RESEARCH ARTICLE

The Complete Mitochondrial Genome of Delia antiqua and Its Implications in Dipteran Phylogenetics

Nai-Xin Zhang, Guo Yu, Ting-Jing Li, Qi-Yi He, Yong Zhou, Feng-Ling Si, Shuang Ren, Bin Chen*

Institute of Entomology and Molecular Biology, College of Life Sciences, Chongqing Normal University, Chongqing, 401331, China

* bin.chen@cqnu.edu.cn

Abstract

Delia antiqua is a major underground agricultural pest widely distributed in Asia, Europe and North America. In this study, we sequenced and annotated the complete mitochondrial genome of this species, which is the first report of complete mitochondrial genome in the family Anthomyiidae. This genome is a double-stranded circular molecule with a length of 16,141 bp and an A+T content of 78.5%. It contains 37 genes (13 protein-coding genes, 22 tRNAs and 2 rRNAs) and a non-coding A+T rich region or control region. The mitochondrial genome of Delia antiqua presents a clear bias in nucleotide composition with a positive AT-skew and a negative GC-skew. All of the 13 protein-coding genes use ATN as an initiation codon except for the COI gene that starts with ATCA. Most protein-coding genes have complete termination codons but COII and ND5 that have the incomplete termination codon T. This bias is reflected in both codon usage and amino acid composition. The protein-coding genes in the D. antiqua mitochondrial genome prefer to use the codon UUA (Leu). All of the tRNAs have the typical clover-leaf structure, except for tRNAser(AGN) that does not contain the dihydrouridine (DHU) arm like in many other insects. There are 7 mismatches with U-U in the tRNAs. The location and structure of the two rRNAs are conservative and stable when compared with other insects. The control region between 12S rRNA and tRNAIle has the highest A+T content of 93.7% in the D. antiqua mitochondrial genome. The control region includes three kinds of special regions, two highly conserved poly-T stretches, a (TA)n stretch and several G(A)nT structures considered important elements related to replication and transcription. The nucleotide sequences of 13 protein-coding genes are used to construct the phylogenetics of 26 representative Dipteran species. Both maximum likelihood and Bayesian inference analyses suggest a closer relationship of D. antiqua in Anthomyiidae with Calliphoridae, Calliphoridae is a paraphyly, and both Oestroidea and Muscoidea are polyphyletic.
Introduction

The mitochondrion is an important organelle in eukaryotic cells. It is connected with energy metabolism, apoptosis, aging, and disease and is a location for oxidative phosphorylation [1]. The mitochondrion is known as the cell's "powerhouse" or "power station" because it provides energy for cells through oxidative phosphorylation. The growth and proliferation of mitochondria are controlled by both the nuclear genome and its own genome, so it is considered a semi-autonomous organelle [2].

The mitochondrial genome is a covalently closed circular double-stranded molecule with a small molecular weight. It has a high copy number, does not contain introns, has a compact gene arrangement, and is lack of recombination [3]. There are significant differences in the size of the mitochondrial genome among different organisms. The insect mitochondrial genome is 13–19 kb in length and is composed of an encoding region containing 37 genes (13 protein coding genes, 22 tRNA genes and 2 rRNA genes) and a non-coding A+T rich region. The non-coding A+T rich region, also called as the control region (CR), is considered to control the replication and transcription of the mitochondrial genome [4]. The length variation among insect mitochondrial genomes is mainly determined by variation in the A+T rich region, which varies from 70 to 13 kb in length [5].

The mitochondrial genome is widely reported for its difference from the nuclear genome in its nucleotide composition, codon usage, gene sequencing and tRNA secondary structure [6–8]. Mitochondrial genomes are widely used in phylogenetics as well as in the study of the comparative and evolutionary genomics of insects. Mitochondrial genomes are also ideal molecular markers in population genetics and molecular evolution. All of these are due to mitochondria having a matrilineal inheritance, lack of extensive recombination, a conservative gene structure and composition, a low mutation rate and a faster evolution than nuclear genomes [9–10]. In recent years, partial coding genes of the mitochondrial genome, such as COI, COII, have become widely used in molecular phylogenetic analysis. The genome order has also been used as genetic markers to solve the phylogenetic relationships among distantly related taxa [11].

