The first complete mitochondrial genome sequences for Ulidiidae and phylogenetic analysis of Diptera

kun xing (xingkun1215@126.com)
Shanxi Agricultural University
Chen Kang
Shannon Technology Group
Fei Zhao
Shanxi Agricultural University

Research Article

Keywords: Tetanops sintenisi, Diptera, Ulidiidae, mitochondrial genome, Diptera phylogenetics

Posted Date: March 7th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1406595/v1

The first complete mitochondrial genome sequences for Ulidiidae and phylogenetic analysis of Diptera

Kun Xing [1*], Chen Kang [2*], and Fei Zhao [1,]**

[1]College of Plant Protection, Shanxi Agricultural University, Taiyuan 030031, China; xingkun1215@126.com (K.X.);
[2]Shannon Technology Group, Taigu 0300800, China; kangchen6726@126.com (K.C.)
[*]These authors contributed equally to this work and should be considered co-first authors.
[**]Correspondence: zhaofei12@126.com; Tel.: +86 0351 713663

Abstract

Background: *Tetanops sintenisi* is a pest that mainly damages the roots of quinoa (*Chenopodium quinoa*) and it is first discovered in China in 2018. Methods and Results: Here, the complete mitochondrial genome (mitogenome) of *T. sintenisi* was sequenced and compared with the mitogenomes of other Diptera species. The results revealed that the mitogenome of *T. sintenisi* is 15763 bp in length (GenBank accession number: MT795181) and is comprised of 13 protein-coding genes (PCGs), 22 transfer RNA (tRNA) genes, 2 ribosomal RNA genes, and a non-coding A+T-rich region (959 bp). The highly conserved gene arrangement of the mitogenome of *T. sintenisi* was identical to that of other Diptera insects. Twelve PCGs contained the typical insect start codon ATN, while *cox1* had CGA as the start codon. The genes *cox2*, *nad4*, and *nad1* contained an incomplete termination codon T; *nad3*, *nad5*, and *cob* contained the complete termination codon TAG; and the remaining seven PCGs contained the termination codon TAA. All tRNA genes were predicted to fold into the typical cloverleaf secondary structure. Phylogenetic analysis of 48 species based on the mitogenome sequence revealed that *T. sintenisi* clustered with the Tephritidae family, indicating that *T. sintenisi* and Tephritidae have a close phylogenetic relationship. Conclusions: The phylogenetic relationship of *T. sintenisi* based on the mitogenome was consistent with the traditional morphological taxonomy, according to which *T. sintenisi* belongs to the family Otitidae, which is closely related to the family Muscidae.

Keywords *Tetanops sintenisi* · Diptera · Ulidiidae · mitochondrial genome · Diptera phylogenetics

Introduction

Mitochondrion is an important organelle that exists in all eukaryotic cells. In addition to providing energy for cells, mitochondria are involved in processes such as cellular differentiation, messaging, and apoptosis, and they regulate cell growth and the cell cycle. Mitochondrial genome contains genetic information that is independent of nuclear genes, and it encodes the proteins, tRNAs, and rRNAs required by the mitochondria. The insect mitochondrial genome (mitogenome) is a circular double-stranded DNA molecule with a size of approximately 14-20 kb, consisting of 37 conservative genes, comprising 2 ribosomal RNA (lrRNA and srRNA), 22 tRNA, and 13 protein-coding genes (PCGs). It also contains a major non-coding region called mitochondrial DNA (mtDNA) A+T-rich region (or control region, CR). Due to the simple structure, high copy number, relatively conserved gene arrangement, rapid gene evolution, and almost no recombination of the maternal genetic characteristics, insect mitochondria are currently widely used for the identification of insect species and determination of molecular phylogeny, biogeography, and genetic structure of the insect population.

To date, 255 insect mitogenome sequences representing 23 insect orders are available in the Insect Mitochondrial Genome Database (IMGD; http://www.imgd.org/). Among them, the number of sequenced mitogenomes of Diptera species is the highest, at 75 mitogenomes, accounting for 29.4% of the total [1], and it includes the mitogenomes of *Helophilus virgatus* [2], *Hydrotaea dentipes* [3], *Delia antiqua* [4], *Bactrocera zonata* [5], *Bactrocera tsuneonis* [6], *Anopheles funestus* [7], *Ceratitis capitata* [8], *Cochliomyia hominivorax* [9], *Lucilia sericata* [10], *Chrysomya putoria* [11], *Mayetiola destructor* [12], *Stomoxys calcitrans* [13], and *Drosophila littoralis* [14]. However, information on the *Tetanops* mitogenome is still limited. Therefore, further studies on the phylogeny and evolution of Diptera should be conducted to determine more relevant species sequences.

Tetanops sintenisi belongs to the family Ulidiidae of Diptera [15]. Its larvae mainly damage the roots of *Chenopodium* plants, causing the above ground parts to turn yellow and wilt and leading toa decreased growth and eventually death of the plants [16]. In recent years, this pest has broken out on quinoa (*Chenopodium quinoa*) fields in Shanxi Province in China, causing serious losses [16]. It was first captured from *Calluna vulgaris* in Bargerveen, central Holland in 1909 [15], and has been discovered in

Ukraine, Finland, Latvia, Russia, and other regions in Eastern Europe [17]. Since 2000, it has been found in Poland, Germany, United Kingdom, and Belgium in Western Europe [18-21]. In China, it was first discovered in Jingle County, Shanxi Province in 2018 [16]. To our knowledge, the mitogenome of *T. sintenisi* or other species of the family Ulidiidae has not been previously determined.

Here, we characterized the mitogenome sequence of *T. sintenis*i and compared it with the mitogenome sequences of other insects, especially Diptera species. In addition, we used the mitogenome sequences to construct phylogenetic relationships among the 48 families of Diptera.

Materials and Methods

Specimens

On May 10, 2018, *T. sintenisi* larvae (approximately 50) were collected from a quinoa field (112.2°E, 38.4°N) in Jingle County, Shanxi Province, China. The samples were fixed in 99% anhydrous ethanol. The ethanol was replaced with fresh 99% anhydrous ethanol once after the samples were brought back to the laboratory, and the samples were stored in a refrigerator at -30 °C.

Morphological characteristics

The taxonomic status of quinoa root maggots was determined following the morphological identification characteristics reported in the literature [15]. The egg is 1 mm long, white or milky white, fusiform, slightly curved. The larva is 8-10 mm long, white, maggot-shaped, and the head is cone-shaped; the surface of the mature larva is tough, with a pair of spikes at the end of the tail section. The pupa is 7-9 mm long, cocoon-shaped, yellow at the beginning, and brown later. Adult body length is 6-9 mm, wingspan is 11-13 mm, and surface appearance is similar to housefly, with black, shiny body, without obvious stripes or bristles. The wings are transparent, and there is a brown streak at 1/3rd of the body of the subfront vein of the forewing (the identifying feature of this species). The end of the abdomen of the male fly is black and round, and the end of the abdomen of the female fly is dark to dark orange and pointed, with a long ovipositor at the end.

Mitochondrial genome sequencing

Ten larvae were selected and total DNA was extracted using a blood/cell/tissue genomic DNA extraction kit (Tiangen Biochemical Technology (Beijing) Co., Ltd.) according to the manufacturer's instructions. The concentration and purity of the total DNA were detected using a nucleic acid analyzer, and the integrity of the total DNA was evaluated through 1% agargel electrophoresis. After the quality test was completed, the DNA was sent to Shanghai Paisenol Biotechnology Co., Ltd. for sequencing. In this study, paired end (PE) sequencing was performed on these libraries using the Illumina MiSeq Sequencing platform and next generation sequencing (NGS).

Sequence annotation and feature analysis

A5-miseq V20150522 [22] and SPAdesv3.9.0 [23] were used to assemble high-quality second-generation sequencing data de novo and construct contig and scaffold sequences. The sequences were blasted (BLAST V2.2.31+) by comparing with the nt database in NCBI, and the mitochondrial sequences of each splicing result were selected.

Mitochondrial assembly results were integrated, and collinear analysis was performed using Mummer V3.1 [24] to determine the position relationship and fill the gap between contigs. Pilon V1.18 [25] was used to correct the results to obtain the final mitochondrial sequence. Then, we uploaded the complete mitogenome sequence to the MITOS (http://mitos.bioinf.uni-leipzig.de/) and MTannot (http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl) web servers for functional annotation [26]. All parameters were set to the defaults, except Genetic Code, which was set as 04-MOLD. The tRNA genes were identified and their potential secondary structures were inferred using MITOS online software. CGView visualization software was used to generate the whole genome circle map [27].

