was carried out according to the following formulas: AT skew = [A − T] / [A + T] and GC skew = [G − C] / [G + C], where A, T, G and C are the frequencies of the four bases. Intergenic and overlap sequences were identified using SeqBuilder from the DNAStar package (DNAStar Inc., Madison, Wisconsin, USA) and retrieved manually. Graphical representation of the *T. solanivora* mitogenome was performed using this tool as well.

2.4. Phylogenetic analysis

To illustrate the phylogenetic relationship of *T. solanivora* (Lepidoptera: Gelechiidae) to other Lepidoptera families (28), we used a concatenated set of PCGs from a set of 140 additional complete Lepidoptera mitogenomes obtained from GenBank with previous elimination of start and stop codons. The mitogenomes of *Aedes aegypti* (Behura et al., 2011) and *Acrida cinerea* (Liu and Huang, 2010) represented our out-groups (for the complete genome list see Supplementary Table 1). The nucleotide sequences of the 13 individual PCGs were aligned using RevTrans 1.4 (Wernersson and Pedersen, 2003). The resulting alignments were concatenated via a custom Python script to generate the PCG alignment used to reconstruct the phylogenetic relationships within Lepidoptera. Prior to the analysis, partition finder version 1.1.1 (Lanfear et al., 2014) was run on the concatenated datasets to identify an appropriate partitioning scheme and model. The general time-reversible model with proportion of invariable sites and gamma distributed rate variation among sites (GTR + I + G) (Tavaré, 1986) was chosen as the appropriate model of molecular evolution for all partitions. Bayesian Inference (BI) and Maximum Likelihood (ML) analytical approaches were used to infer phylogenetic trees by MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003) and Garli mpi version 2.01 (Bazinet and Cummings, 2008), respectively. Both analyses were performed in the Apolo computer cluster at Universidad EAFIT. BI analysis was conducted under the following conditions: 10,000,000 generations, eight chains (one cold and seven hot chains) and an initial 35% of the sampled trees was discarded as burn-in and consensus tree was constructed using a 50% majority rule. ML analysis was performed using the default parameters and the appropriate molecular evolution model. Confidence values of the ML tree were calculated through the bootstrap test using 1000 iterations. The software TOPD/FMTS (Puigbo et al., 2007) was used to calculate the differences between inferred trees.

3. Results

3.1. Genome structure, organization and base composition

The *T. solanivora* mitogenome is a closed circular 15,251-bp molecule (GenBank accession number KT326187) (Fig. 1). It contains the typical set of 37 genes (13 PCGs, 22 tRNAs and two rRNAs) and a large, 332-bp non-coding region (A + T rich region). A total of 24 genes were transcribed on the heavy-coding strand, while the rest were transcribed on the light-coding strand (Table 1). The typical lepidopteran arrangement of the tRNAs (tRNA-Met; tRNA-Ile; tRNA-Gln) was observed in the *T. solanivora* mitogenome and differs from the order found in ancient insects (Fig. 2). The nucleotide composition of the entire mitogenome (A: 38.6%, T: 39.6%, C: 13.3% and G: 8.4%) is highly A + T-biased (78.2%) and exhibits negative AT-skew (−0.013) and GC-skew (−0.226) values. The full composition and skew of the *T. solanivora* mitogenome are shown in Table 2.

3.2. Protein-coding genes (PCGs)

The protein-coding genes encompassed 11,146 bp of the entire assembled sequence (73.08%) and exhibited an A + T content of 76.3%. Nine of the 13 PCGs are coded on the heavy strand (*ATP6*, *ATP8*, *COI*, *COII*, *COIII*, *Cytb*, *NAD2*, *NAD3* and *NAD6*), while the rest (*NAD1*, *NAD4*, *NAD4L* and *NAD5*) are coded on the light strand. Twelve PCGs are initiated by the canonical putative start codon ATN. The *COI* gene is initiated

Table 1
Summary of mitochondrial genome of *T. solanivora*.

Gene	Direction	Position (bp)	Length (bp)	Anticodon	Start codon	Stop codon
tRNA-Met	Forward	1–68	68	CAT		
tRNA-Ile	Forward	70–134	65	GAT		
tRNA-Gln	Reverse	136–204	69	TTG		
NAD2	Forward	259–1269	1011		ATT	TAA
tRNA-Trp	Forward	1268–1336	69	TCA		
tRNA-Cys	Reverse	1329–1394	66	GCA		
tRNA-Tyr	Reverse	1406–1471	66	GTA		
COI	Forward	1475–3010	1536		CGA	TAA
tRNA-Leu (UUR)	Forward	3006–3073	68	TAA		
COII	Forward	3074–3754	681		ATG	T
tRNA-Lys	Forward	3756–3826	71	CTT		
tRNA-Asp	Forward	3837–3904	68	GTC		
ATP8	Forward	3905–4072	168		ATT	TAA
ATP6	Forward	4066–4743	678		ATG	TAA
COIII	Forward	4743–5531	789		ATG	TAA
tRNA-Gly	Forward	5534–5600	67	TCC		
NAD3	Forward	5601–5954	354		ATT	TAA
tRNA-Ala	Forward	5964–6029	66	TGC		
tRNA-Arg	Forward	6030–6095	66	TCG		
tRNA-Asn	Forward	6101–6166	66	GTT		
tRNA-Ser (AGN)	Forward	6181–6246	66	GCT		
tRNA-Glu	Forward	6247–6315	69	TTC		
tRNA-Phe	Reverse	6314–6380	67	GAA		
NAD5	Reverse	6382–8097	1716		ATT	T
tRNA-His	Reverse	8113–8178	66	GTG		
NAD4	Reverse	8183–9523	1341		ATG	TAA
NAD4L	Reverse	9523–9816	294		ATG	TAA
tRNA-Thr	Forward	9819–9883	65	TGT		
tRNA-Pro	Reverse	9884–9949	66	TGG		
NAD6	Forward	9952–10,479	525		ATA	TAA
Cytb	Forward	10,497–11,642	1146		ATA	TAA
tRNA-Ser	Forward	11,646–11,712	67	TGA		
NAD1	Reverse	11,730–12,665	936		ATA	TAG
tRNA-Leu	Reverse	12,669–12,736	68	TAG		
rRNA-Large	Reverse	12,737–14,065	1329			
tRNA-Val	Reverse	14,089–14,155	67	TAC		
rRNA-Small	Reverse	14,157–14,919	763			
A + T region	-	14,920–15,251	332			