Insects exhibit the most extensive range of taxa on the planet, and insects have also been the subject of more research than other species. To date, there are more than 480 insect mitochondrial genome sequences published, among which there are 77 complete or nearly complete sequences from Diptera [12], accounting for 16% of the total sequences. These dipteran mitochondrial genome sequences provide an important database reference and are the basis for new molecular phylogenetic analyses of insects.

The onion maggot *Delia antiqua*, belonging to the family Anthomyiidae in the superfamily Muscoidea, is a major underground agricultural pest with wide distributed in Asia, Europe and North America. Its larvae damage bulb onions, garlic, chives, shallots, leeks and the bulbs of tulips, and reside in rotting liliaceous vegetables [13]. It naturally enters diapause in the pupal stage in summer or winter seasons just after the head evagination completed, and can serve a good model for insect diapause study [14]. To date, the mitochondrial genome sequence of this species has not been available. The Muscoidea was considered to be a paraphylly and the superfamily Oestroidea was nested within the Muscoidea. The phylogenetic relation of the two superfamilies and the location of Anthomyiidae are still not resolved [15–17].

In this study, we report the complete mitochondrial genome sequence, and investigate the organization, composition, codon usage and RNA secondary structure of the *Delia antiqua* and kown dipteran mitochondrial genomes. Importantly, this is the first report and description of complete mitochondrial genome of the family Anthomyiidae. We constructed the phylogenetic relationship of 26 representative species of known dipteran mtgenomes, and provide new
insight in the phylogenetics of the two superfamilies. We found that Anthomyiidae was claded in Calliphoridae in the Oestroidea.

Materials and Methods

Sampling and DNA Extraction

*Delia antiqua* colony was reared in the Institute of Entomology and Molecular Biology, Chongqing Normal University, China at 20 ± 0.2°C under 50–70% relative humidity with a 16L:8D photocycle as previously described [13]. The mitochondrial genomic DNA was extracted from the third instar of larvae with the TIANamp Genomic DNA Kit (TianGen, China).

PCR Amplification and Sequencing

The mitochondrial genome of *D. antiqua* was amplified by overlapping short PCR fragments (<1.2kb) with the extracted genomics DNA. All 26 fragments were amplified using the universal primers for Diptera designed by Zhang et al [18]. All short PCRs were carried out using Takara rTaq DNA polymerase (Takara, China) under the following cycling conditions: denaturation at 94°C for 5 min, followed by 35 cycles of denaturation at 94°C for 40 s, annealing at 48–55°C for 45 s, and elongation at 72°C for 1 min. The final elongation step was continued for 10 min at 72°C. These PCR products were analyzed by 1.0% agarose gel electrophoresis. All amplified products were sequenced directly except for the control region, which was sequenced after cloning into pMD-19T Vector. All fragments were sequenced in both directions.

Sequence Assembly, Annotations and Analysis

Sequences obtained were assembled using DANMAN (http://www.lynnon.com/). Protein-coding genes were aligned by Clustal X [19], then identified and translated to amino acids through MEGA version 4.0 [20]. rRNA genes were identified by sequence comparison with other dipteran insect [21]. Almost all tRNAs were recognized by tRNAscan-SE Search Server v.1.21 online [22] and the tRNAs that could not be found by tRNAscan-SE were confirmed by sequence comparison with other dipteran insects. The control region was examined for repeats and special structures with the aid of the Tandem Repeats Finder (http://www.bioinfo.rpi.edu/applications/Mfold) [23]. The nucleotide composition was calculated by the DNA Star (http://www.dnastar.com/,[24]). The relative synonymous codon usage was calculated by MEGA version 4.0 [20]. Strand asymmetry was evaluated by AT Skew and GC Skew using the formulae: AT skew = [A% − T%] / [A% + T%] and GC skew = [G% − C%] / [G% + C%] [22].