Phylogenetic analysis

We used 48 complete mitogenome sequences from Diptera and one Hymenoptera species as outgroups to construct phylogenetic relationships among families of Diptera. The nucleotide sequences of their complete mitogenome sequences were obtained from NCBI, and they were aligned using ClustalW in MEGA-X [28]. Maximum likelihood was used with Kimura 2-parameter model, and the number of

repeated runs was set to 1000 to construct the phylogenetic tree. The Bayesian method was modeled as GTR+I+G, and MrBayes 3 software was used to run the system for 200,000 generations using a burn-in of 25% generations.

Results

Mitochondrial genome structure

The mitogenome of *T. sintenisi* was 15763 bp long (GenBank accession number ID: MT795181) and it consisted of 13 PCGs, 22 tRNA genes, 2 rRNA genes, and 1 non-coding control region. Minor coding strand encoded 23 genes, comprising 9 PCGs and 14 tRNA genes. The majority strand contained four PCGs, eight tRNA genes, and two rRNA genes (Figure 1 and Table 1). The annotated sequence has been deposited into GenBank under the accession number MT795181.

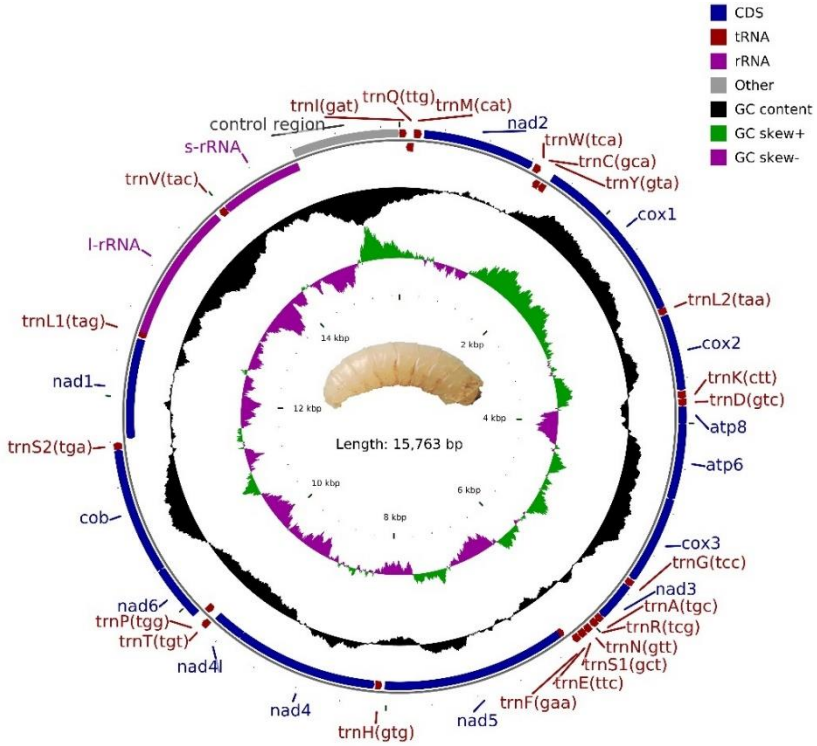


Fig. 1 Mitochondrial genome structure of *Tetanops sintenisi*. Genes outside the map are transcribed in a clockwise direction, whereas those inside the map are transcribed counterclockwise. The second circle shows the GC content and the third shows the GC skew.

Table 1. Organization of the mitochondrial genome of *Tetanops sintenisi*.

Feature	Strand	Position	Length(bp)	Start codon	Stop codon	Anticodon	Intergenic nucleotide
trnI	N	1-64	64			GAT	-3
trnQ	J	62-130	69			TTG	-1
trnM	N	130-198	69			CAT	21
nad2	N	220-1221	1002	ATT	TAA		18
trnW	N	1240-1306	67			TCA	-8
trnC	J	1299-1360	62			GCA	1
trnY	J	1362-1426	65			GTA	1
cox1	N	1428-2963	1536	CGA	TAA		-5
trnL2	N	2959-3023	65			TAA	4
cox2	N	3028-3712	685	ATG	T(AA)		3
trnK	N	3716-3786	71			CTT	-1
trnD	N	3786-3852	67			GTC	

atp8	N	3853-4014	162	ATC	TAA		-7
atp6	N	4008-4685	678	ATG	TAA		-1
cox3	N	4685-5473	789	ATG	TAA		5
trnG	N	5479-5542	64			TCC	
nad3	N	5543-5896	354	ATT	TAG		-2
trnA	N	5895-5958	64			TGC	-1
trnR	N	5958-6020	63			TCG	8
trnN	N	6029-6093	65			GTT	
trnS1	N	6094-6161	68			GCT	
trnE	N	6162-6226	65			TTC	18
trnF	J	6245-6308	64			GAA	-20
nad5	J	6289-8019	1731	ATT	TAG		24
trnH	J	8044-8107	64			GTG	6
nad4	J	8114-9446	1333	ATG	T(AA)		-7
nad4l	J	9440-9736	297	ATG	TAA		2
trnT	N	9739-9802	64			TGT	
trnP	J	9803-9867	65			TGG	23
nad6	N	9891-10394	504	ATT	TAA		-1
cob	N	10394-11530	1137	ATG	TAG		-2
trnS2	N	11529-11595	67			TGA	18
nad1	J	11614-12550	937	ATA	T(AA)		10
trnL1	J	12561-12625	65			TAG	-15
rrnL	J	12611-13912	1302				32
trnV	J	13945-14016	72			TAC	-1
rrnS	J	14016-14804	789				0
OH	N	14805-15763	959				0

Gene spacing or overlap was found in 37 genes of the whole mitogenome except the control region. There were 16 intergenic spacers, 194 bp in total, among which the longest intergenic spacer was 32 bp and located between *rrnL* and *trnV*. There were 15 gene overlaps, a total of 75 bp, and the longest overlap (20 bp) was between *trnF* and *nad5*(Table 1). The number of gene spacer regions was larger than that of the overlap regions and most spacer regions were longer than the overlap regions (Figure 1, Table 1).

The nucleotide composition of the genome was A: T: C: G = 38.2%: 31.0%: 19.5%: 11.3%. The A+T content of the whole genome was 69.2%, higher than that of G+C (30.8%), and the A+T content of all protein-codingand rRNA genes in the genome was higher than the G+C content (Table 2). There was a high A+T content bias in the mitogenome sequence, and the highest A+T content was detected in the control area (80.6%), which was much higher than that in the whole sequence; this result was consistent with the characteristics of base nucleotide composition bias in the insect control area.

Table 2. Nucleotide composition of mitochondrial genome.

Region	A%	C%	G%	T%	A+T%	G+C%	AT skew	GC skew
Whole genome	38.2	19.5	11.3	31.0	69.2	30.8	0.103	-0.267
nad2	31.5	19.9	10.8	37.8	69.3	30.7	-0.091	-0.296
cox1	28.3	22.3	17.0	32.4	60.7	39.3	-0.069	-0.132
cox2	31.8	21.3	14.2	32.7	64.5	35.5	-0.014	-0.202
atp8	38.9	17.3	6.8	37.0	75.9	24.1	0.024	-0.436
atp6	30.4	23.6	12.8	33.2	63.6	36.4	-0.044	-0.296
cox3	28.8	24.3	16.2	30.7	59.5	40.5	-0.032	-0.200
nad3	33.0	22.6	10.2	34.2	67.2	32.8	-0.017	-0.379
nad5	24.7	10.8	20.1	44.4	69.1	30.9	-0.286	0.301
nad4	23.3	10.7	20.0	46.0	69.3	30.7	-0.327	0.301
nad4l	25.9	8.7	17.2	48.2	74.1	25.9	-0.300	0.325
nad6	37.7	18.7	7.5	36.1	73.8	26.2	0.022	-0.424
cob	31.3	23.8	14.1	30.8	62.1	37.9	0.008	-0.258
nad1	20.8	9.8	20.9	48.5	69.3	30.7	-0.399	0.361
rrnL	33.3	7.1	15.1	44.5	77.8	22.2	-0.144	0.359
rrnS	33.5	9.1	18.4	39.0	72.5	27.5	-0.077	0.336
OH	42.7	12.1	7.3	37.9	80.6	19.4	0.061	-0.247