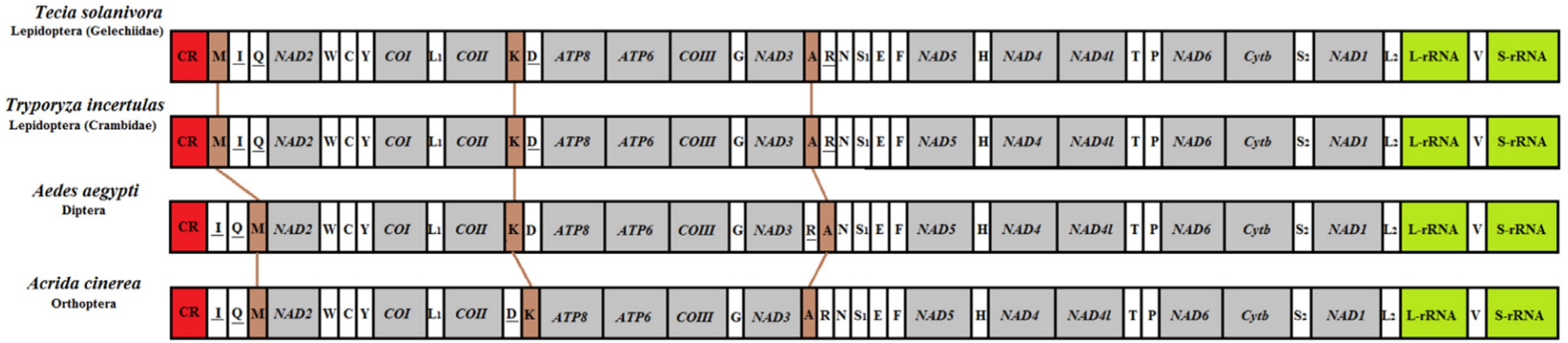

Fig. 2. Gene arrangement of the *T. solanivora* mitogenome. Protein-coding genes are marked by gray, ribosomal RNA genes by green, A + T rich region by red and tRNA genes are designated by the single letter amino acid code inside the white boxes. Brown or white boxes with underscore represent gene clusters that changed positions.

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

nt %	Whole mtDNA	Protein-coding sequence			Concatenated PCGs	rRNAs	tRNAs	IGs	A + T rich region
		1st#	2nd#	3rd#					
A%	38.6	35.3	21.3	38.3	31.6	43.8	40.4	44.7	43.1
T%	39.6	36.6	48.2	49.3	44.7	39.8	40.2	44.2	48.2
C%	13.3	11.1	17	7.2	11.8	5.2	8.1	7.6	6
G%	8.4	17.1	13.5	5.2	11.9	11.2	11.2	3.6	2.7
A + T%	78.2	71.9	69.5	87.6	76.3	83.6	80.6	88.9	91.3
C + G%	21.7	28.2	30.5	12.4	23.7	16.4	19.3	11.2	8.7
AT-Skew%	−0.013	−0.018	−0.387	−0.126	−0.172	0.048	0.002	0.006	−0.056
GC-Skew%	−0.226	0.213	−0.115	−0.161	0.004	0.366	0.161	−0.357	−0.379

by CGA (Arginine). 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, isoleucine codon (ATT) was used to initiate protein synthesis in the *ATP8*, *NAD2*, *NAD3* and *NAD5* genes. Eleven genes shared the complete stop codon TAA and the *NAD1* gene used the TAG stop codon, while the *COII* and *NAD5* genes used a single T as an incomplete stop codon. 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 Fig. 3a. Additionally, the Relative Synonymous Codon Use (RSCU) was determined. This estimated that at the third position, the codons were richer in A or T and consequently, in this position, these codons have less G or C. Finally, we also observed that *T. solanivora* presents all typical codons found in other invertebrates (Fig. 3b).

3.3. Transfer RNA and ribosomal RNA genes

T. solanivora contains a typical set of 22 tRNAs with a high A + T bias, accounting for 80.6% of the tRNAs. The lengths of the tRNA genes ranged from 65 to 71 bp, and the genes were spread over the mitogenome and exhibited positive AT-skew (0.002). Among the tRNA genes, 14 are coded on the heavy strand and eight on the light strand, which is the same coding pattern observed in almost all lepidopteran mitogenomes.

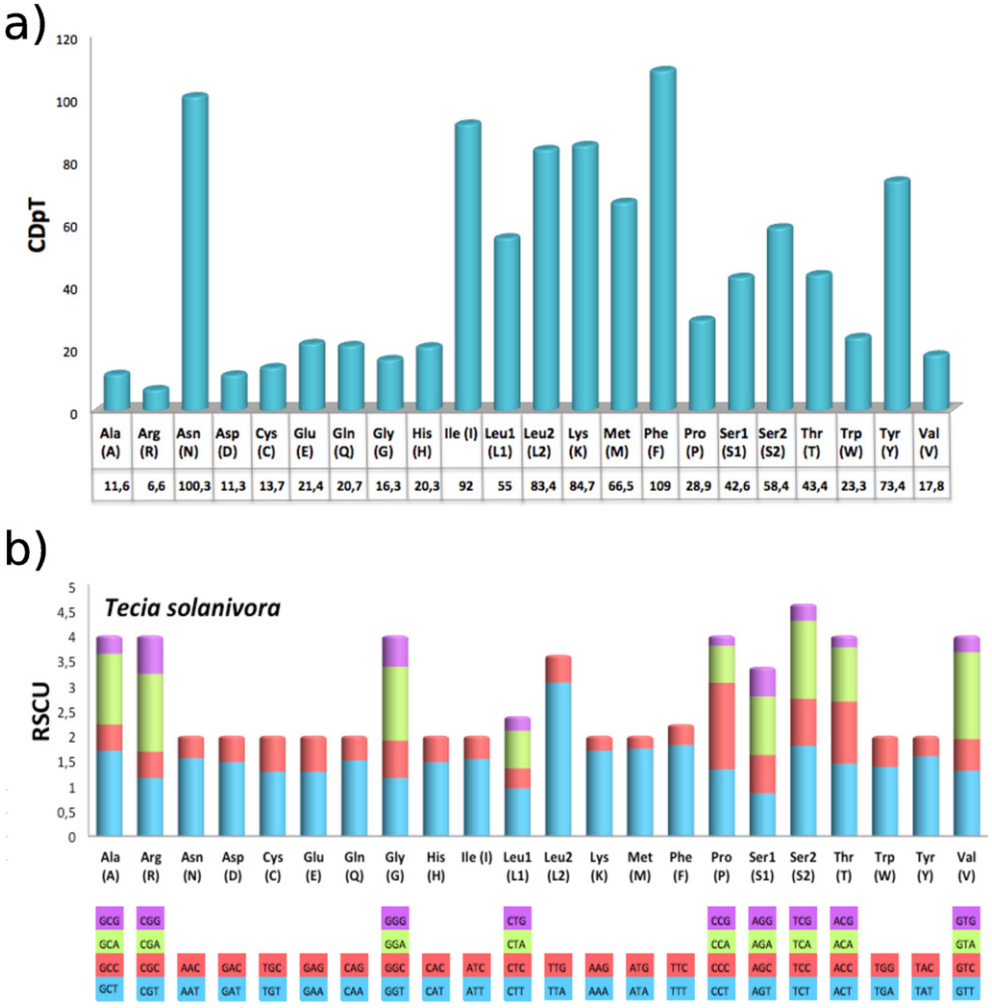

Fig. 3. Codon distribution and Relative Synonymous Codon Usage (RSCU) in *T. solanivora* mitogenome. (a) Codon distribution. (b) RSCU. Codon families are provided on the X axis and the RSCU on the Y axis. This mitogenome presents all possible codon families existing in Lepidoptera.