Phylogenetic Analysis

Phylogenetic analysis was carried out based on 26 complete mitochondrial genome sequences from the known 75 dipteran sequences. *Bombyx mandarina* was selected as the out-group (S1 Table). Phylogenetic trees were built based on the 13 protein-coding genes. First, the alignment of amino acids for every protein-coding gene was carried out using Clustal X [19]. Then, we concatenated the alignment results of individual genes. Model selection was done with Modeltest 3.7 [25] and MrModeltest 2.3 [26] for ML analysis and Bayesian inference, respectively. The results showed that the GTR+I+G model was the most ideal for analysis using nucleotide alignments. The GTR+I+G model was used with MrBayes Version 3.1.1 [27] and a PHYML online web server [28]. The alignments were used to carry out a maximum likelihood (ML) and Bayesian analysis (BI), using PHYML [28] and MrBayes [27]. In Bayesian analysis, the average standard deviation of split frequencies was below 0.01, and about 1,000,000 generations were
conducted for the matrix, and each set was sampled every 200 generations with a burn of 25%. Finally, we removed the aging trees and exported the optimal tree.

Results and Discussion

Genome Organization

The complete mitochondrial genome of *D. antiqua* is a double stranded circular molecule with a length of 16,141 bp (Fig 1, GenBank accession number KT026595). The genome is medium-sized in compared with other Diptera mitochondrial genomes that range from 14,503 bp (*Rhopalomyia pomum*) to 19,517 bp (*Drosophila melanogaster*) in length. It includes 37 genes (13 protein coding genes, 22 tRNAs and 2 rRNAs) and a non-coding region (A+T rich region, also called as the control region) (Table 1). There are 23 genes located on the J-strand (9 protein coding genes and 14 tRNAs) with the other 14 genes on the N-strand (4 protein coding genes, 8 tRNAs and 2 rRNAs). Fourteen intergenic spacers were found to have a total length of 127 bp, ranging in size from 2–26 bp and with the longest intergenic spacer located between tRNA\(^{\text{Arg}}\) and tRNA\(^{\text{Asn}}\). On the other hand, there were 12 gene overlaps in the mitochondrial genome of *D. antiqua* and they involve in a total of 43 bp; the longest overlap was 8 bp and appears between tRNA\(^{\text{Trp}}\) and tRNA\(^{\text{Cys}}\).

The gene order in the *D. antiqua* mitochondrial genome is the same as the gene order in *Dr. melanogaster*, which is the classical structure for Diptera [29]. The gene order of this mitochondrial genome shows the order is highly conserved in Diptera, and only in the Cecidomyiidae do we see the rearrangement in trnA and trnR forming trnR-trnA. Other known dipteran species all have the same gene order as *D. melanogaster*. Rearrangements of the mitochondrial genome are relatively rare as evolutionary events; therefore, this is an important tool to evaluate the phylogenetic relations between different species.

Nucleotide Composition

The nucleotide composition of the mitochondrial genome of *D. antiqua* showed obvious bias towards A and T. The A+T content of the whole genome was 78.5% (A% = 39.6%, T% = 38.9%, G% = 8.9%, C% = 12.6%). The A+T content of isolated PCGs, tRNAs, rRNAs, control region and J-strand, N-strand were all above 70% (Table 2). The control region has the highest A+T content (93.7%). The skew statistics of the whole genome showed that the whole mitochondrial genome of *D. antiqua* is CG-skewed distinctly with almost equal A and T. The protein coding genes and rRNAs are TA-skewed and GC-skewed, tRNAs showed as AT-skewed and GC-skewed, the control region preferred to use T and C. Isolated genes on different strands showed different nucleotide bias (Table 2).

This strand bias in nucleotide composition is a universal phenomenon in metazoan mitochondrial genomes. The strand bias can be indicated by a comparative analysis of (A + T)% vs AT-skew and (G + C)% vs GC-skew. The mitochondrial genome analysis of all known families of Diptera is shown in Fig 2. The average AT-skew among the Diptera is 0.032, ranging from -0.034 in *Arachnocampa flavo* to 0.131 in *Bactrocera minax*, whereas the *D. antiqua* mitochondrial genome shows a quite weak AT-skew (0.009) (Table 2). The average GC-skew among the Diptera is -0.186, ranging from -0.315 in *Bactrocera minax* to -0.110 in *Mayetiola destruc*, and the *D. antiqua* mitochondrial genome shows a little higher than the average value (-0.172) (Table 3). The AT-skew and GC-skew of most dipteran mitochondrial genomes shows a positive AT-skew and negative GC-skew for the J-strand. AT content and GC content consistently show that the dipteran mitochondrial genomes have higher percentages of A+T. The underlying mechanism of this bias has been generally related to asymmetric mutation and selection pressure during replication and transcription. In the process of DNA replication and
transcription, one chain is a single chain longer than the other strand, the deamination rate of A and C is faster in single chain, and therefore, more deamination of A and C occurs, leading to this bias [30]. This nucleotide bias has significance for the study of replication, transcription and rearrangement of the mitochondrial genome.
Protein-coding Genes