Protein-coding genes

The total sequence length of 13 PCGs was 11145 bp, among which the longest gene was *nad5* (1731

bp) and the shortest gene was *atp8* (162 bp), located in the J-strand and N-strand, respectively. The start codon of *cox2*, *atp8*, *atp6*, *cox3*, *nad4*, *nad4l*, and *cob* was typically ATG. The start codon of *nad2*, *nad3*, *nad5*, and *nad6* was ATT, and those of the remaining genes *cox1* and *nad1* were CGA and ATA, respectively (Table 1). The stop codon of *nad2*, *cox1*, *atp8*, *atp6*, *cox3*, *nad4l*, and *nad6* was TAA, while that of *nad3*, *nad5*, and *cob* was TAG. The *cox2*, nad4, and *nad1* had an incomplete stop codon T (AA). Among the 13 PCGs, only three genes (*atp8*, *nad6*, and cob) showed slight A-bias, ranging from 0.008 to 0.024. The rest of the PCGs were T-skewed. There were four genes with G-bias, namely, *nad5*, *nad4*, *nad4l*, and *nad1*, which ranged from 0.301 to 0.361. The remaining PCGs showed C-bias (Table 2). Compared with the mitogenomes of other related species in Diptera, the A+T content of the PCGs in *T. sintenisi* was the lowest at 66.3%; the A+T content in four other related species was more than 70%; and the highest A+T content was in allium fly (76.5%) (Table 3).

Table 3. Mitochondrial genome comparison of five species in Diptera.

Species		***Tetanops sintenisi***	***Bactrocera oleae***	***Ceratitis capitata***	***Delia antiqua***	***Anopheles funestus***
GeneBank accession number		MT795181	NC_005333.1	NC_000857.1	KT026595	MF775371
Whole genome	Size (bp)	15763	15815	15980	16141	15403
	A+T (%)	69.2	72.6	77.5	87.5	78.4
Protein- coding genes	Size (bp)	11145	11189	11183	11188	11080
	A+T (%)	66.3	70.2	75.6	76.5	76.6
rrnL	Size (bp)	1302	1323	1335	1330	1329
	A+T (%)	77.8	79.0	81.8	82.3	83.4
rrnS	Size (bp)	789	793	788	784	791
	A+T (%)	72.5	74.0	77.7	78.3	80.1
A+T-rich region	Size (bp)	959	949	1004	1266	570
	A+T (%)	80.6	86.9	91.1	93.7	94.1

MEGA-X software was used to analyze the relative synonymous codon usage (RSCU) and distribution of 13 PCGs in the mitogenome of *T. sintenisi* and its four related species (Figures 2 and 3). Among the five Diptera species, the most frequently employed amino acid was Ser2. The loss of synonymous codons was observed in the PCGs of *A. funestus*, *D. antiqua*, and *C. capitata*. The CUC encoding Leu1 was lost in all the three species, the CGC codon encoding Arg and the CUG encoding Leu1 were lost in both *A. funestus* and *D. antiqua*, and the AGG encoding Ser1 was lost in both *A. funestus* and *C. capitata*. In addition, the CUG encoding Leu1 was lost in *A. funestus* and *D. antiqua*, and the CCG encoding Pro was lost in *C. capitata*. Based on these findings, it is evident that the lost codons always ended in C or G. Although no loss of synonymous codons was found in *T. sintenisi* and *Bactrocera oleae*, the content of the above codons was very low (Figure 2). Additionally, it is evident from Figure 2 that the codons with high frequency mostly ended in A or U (the part shown in blue and green in the figure). Therefore, the use of synonymous codons confirmed the biased characteristics of A+T content in insect mitogenome. The distribution trend of the relative synonymous codons of 13 PCGs in *T. sintenisi* and its four related species was approximately the same (Figure 3). Although the mitochondrial PCGs of these five Diptera species had a preference in codon usage, the difference in preference was not significant, indicating that the mitogenome of Diptera species had a certain degree of conservation in the evolutionary process. This could indirectly reflect that more closely related species have more similar pattern of codon use.

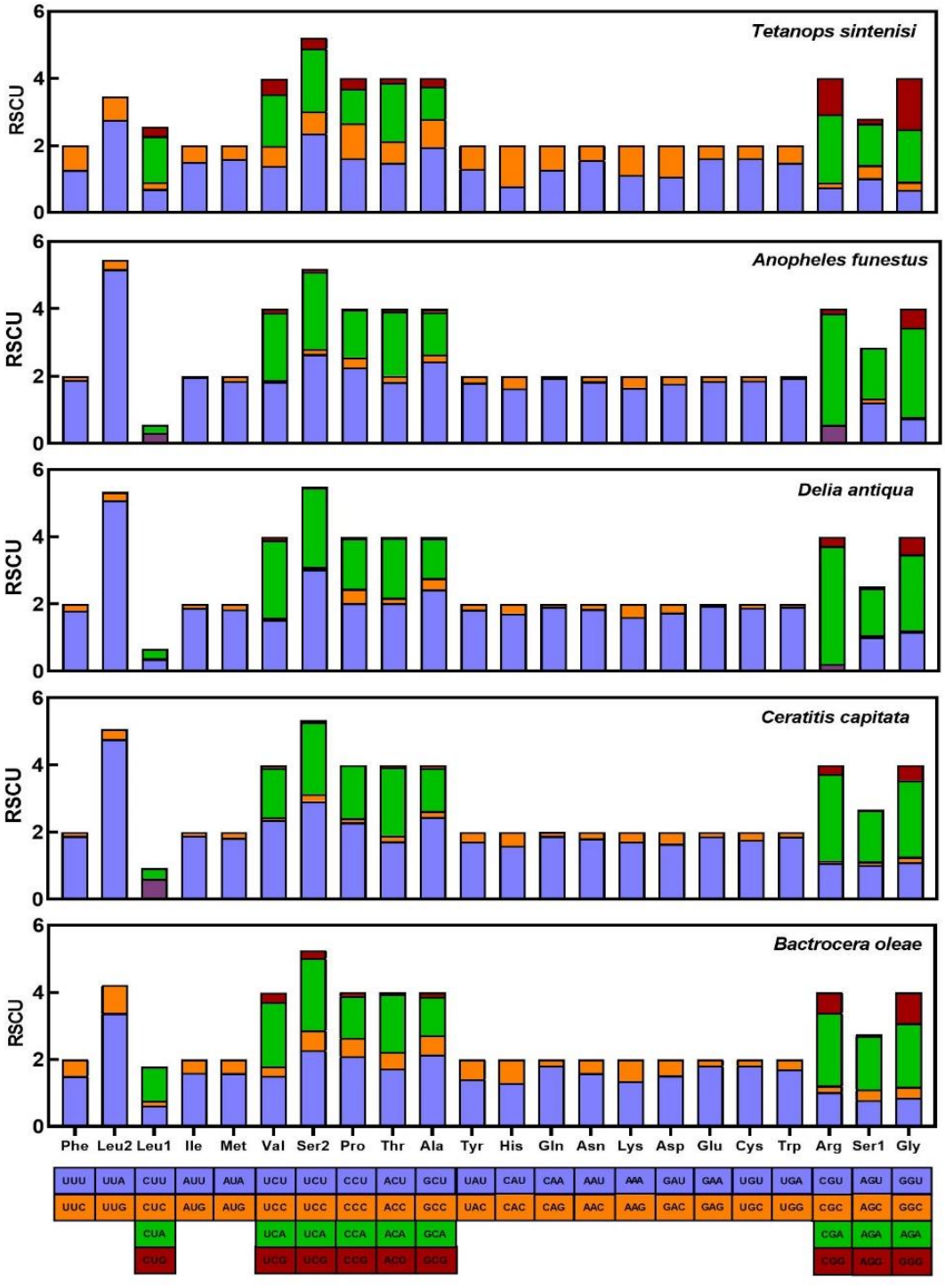


Fig. 2 The relative synonymous codon usage (RSCU) of the mitochondrial genome across five superfamilies in Diptera.