Bombycidae
Sphingidae
Saturniidae
Lasiocampidae
Geometridae
Erebidae
Lymantriidae
Noctuoidae
Nolidae
Notodontidae
Crambidae
Pyralidae
Nymphalidae
Lycaenidae
Riodinidae
Pieridae
Hesperiidae
Papilionidae
Gelechiidae
Oecophoridae
Carposinidae
Cossidae
Zygaenidae
Tortricidae
Yponomeutidae
Lyonetiidae
Plutellidae
Hepialidae
A
BI
Cyana sp.
Vamuna virilis
Amata formosae
Nyctemera arctata albofasciata
Hyphantria cunea
Catocala deuteronympha
Asota plana lacteata
Lymantria dispar
Gynaephora menyuanensis
Euproctis pseudoconspersa
ML
Vamuna virilis
Cyana sp.
Amata formosae
Nyctemera arctata albofasciata
Hyphantria cunea
Lymantria dispar
Gynaephora menyuanensis
Euproctis pseudoconspersa
Asota plana lacteata
Catocala deuteronympha
B
BI
Choaspes benjaminii
Celaenorrhinus maculosa
Daimio tethys
Potanthus flavus
Polytremis nascens
Carterocephalus silvicola
Lobocla bifasciata
Erynnis montanus
ML
Lobocla bifasciata
Carterocephalus silvicola
Celaenorrhinus maculosa
Daimio tethys
Potanthus flavus
Polytremis nascens
Erynnis montanus
Choaspes benjaminii
Bombycoidea
Bombyx mori Xiafang
Bombyx mori
Bombix mori C108
Chinese Bombyx mandarina
Japanese Bombyx mandarina
Bombyx huttoni
Rondotia menciana
Rondotia menciana 1
Manduca sexta
Sphinx morio
Antheraea yamamai
Antheraea pernyi
Actias selene
Eriogyna pyretorum
Saturnia boisduvalii
Samia cynthia cynthia
Samia cynthia ricini
Attacus atlas
Dendrolimus punctatus
Dendrolimus tabulaeformis
Dendrolimus spectabilis
Geometroidea
Biston thibetaria
Biston panterinaria
Biston suppressaria
Apocheima cinerarius
Jankowskia athleta
Phthonandria atrilineata
Celenna sp.
Operophtera brumata
Noctuoidea
Cyana sp.
Vamuna virilis
Amata formosae
Nyctemera arctata albofasciata
Hyphantria cunea
Catocala deuteronympha
Asota plana lacteata
Lymantria dispar
Gynaephora menyuanensis
Euproctis pseudoconspersa
Helicoverpa assulta
Helicoverpa armigera
Helicoverpa punctigera
Sesamia inferens
Agrotis segetum
Spodoptera litura
Ctenoplusia limbirena
Eutelia adulatricoides
Gabala argentata
Risoba prominens
Phalera flavescens
Ochrogaster lunifer
Pyraloidea
Paracymoriza prodigalis
Paracymoriza distinctalis
Elophila interruptalis
Chilo auricilius
Diatraea saccharalis
Scirpophaga incertulas
Maruca vitrata
Maruca testulalis
Glyphodes quadrimaculalis
Dichocrocis punctiferalis
Ostrinia nubilalis
Ostrinia furnacalis
Ephestia kuehniella
Lista haraldusalis
Corcyra cephalonica
Papilionoidea
Sasakia charonda kuriyamaensis
Sasakia charonda
Sasakia funebris
Apatura ilia
Apatura metis
Kallima inachus
Melitaea cinxia
Argynnis hyperbius
Fabriciana nerippe
Issoria lathonia
Acraeas issoria
Athyma sulpitia
Libythea celtis
Hipparchia autonoe
Calinaga davidis
Euploea mulciber
Protantigius superans
Coreana raphaelis
Japonica lutea
Cupido argiades
Lycaena phlaeas
Spindasis takanonis
Curetis bulis
Apodemia mormo
Pieris canidia
Pieris rapae
Artogeia melete
Aporia crataegi
Delias hyparete
Anthocharis bambusarum
Hebomoia glaucippe
Catopsilia pomona
Colias erate
Gonepteryx mahaguru
Leptidea morsei
Choaspes benjaminii
Celaenorrhinus maculosa
Daimio tethys
Potanthus flavus
Polytremis nascens
Carterocephalus silvicola
Lobocla bifasciata
Erynnis montanus
Papilio bianor
Papilio polytes
Papilio machaon
Papilio maraho
Papilio glaucus
Teinopalpus aureus
Lamproptera curius
Graphium chironides
Parnassius bremeri
Parnassius epaphus
Parnassius cephalus
Celastrina hersilia
Gelechioidea
Tecia solanivora
Pectinophora gossypiella
Promalactis suzukiella
Copromorphoidea
Carposina sasakii
Cossoidea
Eogystia hippophaecolus
Zygaenoidea
Rhodopsona rubiginosa
Tortricoidea
Grapholita dimorpha
Grapholita molesta
Cydia pomonella
Spilonota lechriaspis
Adoxophyes honmai
Adoxophyes orana
Prays oleae
Yponomeutoidea
Leucoptera malifoliella
Plutella xylostella
Hepialoidea
Thitarodes gonggaensis
Thitarodes renzhiensis
Thitarodes pui
Ahamus yunnanensis
Napialus hunanensis
Acrida cinerea
Aedes aegypti
0.3

Certain tRNAs were found translocated in the *T. solanivora* mitogenome compared with out-groups, which are described in Fig. 2.

Similar to other mitochondrial sequences from insect species, there were two rRNAs in *T. solanivora* with a total length of 2092 bp and an AT content of 83.6% (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 763 bp (Table 1).

3.4. Non-coding and overlapping regions

The total length of the non-coding regions in the mtDNA of *T. solanivora* is 199 bp. This region is highly A + T-biased (88.9%) (Table 2), and it is composed of 20 intergenic spacer sequences, ranging from 1 to 54 bp. There are four major intergenic spacers at least 17 bp in length (S1, S2, S3 and S4). Intergenic sequence S1 (54 bp) is located between the *tRNA-Gln* and *NAD2* genes, and intergenic sequence S2 (23 bp) is between *rRNA-Large* and *tRNA-Val*. Intergenic sequences S3 and S4 (17 bp) separates genes *NAD6* and *Cytb*, and the *tRNA-Ser2* and *NAD1* genes, respectively. The latter sequence contains the "ATACTAA" motif.