Most of the protein-coding genes use ATN as start codon (four use ATT, six use ATG, and two use ATA). The only exception is the COI gene, which begins with the special quadruplet start codon of ATCA (Table 1). Only COII, ND5 and ND4 genes had incomplete termination.

Table 1. Organization of the Delia antiqua mitochondrial genome.

| Gene      | Direction | Location | Size (bp) | Anticodon | Codon Start | Codon Stop | Intergenic nucleotides* |
|-----------|-----------|----------|-----------|-----------|-------------|------------|-------------------------|
| tRNA^leu(UCN) | R         | 11676–12614 | 939       |           | ATA         | TAA        | 16                      |
| tRNA^leu(UUR) | F         | 12625–12689 | 65        | 12658-12660TAG |             |            | 10                      |
| tRNA^pro    | R         | 12690–14019 | 1330      |           |             |            | 0                       |
| tRNA^val    | R         | 14020–14091 | 72        | 14056-14058TAC |             |            | 0                       |
| srRNA       | R         | 14092–14875 | 784       |           |             |            | 0                       |
| Control region |         | 14876–16141 | 1266      |           |             |            | 0                       |

a Negative numbers indicate that adjacent genes overlap.

doi:10.1371/journal.pone.0139736.t001
codons of T and TA, all others use the complete termination codons TAA (ND2, COI, ATP8, ATP6, COIII, ND3, ND6, ND4L, ND1, CytB) (Table 1).

The nucleotide bias is also reflected in the protein-coding genes. The base composition of each codon position for the 13 protein-coding genes shows that they all have a high A+T percentage. The third codon position (81.8%) was distinctly higher than the other two codon
positions (76.4% and 71.2%). The A+T content of the protein-coding genes on different strands also show a high percentage (Table 2). Different codon positions of protein-coding genes show different skew statistics. The first codon position prefers to use A and G, and the others were TA-skewed and CG-skewed. The genes on the J-strand and in its second and third codon position all showed TA-skew and CG-skew; the first codon position was AT-skewed and GC-skewed; the genes on the N-strand all had a higher frequency of T and G (Table 2).

The bias of amino acids was found in the protein-coding genes. The protein-coding genes and genes on different strands all had an unbalanced percentage of amino acids. They all had a high percentage of Leu, and the least percentage of Cys (Table 4). The relative synonymous codon usage also showed significant biases. The most frequently used codons were UUA, CGA, GGA, GCU, UCA and GUAG, the codons CUC, CUG, CCG, ACG, GGC and GCG most rarely used (Table 5).

In the dipteran mitochondrial genomes, COI initiation codons are variable and include TCG, CCG, ATCA and ATTTAA [31–33]. It is a common phenomenon to use an incomplete codon as a termination codon. They will be supplemented by processing after transcription.

Table 3. Base composition and strand bias across mitochondrial genomes of Dipteran families.