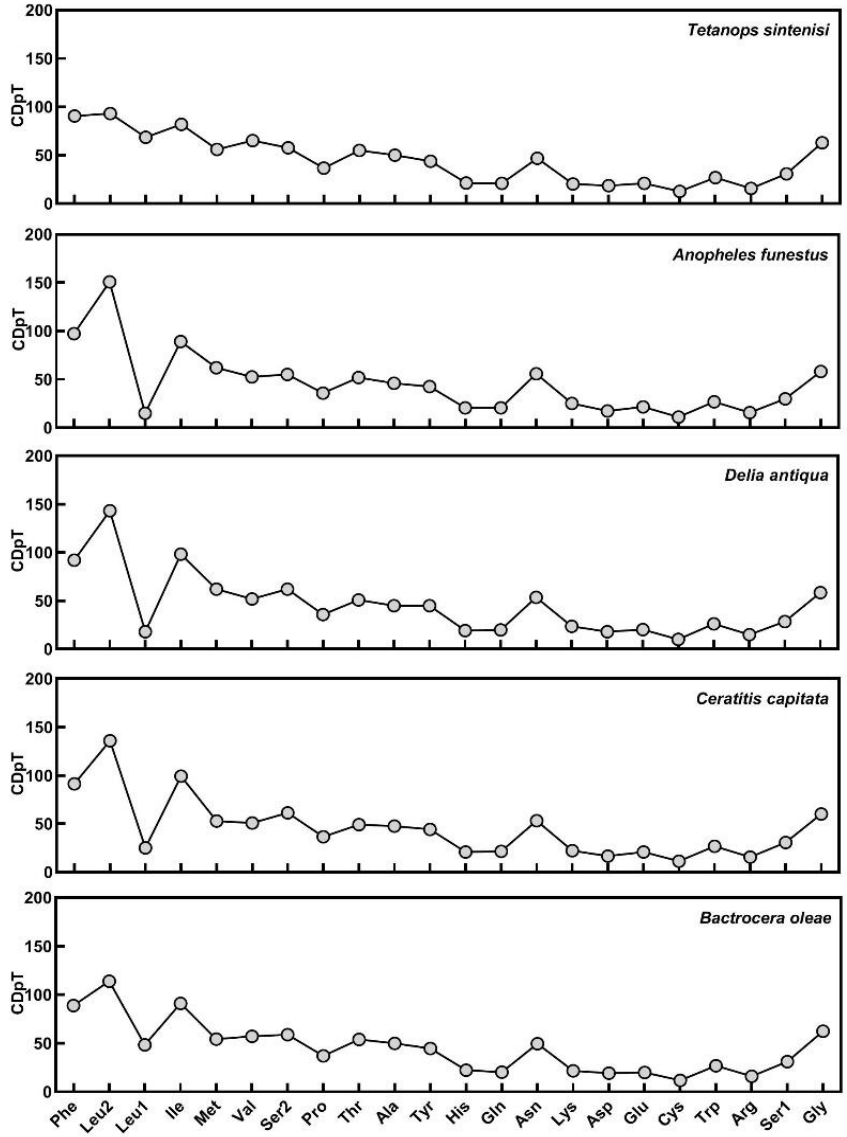


Fig. 3 Codon distribution in Diptera. CDpT, codons per thousand codons.

Transfer RNA and ribosomal RNA genes

The 22 tRNA genes of the mitogenome had a sequence length of 62-72 bp. *trnC* was the shortest tRNA gene with 62 bp, while *trnV* was the longest with 72 bp. Among the 22 tRNA genes, 14 tRNA genes were located in the N-strand and 8 tRNA genes were located in the J-strand. Twenty-three mismatched base pairs were identified in the tRNA genes, and these were all G-U pairs, comprising seven pairs located in the DHU arm of *trnQ*, *trnY*, *trnG*(2), *trnH*, *trnP*, and *trnF*, nine pairs located in the acceptor stem of *trnC*, *trnY*, *trnS1*, *trnE*, *trnH*(2), *trnP*, *trnF*, and *trnA*, five pairs located in the anticodon arm of *trnE*, *trnH*, *trnT*, *trnF*, and *trnV*, and two pairs located in the TψC arm of *trnP* and *trnF* (Figure 4).

The two rRNA genes *rrnL* and *rrnS* were separated by *trnV*. The length of the *rrnL* sequence was 1302 bp, the intergenic spacer between *rrnL* and *trnV* was 32 bp, and the A+T content was 77.8%. The length of the *rrnS* sequence was 789 bp, which overlapped with *trnV* by 1 bp, and the A+T content was 72.5% (Tables 1 and 2). The length of *rrnL* of *T. sintenisi* was the shortest among the five relative species and its A+T content was the lowest. The length of *rrnS* was the shortest in *D.antiqua* (784 bp), while the *rrnS* of *T. sintenisi* was only five bases longer than that of *D. antiqua*, and the A+T content of *rrnS* was still the lowest in *T. sintenisi* (Table 3).

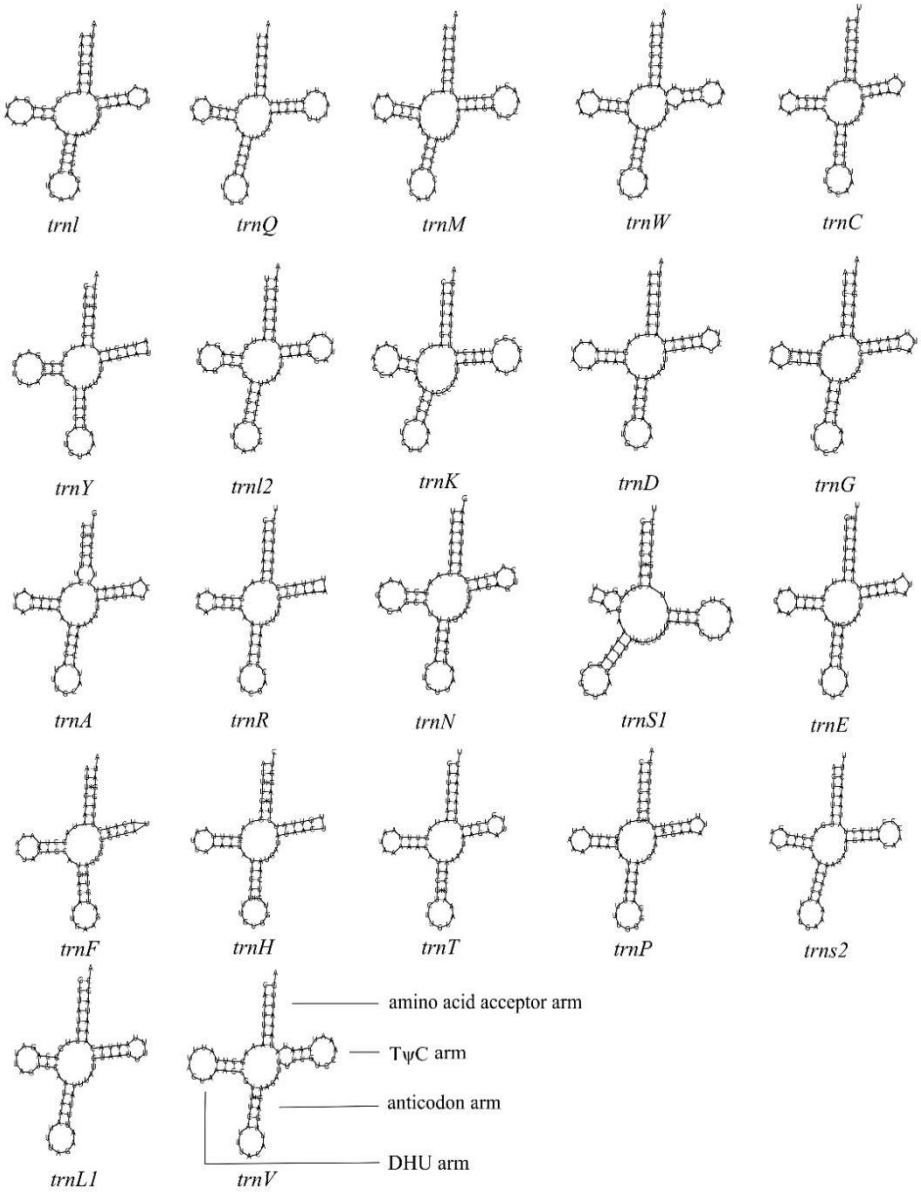


Fig. 4 Predicted secondary structure of tRNA genes in the mitochondrial genome of *Tetanops sintenisi*. The tRNAs are labeled with the abbreviations of their corresponding amino acids. Structural elements in tRNA arms and loops are illustrated as for *trnV*. Dashes (-) indicate Watson-Crick bonds and stars (*) indicate mistaken bonds.