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-Trp* and *tRNA-Cys* genes. This sequence presents the greatest length; with a total of 8 bp. OLS2 was found between the *ATP8* and *ATP6* genes, with a total length of 7 bp and the 5-bp OLS3 overlapped genes *COI* and *tRNA-Leu2*.

3.5. The A + T-rich region

The A + T-rich region is a non-coding region of 332 bp, located between *rRNA-Small* and *tRNA*-Met. The region contains 91.3% AT nucleotides, with negative AT- and GC-skew values (Table 2). This is a conserved structure that includes the "ATAGA" motif followed by a 17-bp poly-T stretch. A microsatellite sequence, $(AT)_8$, was also found here and was preceded by the 'ATTTA' motif. In addition, the $(TA)_4$ and $(TAT)_7$ microsatellites as well as the $(T)_6$ and $(A)_{10}$ mononucleotide repetitive motifs were found.

3.6. Phylogenetic relationships

The phylogenetic relationship among the eight Lepidoptera superfamilies was inferred using both Bayesian Inference and Maximum Likelihood methods. The results obtained with both methods produced concordant topologies. The Gelechiidae family forms a monophyletic group composed by *T. solanivora* and *P. gossypiella*. All other taxa analyzed were distributed in clades according to their traditional taxonomic classification with high confidence, as their posterior probabilities were higher than 90% with bootstrapping values greater than 75%, except two clades of superfamilies Noctuoidea and Papilionoidea (Fig. 4).

Phylogenetic analyses resulted in high support values for the majority of the nodes and thus the interrelationships are well resolved within order Lepidoptera. Using *A. aegypti* (Diptera) and *A. cinerea* (Orthoptera) as out-groups. Lepidopteran clades are revealed in phylogenetic trees. Species of the Papilionoidea, Noctuoidea, Bombycidea, Geometroidea, Pyraloidea, Gelechioidea, Tortricoidea, Yponomeutoidea, Hepialoidea, Zygaenoidea, Cossoidea and Copromorphoidea superfamilies cluster as monophyletic groups, with strongly supported bootstrap (72–100) and posterior probabilities (0.926–1).

The optimal cladograms inferred by Bayesian Inference and Maximum Likelihood methods, produced nearly identical topologies regardless of partitioning schemes, with the exception of the relationships among three taxa from Hesperidae family and the Lymantriidae–Erebidae families. In the BI analysis *Carterocephalus silvicola* was grouped within a small clade composed by two skippers of Hesperiidae family (*Carterocephalus silvícola* + (*Potanthus flavus* + *Polytremis nascens*)) with a nodal support equal to 0.6959 posterior probabilities, while in the ML analysis *C. silvicola* was grouped with *Lobocla bifasciata* in the same clade (*C. silvicola* + *L. bifasciata*) with a nodal bootstrap probability support of 49.

Erynnis montanus was grouped with *L. bifasciata* in the BI analysis with a posterior probability support equal to 0.7739.

On the other hand, the Lymantriidae family is composed by (*Euproctis pseudoconspersa* + (*Lymantria dispar* + *Gynaephora menyuanensis*)), and shares a common ancestor with all species of the Erebidae family in the tree inferred by BI with a nodal support of 1, while in the ML's tree the Lymantriidae family is not differentiated and appear grouped within Erebidae family with low bootstrap support (52).

4. Discussion

The gene order and orientation of the mitochondrial genes of *T. solanivora* were identical to other lepidopteran moths, including *Tryporyza incertulas* (Cao and Du, 2014), *Corcyra cephalonica* (Wu et al., 2012a, b), *Adoxophyes honmai* (Lee et al., 2006), *Apocheima cinerarius* (Liu et al., 2014), *Amata emma* (Lu et al., 2013), *Attacus atlas* (Chen et al., 2014), *Bombyx mori* (Dai et al., 2013), *Caligula boisduvalii* (Hong et al., 2008), *Chilo auricilius* (Cao and Du, 2014), *Diaphania pyloalis* (Zhu et al., 2013), *Manduca sexta* (Cameron and Whiting, 2008), *Ostrinia nubilalis*, *Ostrinia furnacalis* (Coates et al., 2005), *Samia cynthia ricini* (Kim et al., 2012) and *Sasakia funebris* (Wang et al., 2013), among others. However, *T. solanivora* presents several differences from the ancestral organization of the *tRNA-Met* region (A + T-rich region, *tRNA-Ile*, *tRNA-Gln*, *tRNA-Met*) (Wu et al., 2012a, b). 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* (Fig. 2).

The nucleotide composition of the *T. solanivora* mitogenome presents an adenine (A) and thymine (T) relative content within the reported ranges for other Lepidoptera mitogenomes. These Lepidoptera mitogenomes, including *T. solanivora*, present negative AT- and GC-skew values (Table 3). However, other Lepidoptera have shown higher percentages of A than T, including *O. nubilalis* (A: 41.3%, T: 38.8%), *O. furnacalis* (A: 41.46%, T: 38.92%), *B. mori* (A: 43.06%, T: 38.30%) (Yukuhiro et al., 2002), *Phthonandria atrilineata* (A: 40.78%, T: 40.24%) (Yang et al., 2009), *Ochrogaster lunifer* (A: 40.09%, T: 37.75%) (Salvato et al., 2008), Chinese *Bombyx mandarina* (A: 43.11%, T: 38.48%) (Pan et al., 2008) and *A. atlas* (A: 39.8%, T: 39.5%), among others.

The cytosine (C) content in the *T. solanivora* mitogenome was greater than guanine (G), which is common in other recently discovered Lepidoptera mitogenomes (Table 3), except for *Antheraea yamamai* (G: 10.71%, C: 10.35%) (Kim et al., 2009), *Eriogyna pyretorum* (G: 10.61%, C: 7.45) and *Artogeia melete* (G: 11.33%, C: 8.65%) (Hong et al., 2009).

The A + T content was calculated for the three-codon positions in the protein coding genes, and they showed few differences from other Lepidoptera mitogenomes. In *T. solanivora*, the A + T content for the first position was 71.9%, and similar values were observed in *A. emma* (73.1%), *Antheraea pernyi* (72.9%) (Liu et al., 2008), *C. boisduvalii* (73.8%), *S. cynthia ricini* (72.9%), *B. mandarina* (75.0%) (Yukuhiro et al., 2002) and *M. sexta* (74.8%). In the second position, a lower A + T

Fig. 4. Inferred phylogenetic relationships between Lepidoptera based on nucleotide sequences of mitochondrial 13 PCGs using Bayesian inference (BI) and Maximum Likelihood (ML). Numbers at nodes indicate Bayesian posterior probabilities (first value) and Bootstrap values (second value). The dipteran *A. aegypti* and the Orthopteran *A. cinerea* were used as outgroups. Color of the branches represents Lepidopteran families and color-legend is shown in the top left. Topology differences between methods are shown in A (Noctuoidea superfamily) and B (Papilionoidea superfamily) subfigures.