Control region

The control region was the main non-coding region in the mitogenome sequence of *T. sintenisi*. The control region of *T. sintenisi* was 959 bp, located between *rrnS* and *trnI*, with the A, T, C, and G base contents of 42.7%, 37.9%, 12.1%, and 7.3%, respectively, and the A+T content of 80.6%, thereby showing significant AT base bias (Tables 1 and 2). The control region included a total of nine (TA)n regions, of which five were $(TA)_3$, two were $(TA)_4$, one was $(TA)_5$, and one was $(TA)_6$. Moreover, there were six poly-T regions with a length of 5 bp or more, among which the longest was 17 bp. In addition, there were two repeat elements with a length of 13 bp at 14868 bp and 15037 bp, and an interval of 156 bp between the repeat elements. There were two repeat elements with a length of 17 bp at 15305 bp and 15357 bp, and an interval of 35 bp between the repeat elements (Figure 5).

The control region length of these five related species ranged from 570 bp to 1266 bp, with that of *T. sintenisi* being intermediate. The A+T content in the control region ranged from 80.6% to 94.1%, and the A+T content in *T. sintenisi* was the lowest (Table 3).

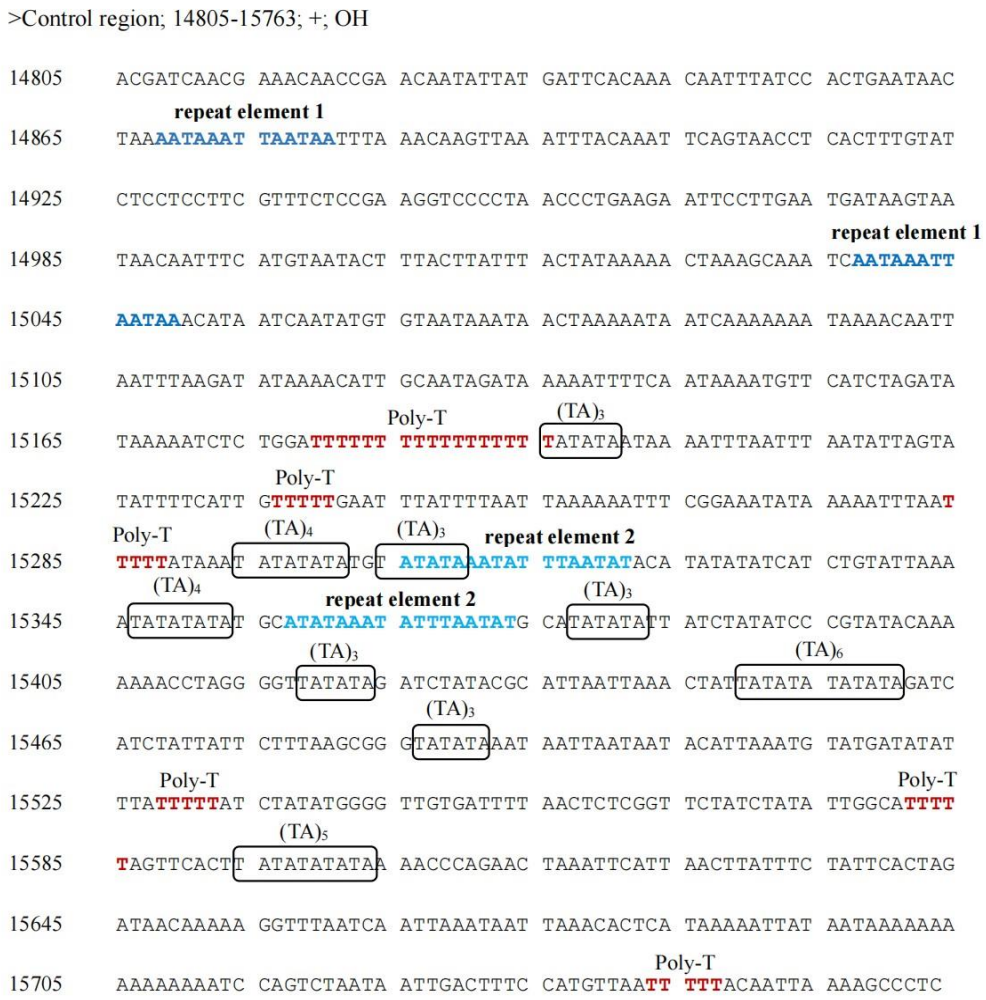
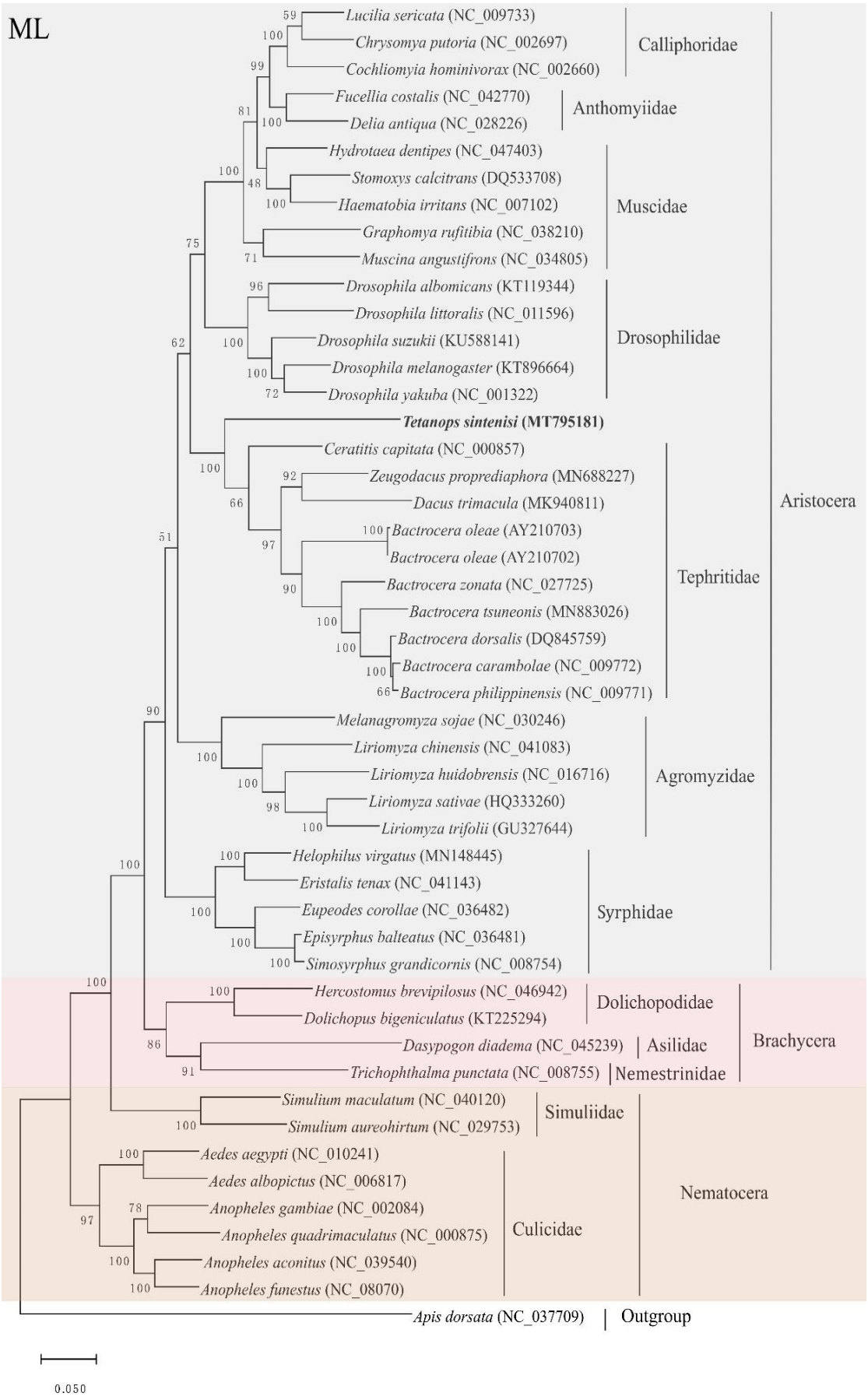


Fig. 5 Structure of the control region of the mitochondrial genome of *Tetanops sintenisi*. The (TA)nis marked with black box, the poly-T stretch is shown in red, and the repeat elements are shown in blue.