Table 3
Comparison of nucleotide composition and skewness between *T. solanivora* and other lepidopteran mitogenomes.

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

percentage was found (69.5%) (Table 2). The negative AT- and GC-skew values indicate a greater inclination for the nitrogen bases, T and C. Moreover, the second and third positions showed negative values for both AT-skew and GC-skew, while the first position showed a slightly positive value for the GC-skew (Table 2), indicating a greater bias for G than C. On the other hand, the nucleotide composition in tRNAs exhibit a higher inclination for nitrogen bases A and G than for T and C. Similar results were reported by Jiang et al. (2009) in *E. pyretorum* (AT-skew = 0.039 and GC-skew = 0.174) and Hong et al. (2009) in *A. melete* (AT-skew = 0.034 and GC-skew = 0.142).

Most PCGs of *T. solanivora* mitogenome showed typical ATN initiation codons (isoleucine and methionine). These initiation codons are frequently found in the majority of Lepidoptera insects (Liu et al., 2008; Cameron and Whiting, 2008), except for the *COI* gene that presented a CGA initiation codon (arginine). This codon represents the first codon in mature mRNA for this gene and is a conserved feature in the order Lepidoptera (Fenn et al., 2007; Cameron and Whiting, 2008), for which reason it is considered a synapomorphy of this group of insects (Kim et al., 2014a, b). It is important to point out that the high A + T percentages in insect mitogenomes result in high probabilities of finding a non-coding triplet or a coding triplet within the *tRNA-Tyr* gene, a result that could potentially produce generalized annotation errors for the gene *COI* (Stewart and Beckenbach, 2009).

The TAA codon was found in eleven of the PCGs, in accordance with the mitogenomes of other Lepidoptera, including *T. incertulas*, *S. funebris*, S. *cynthia* and *A. emma* (Kim et al., 2014a, b). The stop codon TAG was identified in *NAD1* gene, and similar results were obtained in five species of the family Hesperiidae (Kim et al., 2014a, b). In summary, 12 of the 13 PCGs presented complete stop codons, and for the *COII* gene, we identified an incomplete stop codon, T, which is commonly found in the majority of Lepidoptera species to date (Cameron and Whiting, 2008; Kim et al., 2014a, b). This truncated codon could be 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) (Ojala et al., 1981; Cameron and Whiting, 2008; Cao and Du, 2014).

The codon distribution analysis (Fig. 3a) shows that in *T. solanivora*, the families of codons that represent the amino acid phenylalanine (*Phe*) are more abundant, instead of Leucine 2 (*Leu2*), which dominates in other Lepidoptera mitogenomes (Salvato et al., 2008). The relative synonymous codon use in *T. solanivora*, as in other Lepidoptera, prefers codons with an A or T in their third position (Lu et al., 2013).

The rRNA lengths were within the range of values reported for other Lepidoptera, as their values range between 1314 bp in *Euploea mulciber* (Nymphalidae) (Hao et al., 2013) and 1330 bp in *Coreana raphaelis* (Lycaenidae). For the *rRNA-Large* and *rRNA-Small*, the length ranges from 739 bp for *Protantigius superans* (Lycaenidae) to 788 bp in *Acraea issoria* (Nymphalidae) (Kim et al., 2014a, b). 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%) (Wu et al., 2012a, b).

Most of the intergenic regions in this mitogenome were short (≤15 bp), however, four longer overlap regions were found. The S1 intergenic sequence is commonly found in the Lepidoptera mitogenome, and this region has not been identified in insects that belong to other orders (Cameron and Whiting, 2008). The length of this sequence ranges between 38 bp in *T. incertulas* and 88 bp in *Sasakia charonda* (Wu et al., 2012a, b). This sequence can be considered a mitogenome marker for the order Lepidoptera, and it most likely originated from a partial *NAD2* gene duplication (Cao and Du, 2014).

Intergenic sequence S4 (17 bp) contains the "ATACTAA" motif typically found in other lepidopterans (Cameron and Whiting, 2008; Cao et al., 2012; Lu et al., 2013). This motif plays an apparent role as a recognition site for the protein implicated in mitochondrial transcription termination (mtTERM) (Taanman, 1999). Furthermore, this sequence has been recognized for being highly conserved, with a length ranging between 17 and 20 bp (Lu et al., 2013).

The 8-bp OLS1 overlap sequence and OLS2 (7 bp) are consistent with the same genes found in other lepidopterans, although they differ in length (Kim et al., 2009; Taanman, 1999). In addition, the overlap region OLS3 between *COI* and *tRNA-leu2* genes was previously found in *Maruca vitrata* with a length of 2 bp. This study was based on transcription and expression of mitochondrial genes (Margam et al., 2011).

The A + T-rich region is 332 bp in length with 91.3% A + T content, and it has a negative skew value (−0.056), meaning that 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 (Lee et al., 2006). The length of this region is variable in the order Lepidoptera, and it can be as long as 1270 bp, as reported in *Papilio bianor* (Papilionidae) (Hou et al., 2014). This region represents a conserved structural region commonly found in the order Lepidoptera, which includes the "ATAGA" motif followed by a segment with a variable poly-T sequence (15 to 17 bp), and this motif is immediately followed by the *tRNA-Met* gene (Salvato et al., 2008; Zhao et al., 2011; Kim et al., 2014a, b). In *T. solanivora*, this segment is composed of 17 thymine nucleotides. Furthermore, this segment has been considered a fundamental sequence for gene regulation and it apparently functions as a possible recognition site for the replication initiation of the minor strain of mtDNA (Cao and Du, 2014; Wang et al., 2013; Kim et al., 2014a, b). In addition, eight microsatellite regions were identified within the mitogenome of *T. solanivora*, referred to as $(TAA)_4$, $(AT)_8$ and $(TAT)_7$. These were the most representative microsatellites found in the species, although the mononucleotide sequences, $(T)_6$ and $(A)_{10}$, were also identified (Van Oppen et al., 1999). These represent the relevant regions of this genome for future studies. Also, in most lepidopteran mitogenomes, the $(AT)_8$ microsatellite has been previously reported. This microsatellite is preceded by the "ATTTA" motif that is commonly found in other mitogenomes (Cameron and Whiting, 2008; Kim et al., 2014a, b). In *S. funebris*, the same $(AT)_8$ microsatellite was identified as that found in *T. solanivora*. Nevertheless, most lepidopteran mitogenomes report $(AT)_n$, where n ranges from 7 to 12 (Lu et al., 2013; Cao and Du, 2014).

The results obtained in the phylogenetic analysis are well supported and generated the first phylogeny of the family Gelechiidae based on the utilization of mitogenome data, showing to *T. solanivora* and *P. gossypiella* in a monophyletic group with high support values in both trees (1/96, BI/ML). This relationship it is accordance with traits such as both are monophagous pests and they feed on internal parts of the crops. However, other molecular markers are necessary to complement this work, including nuclear markers. Also it is important to increase the number of taxa to study the Gelechiidae family.