Phylogenetic relationships

In the traditional morphological classification, *T. sintenisi* is placed within Ulidiidae, but there are no reports on the complete mitogenome sequence of Ulidiidae species. Therefore, phylogenetic trees based on Maximum likelihood (ML) and Bayesian inference (BI) were constructed using complete mitogenome sequences of 48 species belonging to argonauts and other Diptera, with Apisdorsata (Hymenoptera) as the outgroup. The results showed that the topological structures of phylogenetic trees constructed with the two methods were basically the same, and the 48 Diptera species were grouped into three clusters (Aristocera, Nematocera, and Brachycera). Phylogenetic relationships were slightly different among the Aristocera. The phylogenetic relationship among Aristocera families constructed using ML method was: ((((((Calliphoridae + Anthomyiidae) + Muscidae) + Drosophilidae) + (*T. sintenisi* + Tephritidae)) + Agromyzidae) + Syrphidae) (Figure 6), and that constructed using BI method was: (((*T. sintenisi* + Tephritidae) + Drosophilidae) + (((Muscidae + (Calliphoridae + Anthomyiidae)) + Agromyzidae) + Syrphidae)) (Figure 6). However, in both the phylogenetic trees, Ulidiidae and Tephritidae were clustered together, indicating that the two families have a close relationship.


ML
Lucilia sericata (NC_009733)
Chrysomya putoria (NC_002697)
Cochliomyia hominivorax (NC_002660)
Calliphoridae
Fucellia costalis (NC_042770)
Delia antiqua (NC_028226)
Anthomyiidae
Hydrotaea dentipes (NC_047403)
Stomoxys calcitrans (DQ533708)
Haematobia irritans (NC_007102)
Graphomya rufitibia (NC_038210)
Muscina angustifrons (NC_034805)
Muscidae
Drosophila albomicans (KT119344)
Drosophila littoralis (NC_011596)
Drosophila suzukii (KU588141)
Drosophila melanogaster (KT896664)
Drosophila yakuba (NC_001322)
Drosophilidae
Tetanops sintenisi (MT795181)
Ceratitis capitata (NC_000857)
Zeugodacus propredіaphora (MN688227)
Dacus trimacula (MK940811)
Bactrocera oleae (AY210703)
Bactrocera oleae (AY210702)
Bactrocera zonata (NC_027725)
Bactrocera tsuneonis (MN883026)
Bactrocera dorsalis (DQ845759)
Bactrocera carambolae (NC_009772)
Bactrocera philippinensis (NC_009771)
Tephritidae
Melanagromyza sojae (NC_030246)
Liriomyza chinensis (NC_041083)
Liriomyza huidobrensis (NC_016716)
Liriomyza sativae (HQ333260)
Liriomyza trifolii (GU327644)
Agromyzidae
Helophilus virgatus (MN148445)
Eristalis tenax (NC_041143)
Eupeodes corollae (NC_036482)
Episyrphus balteatus (NC_036481)
Simosyrphus grandicornis (NC_008754)
Syrphidae
Aristocera
Hercostomus brevipilosus (NC_046942)
Dolichopus bigeniculatus (KT225294)
Dolichopodidae
Dasypogon diadema (NC_045239)
Asilidae
Trichophthalma punctata (NC_008755)
Nemestrinidae
Brachycera
Simulium maculatum (NC_040120)
Simulium aureohirtum (NC_029753)
Simuliidae
Aedes aegypti (NC_010241)
Aedes albopictus (NC_006817)
Anopheles gambiae (NC_002084)
Anopheles quadrimaculatus (NC_000875)
Anopheles aconitus (NC_039540)
Anopheles funestus (NC_08070)
Culicidae
Nematocera
Apis dorsata (NC_037709)
Outgroup
0.050

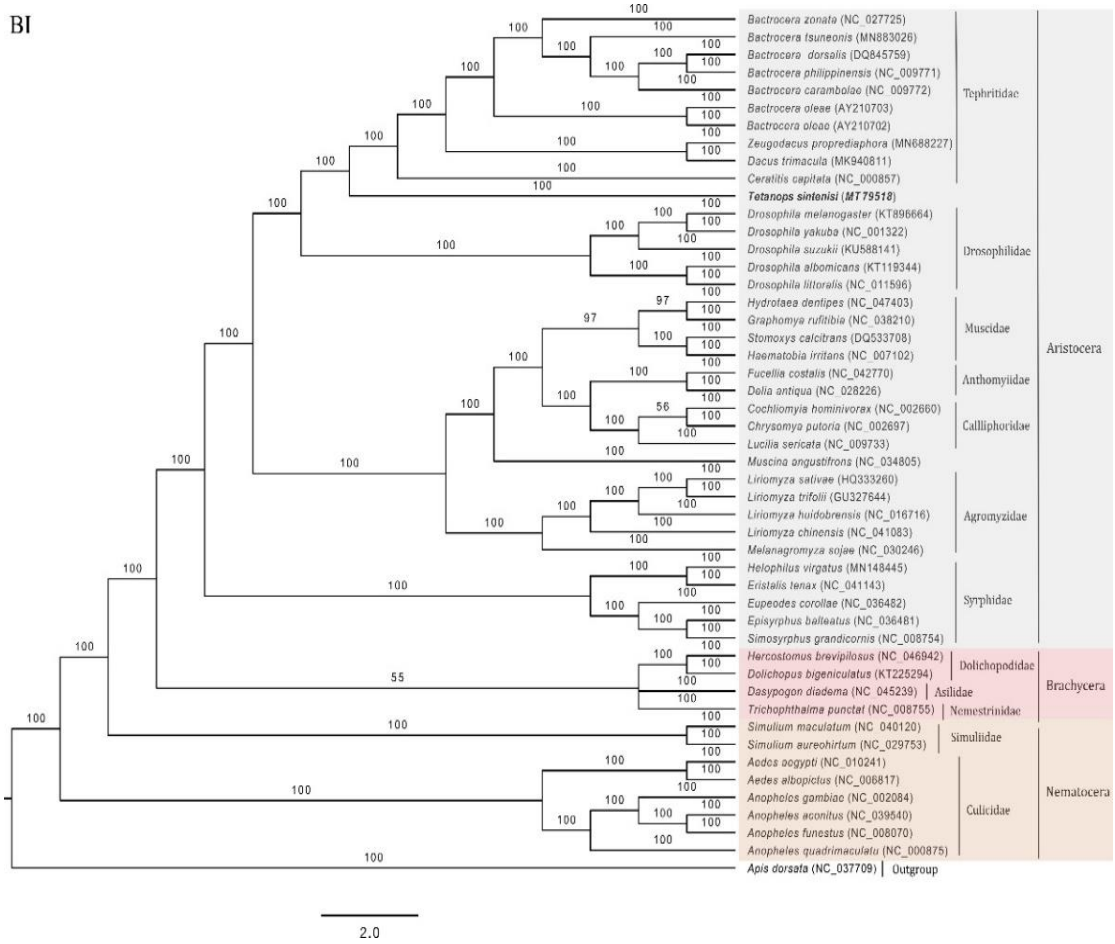


Fig. 6 Phylogenetic relationships of 48 Diptera, including *Tetanops sintenisi*, based on the whole-genome sequence data. The numbers beside the nodes are bootstrap values (maximum likelihood) and posterior probabilities (Bayesian inference), respectively.

Discussion

The mitogenome sequence of *T. sintenisi* was consistent with that of *D. antiqua*, *D.melanogaster*, and *B. oleae*, which are other known Diptera species [4, 29, 30], and exhibited the typical structure observed in Diptera. In evolution, rearrangement in the mitogenome of Diptera is relatively rare [4], which indicates the evolutionary conservatism of the mitogenome in Diptera. In the mitogenome of *T. sintenisi*, all the 12 PCGs contained the typical start codon ATN, but that of *cox1* was CGA (Table 1). The start codon of *cox1* is diverse in insects, it is a tetranucleotide (TTAG, ATAA, ATTA) [14, 31, 32], hexanucleotide (TATTAG, TTTTAG) [33], CGA [34], or TTG [35]. Although, it is commonly nearly identical among related insect groups, such as those of Lepidoptera [34] and Polyphaga [36], the present study found that start codon of *cox1* in *T. sintenisi* was not consistent with that in the known Diptera species (Table 4).

Table 4. Classification of the start codons of *cox1*.