An study made by Karsholt et al. (2013), was based on sequencing seven nuclear genes and one mitochondrial gene (*COI*) and demonstrated that pest species of this family are scattered among different lineages, dividing this family in six distinct clades which seem be associated with traits such as larval mode of life and external or internal feeding in the crops within the family.

Additionally, the Gelechiidae family shares a common ancestor with *Promalactis suzukiella*, member of Oecophoridae family from Gelechioidea superfamily, which is not shared with other families of the order Lepidoptera, even so, this hypothesis must be tested using other genes and a larger number of analyzed taxa, including individuals from the Cosmopterigidae family, because previous studies based both on morphological features (Hodges, 1998) and molecular analysis (Kaila et al., 2011; Karsholt et al., 2013) support the sister group relationship between Cosmopterigidae and Gelechiidae families.

Recent molecular phylogenetic studies of the insect order Lepidoptera using datasets of concatenated mitochondrial protein coding genes and rRNAs (Timmermans et al., 2014) and nuclear genes (Sohn et al., 2015) support a relationship between Gelechioidea and the advanced clade Obtectomera sensu, These studies shown an extended clade Obtectomera, including Gelechioidea, Thyridoidea and Pterophoroidea, in addition to the conventional clade composition: Papilionoidea, Pyraloidea and Macroheterocera (Bombycoidea, Noctuoidea, Geometroidea, Depranoidea) (Bazinet et al., 2013), which seems consistent with our study.

5. Conclusion

In this study, the mitogenome of *T. solanivora* was sequenced, analyzed and compared with other lepidopteran insects, making this the second mitogenome record of the family Gelechiidae. 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 12 clades of the order Lepidoptera were shown in this work using Bayesian and Maximum Likelihood inference, which provided well-supported results consistent with molecular and morphological traits. In addition, the taxonomic status of the *T. solanivora* species was shown, detecting its close relationship with another pest: *P. gossypiella*. This study provides relevant information on the phylogeny of the family Gelechiidae, which is composed of pest species that produce economic loss in several countries in the Western Hemisphere, particularly with potato crops, as is the case with *T. solanivora* or the potato tuber moth. However, Gelechiidae family continues being little explored.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gene.2016.01.031.

Acknowledgments

The authors thank the Scientific Computing Center Apolo of Univerisidad EAFIT for the contribution on computing resource and Elsevier Webshop for the English revision of this manuscript. We thank Nicolás Pinel at the Department of Biological Sciences for critical reading of the manuscript.

This study received financial support for the genome sequencing from the Universidad Nacional de Colombia to Clara Inés Saldamando-Benjumea with the grant number 18608.

References

Avise, J.C., 1994. Molecular Markers, Natural History and Evolution. Chapman & Hall, New York.

Bazinet, A.L., Cummings, M.P., 2008. The Lattice Project: a grid research and production environment combining multiple grid computing models. In: Weber, M.H.W. (Ed.), Distributed & Grid Computing — Science Made Transparent for Everyone, pp. 2–13.

Bazinet, A.L., Cummings, M.P., Mitter, K.T., Mitter, C.W., 2013. Can RNA-Seq resolve the rapid radiation of advanced moths and butterflies (Hexapoda Lepidoptera: Apoditrysia)? An exploratory study. PLoS One 8, e82615.

Behura, S.K., Lobo, N.F., Haas, B., deBruyn, B., Lovin, D.D., Shumway, M.F., Puiu, D., Romero-Severson, J., Nene, V., Severson, D.W., 2011. Complete sequences of mitochondria genomes of *Aedes aegypti* and *Culex quinquefasciatus* and comparative analysis of mitochondrial DNA fragments inserted in the nuclear genomes. Insect Biochem. Mol. Biol. 41, 770–777. http://dx.doi.org/10.1371/journal.pone.0021504.

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., Stadler, P.F., 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. 69 (2), 313–319.

Boore, J.L., 1999. Animal mitochondrial genomes. Nucleic Acids Res. 27 (8), 1767–1780.

Cameron, S.L., 2014. How to sequence and annotate insect mitochondrial genomes for systematic and comparative genomics research. Syst. Entomol. 39, 400–411. http://dx.doi.org/10.1111/syen.12071.

Cameron, S.L., Whiting, M.F., 2008. 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 408, 112–123. http://dx.doi.org/10.1016/j.gene.2007.10.023.

Cao, S.S., Du, Y.Z., 2014. Characterization of the complete mitochondrial genome of *Chilo auricilius* and comparison with three other rice stem borers. Gene 548, 270–276. http://dx.doi.org/10.1016/j.gene.2014.07.045.

Cao, Y.Q., Ma, C., Chen, J.Y., Yang, D.R., 2012. The complete mitochondrial genomes of two ghost moths, *Thitarodes renzhiensis* and *Thitarodes yunnanensis*: the ancestral gene arrangement in Lepidoptera. BMC Genomics 13, 276. http://dx.doi.org/10.1186/1471-2164-13-276.

Chen, M.M., Li, Y., Chen, M., Wang, H., Li, Q., Xia, R.X., Zeng, C.Y., L.Y., 2014. Complete mitochondrial genome of the atlas moth, *Attacus atlas* (Lepidoptera: Saturniidae) and the phylogenetic relationship of Saturniidae species. Gene 545, 95–101.

Coates, B.S., Sumerford, D.V., Hellmich, R.L., Lewis, L.C., 2005. Partial mitochondrial genome sequences of *Ostrinia nubilalis* and *Ostrinia furnicalis*. Int. J. Biol. Sci. 1, 13–18.

Dai, L.S., Zhu, B.J., Liu, Q.N., Wei, G.Q., Liu, C.L., 2013. Characterization of the complete mitochondrial genome of *Bombyx mori* strain H9 (Lepidoptera: Bombycidae). Gene 519, 326–334.

Fenn, J.D., Cameron, S.L., Whiting, M.F., 2007. The complete mitochondrial genome sequence of the Mormon cricket (*Anabrus simplex*: Tettigoniidae: Orthoptera) and an analysis of control region variability. Insect Mol. Biol. 16, 239–252. http://dx.doi.org/10.1111/j.1365-2583.2006.00721.x.

Hao, J., Sun, M., Shi, Q., Sun, X., Shao, L., Yang, Q., 2013. Complete mitogenomes of *Euploea mulciber* (Nymphalidae: Danainae) and *Libythea celtis* (Nymphalidae: Libytheinae) and their phylogenetic implications. ISRN Genomics 14. http://dx.doi.org/10.1155/2013/491636.

Hodges, R.W., 1998. Lepidoptera, Moths and Butterflies, Vol 1: Evolution, Systematics, and Biogeography. Handbook of Zoology, Vol. IV: Arthropoda: Insecta, Part 35 (ed. by N.P. Kristensen), pp. 131–158.

Hong, M.Y., Lee, E.M., Jo, Y.H., Park, H.C., Kim, S.R., Hwang, J.S., Jin, B.R., Kang, P.D., Kim, K.G., Han, Y.S., Kim, I., 2008. Complete nucleotide sequence and organization of the mitogenome of the silkmoth *Caligula boisduvalii* (Lepidoptera: Saturniidae) and comparison with other lepidopteran insects. Gene 413, 49–57.

Hong, G.Y., Jiang, S., Yu, M., Yang, Y., Li, F., Xue, F., Wei, Z., 2009. The complete nucleotide sequence of the mitochondrial genome of the cabbage butterfly, *Artogeia melete* (Lepidoptera: Pieridae). Acta Biochim. Biophys. Sin. 41, 446–455. http://dx.doi.org/10.1093/abbs/gmp030.

Hou, L.X., Ying, S., Yang, X.W., Yu, Z., Li, H.M., Qin, X.M., 2014. The complete mitochondrial genome of *Papilio bianor* (Lepidoptera: Papilionidae), and its phylogenetic position within Papilionidae. Mitochondrial DNA http://dx.doi.org/10.3109/19401736.2013.873923.

Jiang, S.T., Hong, G.Y., Yu, M., Li, N., Yang, Y., Liu, Y.Q., Wei, Z.J., 2009. Characterization of the complete mitochondrial genome of the giant silkworm moth, *Eriogyna pyretorum* (Lepidoptera: Saturniidae). Int. J. Biol. Sci. 5, 351–365.

Kaila, L., Mutanen, M., Nyman, T., 2011. Phylogeny of the mega-diverse Gelechioidea (Lepidoptera): adaptations and determinants of success. Mol. Phylogenet. Evol. 61, 801–809. http://dx.doi.org/10.1016/j.ympev.2011.08.016.

Karsholt, O., Mutanen, M., Lee, S., Kaila, L., 2013. A molecular analysis of the Gelechiidae (Lepidoptera, Gelechioidea) with an interpretative grouping of its taxa. Syst. Entomol. 38, 334–348.

Kim, S.R., Kim, M.I., Hong, M.Y., Kim, K.Y., Kang, P.D., Hwang, J.S., Han, Y.S., Jin, B.R., Kim, I., 2009. The complete mitogenome sequence of the Japanese oak silkmoth, *Antheraea yamamai* (Lepidoptera: Saturniidae). Mol. Biol. Rep. 36, 1871–1880. http://dx.doi.org/10.1007/s11033-008-9393-2.

Kim, J.S., Park, J.S., Kim, M.J., Kang, P.D., Kim, S.G., Jin, B.R., Han, Y.S., Kim, I., 2012. Complete nucleotide sequence and organization of the mitochondrial genome of eri-silkworm, *Samia cynthia ricini* (Lepidoptera: Saturniidae). J. Asia Pac. Entomol. 15, 162–173. http://dx.doi.org/10.1016/j.aspen.2011.10.002.

Kim, M.J., Wang, A.R., Park, J.S., Kim, I., 2014a. Complete mitochondrial genomes of five skippers (Lepidoptera: Hesperiidae) and phylogenetic reconstruction of Lepidoptera. Gene 549 (1), 97–112. http://dx.doi.org/10.1016/j.gene.2014.07.052.

Kim, M.J., Jun, J., Kim, I., 2014b. Complete mitochondrial genome of the mulberry white caterpillar *Rondotia menciana* (Lepidoptera: Bombycidae). Mitochondrial DNA 1-3. http://dx.doi.org/10.3109/19401736.2014.913163.

Lanfear, R., Calcott, B., Kainer, D., Mayer, C., Stamatakis, A., 2014. Selecting optimal partitioning schemes for phylogenomic datasets. BMC Evol. Biol. 14, 82. http://dx.doi.org/10.1186/1471-2148-14-82.

Lee, E.S., Shin, K.S., Kim, M.S., Park, H., Cho, S., Kim, C.B., 2006. The mitochondrial genome of the smaller tea tortrix *Adoxophyes honmai* (Lepidoptera: Tortricidae). Gene 373, 52–57.

Liu, N., Huang, Y., 2010. Complete mitochondrial genome sequence of *Acrida cinerea* (Acrididae: Orthoptera) and comparative analysis of mitochondrial genomes in Orthoptera. Comp. Funct. Genomics 1-16. http://dx.doi.org/10.1155/2010/319486.

Liu, Y., Li, Y., Pan, M., Dai, F., Zhu, X., Lu, C., Xiang, Z., 2008. The complete mitochondrial genome of the Chinese oak silkmoth, *Antheraea pernyi* (Lepidoptera: Saturniidae). Acta Biochim. Biophys. Sin. 40, 693–703. http://dx.doi.org/10.1093/abbs/40.8.693.

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

Lu, H.F., Su, T.J., Luo, A.R., Zhu, C.D., Wu, C.S., 2013. Characterization of the complete mitochondrion genome of diurnal moth *Amata emma* (Butler) (Lepidoptera: Erebidae) and its phylogenetic implications. PLoS One 8, 1–14. http://dx.doi.org/10.1371/journal.pone.0072410.

Margam, V.M., Coates, B.S., Hellmich, R.L., Agunbiade, T., Seufferheld, M.J., et al., 2011. Mitochondrial genome sequence and expression profiling for the legume pod borer *Maruca vitrata* (Lepidoptera: Crambidae). PLoS One 6 (2), 1–9. http://dx.doi.org/10.1371/journal.pone.0016444.

Meusel, M., Moritz, R., 1993. Transfer of paternal mitochondrial DNA during fertilization of honeybee (*Apis mellifera L.*) eggs. Curr. Genet. 24 (6), 539–543. http://dx.doi.org/10.1007/BF00351719.

Ojala, D., Montoya, J., Attardi, G., 1981. TRNA punctuation model of RNA processing in human mitochondria. Nature 290, 470–474. http://dx.doi.org/10.1038/290470a0.

Pan, M., Yu, Q., Xia, Y., Dai, F., Liu, Y., Lu, C., Zhang, Z., Xiang, Z., 2008. Characterization of mitochondrial genome of Chinese wild mulberry silkworm, *Bomyx mandarina* (Lepidoptera: Bombycidae). Sci. China C Life Sci. 51, 693–701. http://dx.doi.org/10.1007/s11427-008-0097-6.

Pitnick, S., Karr, T.L., 1998. Paternal products and by-products in *Drosophila* development. Rural. Soc. 265, 821–826 (Published online http://dx.doi.org/10.1098/rspb.1998.0366).

Puigbo, P., Garcia-Vallve, S., McInerney, J.O., 2007. TOPD/FMTS: a new software to compare phylogenetic trees. Bioinformatics 23, 1556–1558. http://dx.doi.org/10.1093/bioinformatics/btm135.