Type	Codon	Species	References
Triplet	CGA	*Tetanops sintenisi* *Limenitis helmanni*	Wang et al., 2017[34]
	TTG	Rhyparochromidae *Haematobia irritans*	Li et al., 2016[35] Oliveira et al., 2005[37]
	TCG	*Stomoxy scalcitrans* *Musca domestica* *Bactrocera oleae*	Nardi et al., 2003[30]
Tetranucleotide	TTAG	*Coreana raphaelis*	Kim et al., 2006[32]
	ATAA	*Drosophila yakuba*	Clary and Wolstenholme, 1985[14]
	ATTA	*Drosophila melanogaster*	De Bruijn, 1983[31]

	ATCA	*Delia antiqua* *Liriomyza sativae*	Zhang et al., 2015[4] Yang et al., 2011[38]
Hexanucleotide	TATTAG	*Ostrinia nubilalis* *Ostrinia furnicalis*	Coates et al., 2005[33]
	TTTTAG	*Bombyx mori*	Coates et al., 2005[33]

The *cox2*, *nad4*, and *nad1* genes of *T. sintenisi* are terminated by a single T (Table 1). This partially coincides with other Diptera species (Table 5), suggesting that the usage of incomplete codons as stop codons is a common phenomenon in the Diptera [11, 14, 39], and verifying the characteristics of A+T content bias in the stop codon.

Table 5. Classification of the stop codons.

Species	Codon	Genes	References
Tetanops Sintenisi	T	*cox2*, *nad4*, *nad1*	
Bactrocera tsuneoni	T	*nad3*, *nad5*, *CYTB*	Yue et al., 2018[40]
	TA	*cox1*	
Bactrocera zonata	T	*nad3*, *nad5*, *nad1*	Jaipal et al., 2015[5]
	TA	*cox1*	
Neoceratitis asiatica	T	*nad1*	Su et al., 2017[41]
Liriomyza sativae	T	*nad2*, *CYTB*	Yang et al., 2011[38]
	TA	*nad5*	
Delia antiqua	T	*nad5*, *cox2*	Zhang et al., 2015[4]
	TA	*nad4*	

The location of the control region (A+T-rich region) in the mitogenome was relatively conserved but its length varied greatly from species to species. There were five common structures in the control region: a poly-thymidine stretch, a $[TA(A)]_n$-like stretch, a stem and loop structure, a $G(A)_nT$ structure, and a G+A-rich sequence block [42]. In the present study, the length of the control region in *T. sintenisi* was 959 bp, and many $(TA)_n$ and poly-T structures were found in the control region. This was slightly different from that of some Diptera species, such as *D. antiqua*, whose control region is 1266 bp and comprised of three tandem replicates, namely $(TA)_n$, and poly-T, which were speculated to be involved in the control of transcription or replication [4]. The length of the control region in *Liriomyza sativae* is 741 bp, including poly-T and $G(A)_nT$ structures [38]. This could be attributed to the difference in the number of copies of tandem repeatsin the control region between related species [42].This feature of high A+T content and length difference in the control region is of great significance for the study of molecular evolution.

In the ML- and BI-based phylogenetic tree topologies, the 48 Diptera species were divided into three clusters, Aristocera, Nematocera, and Brachycera, and both topologies strongly supported the close relationship between *T. sintenisi* and Tephritidae (Figure 5). This result is consistent with the traditional morphological taxonomy. However, to date, the complete mitogenome sequences of Ulidiidae have been limited; therefore, further mitogenome sequencing and in-depth analyses of Ulidiidae and other families is required in future studies to determine a more comprehensive evolutionary relationship among the families of Aristocera.

To the best of our knowledge, this is the first report of the complete mitochondrial genome in the family Ulidiidae. Comparative analysis showed that the gene size, gene order, base content, and base composition are comparatively conserved, similar to other dipteran mitochondrial genomes. ATN is the initiation codon in all the 13 PCGs, except for *cox1*, which starts with CGA. All tRNAs have the typical cloverleaf structure, and mismatched base pairs, all of which were G-U pairs, were identified in some tRNA genes. The location of the two rRNAs were conservative and the A+T content of the rRNAs was the lowest among the five related insect species. Two special structures were found in the control region, poly-T stretches and a $(TA)_n$ stretch, which are considered important elements related to replication and transcription. Both ML and BI analyses using nucleotide sequences of the whole mitogenome strongly suggest a closer relationship of *T. sintenisi* with Ulidiidae and Tephritidae. The whole mitogenome sequences have also been demonstrated as an effective method for resolving phylogenetic relationships [43]. The comparison of genomic features in related species should contribute to a comprehensive

understanding of the evolutionary process of Diptera. The sequencing of the *T. sintenisi* mitogenome provides not only an important opportunity for ecological identification of Ulidiidae, but also valuable information for future evolutionary and molecular studies of Ulidiidae.

Acknowledgements We thank Wenna Wang for assistance in completing the experiments. This research was financially supported by the Foundation of Shanxi Academy of Agricultural Sciences (YCX2018D2T02) and the Province Key Research and Development Program of Shanxi (201903D221014).

Data availability statement The genome sequence data that support the findings of this study are openly available in GenBank of NCBI at https://www.ncbi.nlm.nih.gov/genbank under the accession no. MT795181. The associated BioProject, SRA, and Bio-Sample numbers are PRJNA757578, SRR15725549, and SAMN20968555, respectively.

Author contributions Conceptualization, K.X.; Formal analysis, K.X. and C.K.; Funding acquisition, F.Z.; Methodology, K.X and C.K.; Project administration, F.Z.; Writing-original draft, K.X. and C.K.; Writing-review and editing, F.Z. All authors have read and agreed to the published version of the manuscript.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Zhang NX, Zhang YJ, Guo YU, Chen B (2013) Structure characteristics of the mitochondrial genomes of Diptera and design and application of universal primers for their sequencing. Acta Entomol Sin 56(4): 398-407. https://doi.org/10.16380/j.kcxb.2013.04.013
2. Li H, Li J (2019) The complete mitochondrial genome of *Helophilus virgatus* (Diptera: Syrphidae: Eristalinae) with a phylogenetic analysis of Syrphidae. Mitochondrial DNA Part B 4(2): 3106-3107. https://doi.org/10.1080/23802359.2019.1667890
3. Yan W, Shang Y, Ren L, Zhang X, Wang Y (2019) The complete mitochondrial genome of *Hydrotaea dentipes* (Diptera: Muscidae). Mitochondrial DNA Part B 4(1): 2044-2045. https://doi.org/10.1080/23802359.2019.1618217
4. Zhang NX, Guo Y, Li TJ, He QY, Zhou Y, Si FL, Shuang R, Chen B, Yue BS (2015) The complete mitochondrial genome of *Delia antiqua* and its implications in Dipteran Phylogenetics. PLoS ONE 10(10): e0139736. https://doi.org/10.1371/journal.pone.0139736
5. Choudhary JS, Naaz N, Prabhakar CS, Rao MS, Das B (2015) The mitochondrial genome of the peach fruit fly, *Bactrocera zonata* (Saunders) (Diptera: Tephritidae): complete DNA sequence, genome organization, and phylogenetic analysis with other tephritids using next generation DNA sequencing. Gene 569(2): 191-202. https://doi.org/10.1016/j.gene.2015.05.066
6. Wang T, Ren YL, Zhong Y, Yang MF (2020) Sequencing and analysis of the complete mitochondrial genome of *Bactrocera cheni* from China and its phylogenetic analysis. Mitochondrial DNA Part B 5(1): 780-781. https://doi.org/10.1080/23802359.2020.1715882
7. Mitchell SE, Cockburn AF, Seawright JA (1993) The mitochondrial genome of *Anopheles quadrimaculatus* species A: complete nucleotide sequence and gene organization. Genome 36(6): 1058-1073. https://doi.org/10.1139/g93-141
8. Spanos L, Koutroumbas G, Kotsyfakis M, Louis C (2010) The mitochondrial genome of the Mediterranean fruit fly, *Ceratitis capitata*. Insect Mol Biol 9(2): 139-144. https://doi.org/10.1046/j.1365-2583.2000.00165.x
9. Lessinger AC, Junqueira A, Lemos TA, Kemper EL, Azeredo-Espin A (2010) The mitochondrial genome of the primary screwworm fly *Cochliomyia hominivorax* (Diptera: Calliphoridae). Insect Mol Biol 9(5): 521-529. https://doi.org/10.1046/j.1365-2583.2000.00215.x