Pulliandre, N., Dupas, S., Dangles, O., Zeddam, J., Capdevielle-Dulac, C., Barbin, K., Torres-Leguizamon, M., Silvain, J.F., 2008. Genetic bottleneck in invasive species: the potato tuber moth adds to the list. Biol. Invasions 10, 319–333.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Salvato, P., Simonato, M., Battisti, A., Negrisolo, E., 2008. The complete mitochondrial genome of the bag-shelter moth *Ochrogaster lunifer* (Lepidoptera, Notodontidae). BMC Genomics 9, 331. http://dx.doi.org/10.1186/1471-2164-9-331.

Sohn, J.C., Regier, J.C., Mitter, C., Adamski, D., Landry, J.F., Heikkila, M., Park, K.T., Harrison, T., Mitter, K., Zwick, A., Kawahara, A.Y., Cho, S., Cummings, M.P., Schmitz, P., 2015. Phylogeny and feeding trait evolution of the mega-diverse Gelechioidea (Lepidoptera: Obtectomera): new insight from 19 nuclear genes. Syst. Entomol. http://dx.doi.org/10.1111/syen.12143.

Stewart, J.B., Beckenbach, A.T., 2009. Characterization of mature mitochondrial transcripts in *Drosophila*, and the implications for the tRNA punctuation model in arthropods. Gene 445, 49–57. http://dx.doi.org/10.1016/j.gene.2009.06.006.

Taanman, J., 1999. The mitochondrial genome: structure, transcription. Biochim. Biophys. Acta Bioenerg. 1410, 103–123. http://dx.doi.org/10.1016/S0005-2728(98)00161–3.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 1–9 http://dx.doi.org/10.1093/molbev/msr121.

Tavaré, S., 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Lect. Math. Life Sci. Am. Math. Soc. 17, 57–86.

Timmermans, M.J.T.N., Lees, D.C., Simonsen, T.J., 2014. Towards a mitogenomic phylogeny of Lepidoptera. Mol. Phylogenet. Evol. 79, 169–178.

Torres-Leguizamón, M., Dupas, S., Dardon, D., Gómez, Y., Niño, L., Carnero, A., Padilla, A., Merlin, I., Fossoud, A., Zeddam, J., Lery, X., Capdevielle-Dulac, C., Dangles, O., Silvain, J.F., 2011. Inferring native range and invasion scenarios with mitochondrial DNA: the case of *T. solanivora* successive north–south step-wise introductions across Central and South America. Biol. Invasions 13, 1505–1519.

Van Oppen, M., Hislop, N.R., Hagerman, P.J., Miller, D.J., 1999. Gene content and organization in a segment of the mitochondrial genome of the scleractinian coral *Acropora tenuis*: major differences in gene order within the anthozoan subclass zoantharia. Mol. Biol. Ecol. 16, 1812–1815.

Villanueva-Mejía, D.F., Ramírez-Ríos, V., Arango, R.E., Saldamando-Benjumea, C.I., 2015. Genetic differentiation of *Tecia solanivora* from Colombia based on two mitochondrial genes: cytochrome b and cytochrome oxydase I. J. Insect Sci.

Wang, J.P., Cao, T.W., Xuan, S.B., Wang, H., Zhang, M., Ma, E., 2013. The complete mitochondrial genome of *Sasakia funebris* (Leech) (Lepidoptera: Nymphalidae) and comparison with other apaturinae insects. Gene 526, 336–343. http://dx.doi.org/10.1016/j.gene.2013.05.036.

Wernersson, R., Pedersen, A.G., 2003. RevTrans: multiple alignment of coding DNA from aligned amino acid sequences. Nucleic Acids Res. 31, 3537–3539.

Wolstenholme, D.R., 1992. Animal Mitochondrial DNA: Structure and Evolution. Int. Rev. Cytol., pp. 173–216. http://dx.doi.org/10.1016/S0074-7696(08)62066-5.

Wu, Y.P., Zhao, J.L., Su, T.J., Li, J., Yu, F., Chesters, D., Fan, R.J., Chen, M.C., Wu, C.S., Zhu, C.D., 2012a. The complete mitochondrial genome of *Leucoptera malifoliella* Costa (Lepidoptera: Lyonetiidae). DNA Cell Biol. 31, 1508–1522. http://dx.doi.org/10.1089/dna.2012.1642.

Wu, Y.P., Li, J., Zhao, J.L., Su, T.J., Luo, A.R., Fan, R.J., Chen, M.C., Wu, C.S., C.D., Z., 2012b. The complete mitochondrial genome of the rice moth, *Corcyra cephalonica*. J. Insect Sci. 12, 1536–2442.

Wyman, S.K., Jansen, R.K., Boore, J.L., 2004. Automatic annotation of organellar genomes with DOGMA. Bioinforma. Appl. Note. 20, 3252–3255. http://dx.doi.org/10.1093/bioinformatics/bth352.

Yang, L., Wei, Z.J., Hong, G.Y., Jiang, S.T., Wen, L.P., 2009. The complete nucleotide sequence of the mitochondrial genome of *Phthonandria atrilineata* (Lepidoptera: Geometridae). Mol. Biol. Rep. 36, 1441–1449. http://dx.doi.org/10.1007/s11033-008-9334-0.

Yukuhiro, K., Sezutsu, H., Itoh, M., Shimizu, K., Banno, Y., 2002. Significant levels of sequence divergence and gene rearrangements have occurred between the mitochondrial genomes of the wild mulberry silkmoth, *Bombyx mandarina*, and its close relative, the domesticated silkmoth, *Bombyx mori*. Mol. Biol. Evol. 19, 385–1389.

Zeddam, J.L., Orbe, K., Léry, X., Dangles, O., Dupas, S., Silvain, J., 2008. An isometric virus of the potato tuber moth *Tecia solanivora* (Povolny) (Lepidoptera: Gelechiidae) has a tri-segmented RNA genome. J. Invertebr. Pathol. 99, 204–211.

Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18, 821–829. http://dx.doi.org/10.1101/gr.074492.107.

Zhang, B.C., 1994. Index of Economically Important Lepidoptera. CAB International, Wallingford.

Zhao, J.L., Zhang, Y.Y., Luo, A.R., Jiang, G.F., Cameron, S.L., Zhu, C.D., 2011. The complete mitochondrial genome of *Spilonota lechriaspis* Meyrick (Lepidoptera: Tortricidae). Mol. Biol. Rep. 38, 3757–3764. http://dx.doi.org/10.1007/s11033-010-0491-6.

Zhu, B.J., Liu, Q.N., Dai, L.S., Wang, L., Sun, Y., Lin, K.Z., Wei, G.Q., Liu, C.L., 2013. Characterization of the complete mitochondrial genome of *Diaphania pyloalis* (Lepidoptera: Pyralididae). Gene 527, 283–291.