10. Stevens JR, West H, Wall R (2010) Mitochondrial genomes of the sheep blowfly, *Lucilia sericata*, and the secondary blowfly, *Chrysomya megacephala*. Med Vet Entomol 22(1): 89-91. https://doi.org/10.1111/j.1365-2915.2008.00710.x
11. Junqueira A, Lessinger AC, Torres TT, Silva F, Espin A (2004) The mitochondrial genome of the blowfly *Chrysomya chloropyga* (Diptera : Calliphoridae). Gene 339(1): 7-15. https://doi.org/10.1016/j.gene.2004.06.031
12. Beckenbach AT, Joy JB (2009) Evolution of the mitochondrial genomes of *Gall midges* (Diptera: Cecidomyiidae): rearrangement and severe truncation of tRNA genes. Genome Biol Evol 1(1): 278-287. https://doi.org/10.1093/gbe/evp027
13. Oliveira MT, Barau JG, Junqueira A, Feijo PC, Lessinger AC (2008) Structure and evolution of the mitochondrial genomes of *Haematobia irritans* and *Stomoxys calcitrans*: the Muscidae (Diptera: Calyptratae) perspective. Mol Phylogenet Evol 48(3): 850-857. https://doi.org/10.1016/j.ympev.2008.05.022
14. Clary DO, Wolstenholme DR (1985) The mitochondrial DNA molecular of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. J Mol Evol 22(3): 252-271. https://doi.org/10.1007/BF02099755
15. Smit JT (2005) De prachtvlieg *Tetanops sintenisi* nieuw voor Nederland (Diptera: Ulidiidae). Nederlandse Faunistische Mededelingen 22: 91-94.
16. Xing K, Fei Z, Zhao X, Hui Y, Zhou J, Hong L (2018) First report of root maggot *Tetanops sintenisi* on quinoa. China Plant Protection 38(12): 38-40.
17. Mortelmans J, Bree ED, Hendrix J (2012) Four new additions to the *Belgian fauna* (Diptera: Conopidae, Tabanidae, Sciomyzidae, Ulidiidae). Bulletin van de Koninklijke Belgische vereniging voor entomologie 148: 193-196.
18. Stuke JH, Merz B (2005) Drei für Deutschland neu nachgewiesene acalyptrate Fliegen (Diptera: Lauxaniidae, Pallopteridae, Ulidiidae). Studia Dipterologica 12: 242-254.
19. Stuke JH (2006) Bemerkenswerte Zweiflügler aus Niedersachsen und Bremen (Insecta: Diptera) - 1. Teil. Drosera 12: 67-72.
20. Stuke JH (2008) Die Tephritoidea (Diptera) Niedersachsens und Bremens. Abh. Naturwiss. Vereins Bremen 46(2): 329-356.
21. John TS, Gerrit T (2014) Een bijzondere vangst van een bijzondere prachtvlieg (Diptera, Ulidiidae). Entomologische Berichten 74(6): 261-263.
22. Coil D, Jospin G, Darling AE (2014) A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. Bioinformatics 31(4): 587-589. https://doi.org/10.1093/bioinformatics/btu661
23. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19(5): 455-477. https://doi.org/10.1089/cmb.2012.0021
24. Kurtz S, Phillippy A, Delcher AL, Smoot M (2004) Versatile and open software for comparing large genomes. Genome Biol 5(2): R12. https://doi.org/10.1186/gb-2004-5-2-r12
25. Walker BJ, Abeel T, Shea T, Priest M, Earl AM (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE 9(11): e112963. https://doi.org/10.1371/journal.pone.0112963
26. Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M, Stadler PF (2013) MITOS: improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol 69(2): 313-319. https://doi.org/10.1016/j.ympev.2012.08.023

27. Stothard P, Wishart DS (2005) Circular genome visualization and exploration using CGView. Bioinformatics 21(4): 537-539. https://doi.org/10.1093/bioinformatics/bti054
28. Sudhir K, Glen S, Li M, Christina K, Koichiro T (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35(6): 1547-1549. https://doi.org/10.1093/molbev/msy096
29. Lewis DL, Farr CL, Kaguni LS (2010) *Drosophila melanogaster* mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons. Insect Mol Biol 4(4): 263-278. https://doi.org/10.1111/j.1365-2583.1995.tb00032.x
30. Nardi F, Carapelli A, Dallai R, Frati F (2003) The mitochondrial genome of the olive fly *Bactrocera oleae*: two haplotypes from distant geographical locations. Insect Mol Biol 12(6): 605-611. https://doi.org/10.1046/j.1365-2583.2003.00445.x
31. Brujn MHLD (1983) Drosophila melanogaster mitochondrial DNA, a novel organization and genetic code. Nature 304(5923): 234-241. https://doi.org/10.1038/304234a0
32. Kim I, Lee EM, Seol KY, Yun EY, Jin BR (2010) The mitochondrial genome of the Korean hairstreak, *Coreana raphaelis* (Lepidoptera: Lycaenidae). Insect Mol Biol 15(2): 217-225. https://doi.org/10.1111/j.1365-2583.2006.00630.x
33. Coates BS, Sumerford DV, Hellmich RL, Lewis LC (2005) Partial mitochondrial genome sequences of *Ostrinia nubilalis* and *Ostrinia furnicalis*. Int J Biol Sci 1(1): 13-18. https://doi.org/10.7150/ijbs.1.13
34. Wang JP, Cao TW, Zhang Y, Fan RJ, Zhang M, Shi BM, Peng FC (2017) Sequencing and analysis of the complete mitochondrial genome of *Limenitis helmanni* (Lepidoptera: Nymphalidae). Acta Entomol Sin 60(8): 950-961. https://doi.org/10.16380/j.kcxb.2017.08.012
35. Li T, Yang J, Li Y, Cui Y, Xie Q, Bu W, Hillis D M (2016) A mitochondrial genome of Rhyparochromidae (Hemiptera: Heteroptera) and a comparative analysis of related mitochondrial genomes. Sci Rep 6: 35175. https://doi.org/10.1038/srep35175
36. Sun H, Zhao W, Lin R, Zhou Z, Huai W, Yao Y (2020) The conserved mitochondrial genome of the jewel beetle (Coleoptera: Buprestidae) and its phylogenetic implications for the suborder Polyphaga. Genomics 112(5): 3713-3721. https://doi.org/10.1016/j.ygeno.2020.04.026
37. Oliveira M, Lessinger AC (2005) Evolutionary and structural analysis of the cytochrome c oxidase subunit I (*COI*) gene from , and (Diptera: Muscidae) mitochondrial DNA. Mitochondrial DNA Part A 16(2): 156-160. https://doi.org/10.1080/10425170500039901
38. Yang F, Du YZ, Wang LP, Cao JM, Yu WW (2011) The complete mitochondrial genome of the leafminer *Liriomyza sativae* (Diptera: Agromyzidae): great difference in the A+T-rich region compared to *Liriomyza trifolii*. Gene 485(1): 7-15. https://doi.org/10.1016/j.gene.2011.05.030
39. Jin SB, Kim I, Sohn HD, Jin BR (2004) The mitochondrial genome of the firefly, *Pyrocoelia rufa*: complete DNA sequence, genome organization, and phylogenetic analysis with other insects. Mol Phylogenet Evol 32(3): 978-985. https://doi.org/10.1016/j.ympev.2004.03.009
40. Yue Z, Feng S, Zeng Y, Hong N, Liu L, Zhao Z, Fan J, Li Z (2018) The first complete mitochondrial genome of *Bactrocera tsuneonis* (Miyake) (Diptera: Tephritidae) by next-generation sequencing and its phylogenetic implications. Int J Biol Macromol 118(15): 1229-1237. https://doi.org/10.1016/j.ijbiomac.2018.06.099
41. Su Y, Zhang Y, Feng S, He J, Zhao Z, Bai Z, Liu L, Zhang R, Li Z (2017) The mitochondrial genome of the wolfberry fruit fly, *Neoceratitis asiatica* (Becker) (Diptera: Tephritidae) and the phylogeny of *Neoceratitis Hendel* genus. Sci Rep 7(1): 16612. https://doi.org/10.1038/s41598-017-16929-7

42. Zhang DX, Hewitt GM (1997) Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. Biochem Syst Ecol 25(2): 99-120. https://doi.org/10.1016/S0305-1978(96)00042-7

43. Brian MW, Michelle DT, Isaac SW, Norman BB, Jung-Wook K, Christine L, Matthew AB, Brian KC, Keith MB, Alysha MH, Benjamin MW, Kevin JP, Thomas P, Bradley JS, Jeffrey HS, Vladimir B, Jason C, Sujatha NK, Urs S-O, Gail EK, F Christian T, David AG, Andrew TB, Gregory WC, Markus F, Rudolf M, David KY (2011) Episodic radiations in the fly tree of life. PNAS 108(14): 5690-5695. https://doi.org/10.1073/pnas.1012675108