| Family (number of species investigated) | A    | T    | AT-Skew | A+T  | G    | C    | GC-Skew | G+C  |
|----------------------------------------|------|------|---------|------|------|------|---------|------|
| Tipulomorpha                            |      |      |         |      |      |      |         |      |
| Trichoceridae (2)                       | 39.4 | 37.4 | 0.026   | 76.8 | 9.5  | 13.8 | -0.183  | 23.3 |
| Ptychopteridae (2)                      | 39.3 | 37.1 | 0.029   | 76.4 | 9.6  | 14   | -0.185  | 23.6 |
| Anisopodidae (1)                        | 37.9 | 38.4 | -0.007  | 76.3 | 10.1 | 13.5 | -0.144  | 23.6 |
| Tanyderidae (1)                         | 38.8 | 38.3 | 0.006   | 77.1 | 9.5  | 13.3 | -0.167  | 22.8 |
| Tipulidae (1)                           | 38.0 | 36.7 | 0.017   | 74.7 | 10.2 | 15.1 | -0.194  | 25.3 |
| Cecidomyiidae (2)                       | 44.6 | 40.0 | 0.054   | 84.6 | 6.8  | 8.6  | -0.117  | 15.4 |
| Pachyneuridae (1)                       | 38.9 | 37.2 | 0.022   | 76.1 | 9.1  | 14.8 | -0.238  | 23.9 |
| Keroplattidae (1)                       | 39.6 | 42.4 | -0.034  | 82.0 | 7.3  | 10.6 | -0.184  | 17.9 |
| Muscomorpha                             |      |      |         |      |      |      |         |      |
| Syrphidae (1)                           | 40.3 | 40.6 | -0.004  | 80.8 | 10.9 | 19.2 | -0.133  | 19.2 |
| Muscidae (2)                            | 39.6 | 39.4 | 0.002   | 79.0 | 9.1  | 11.9 | -0.132  | 21.0 |
| Oestridae (2)                           | 40.6 | 37.2 | 0.043   | 77.8 | 8.4  | 13.7 | -0.238  | 22.1 |
| Tachinidae (3)                          | 40.2 | 38.5 | 0.021   | 78.7 | 8.7  | 12.6 | -0.184  | 21.3 |
| Agromyzidae (4)                         | 40.9 | 37.5 | 0.043   | 78.3 | 8.9  | 12.8 | -0.180  | 21.7 |
| Tephritidae (12)                        | 39.3 | 34.4 | 0.068   | 73.6 | 10.1 | 16.3 | -0.233  | 26.4 |
| Calliphoridae (13)                      | 39.3 | 37.5 | 0.023   | 76.8 | 9.5  | 13.7 | -0.182  | 23.2 |
| Drosophilidae (18)                      | 40.0 | 39.0 | 0.011   | 79.0 | 8.8  | 12.2 | -0.157  | 21.0 |
| Sarcophagidae (1)                       | 39.4 | 35.3 | 0.055   | 74.7 | 9.6  | 15.6 | -0.238  | 25.2 |
| Fergusoninidae (1)                      | 40.9 | 37.2 | 0.047   | 78.1 | 8.8  | 13.1 | -0.196  | 21.9 |
| Anthomyiidae (1)                        | 39.6 | 38.9 | 0.009   | 78.5 | 8.9  | 12.6 | -0.172  | 21.5 |
| Culicomorpha                            |      |      |         |      |      |      |         |      |
| Culicidae (16)                          | 40.0 | 38   | 0.025   | 78.0 | 9.2  | 12.8 | -0.165  | 22.0 |
| Ceratopogonidae (1)                     | 39.0 | 38.3 | 0.009   | 77.4 | 8.6  | 14.0 | -0.237  | 22.6 |
| Chironomidae (1)                        | 39.1 | 37.8 | 0.017   | 76.9 | 9.4  | 13.7 | -0.186  | 23.1 |
| Tabanomorpha                            |      |      |         |      |      |      |         |      |
| Tabanidae (1)                           | 39.1 | 38.8 | 0.003   | 77.9 | 9.1  | 13.0 | -0.177  | 22.1 |
| Asilomorpha                             |      |      |         |      |      |      |         |      |
| Nembrinidae (1)                         | 40.4 | 33.6 | 0.091   | 74.0 | 9.8  | 16.2 | -0.245  | 26.0 |
| Avg.                                    | 39.8 | 37.4 | 0.032   | 77.2 | 9.3  | 13.6 | -0.186  | 22.8 |

doi:10.1371/journal.pone.0139736.t003
This bias is also reflected in the codon usage and amino acid composition. The protein-coding genes of the *D. antiqua* mitochondrial genome prefer to use codon UUA (Leu) and Leucine. This is expected because there are many transmembrane proteins in the mitochondrial genome and Leucine happens to be a kind of hydrophobic amino acid.

Transfer RNAs

Twenty-two complete tRNAs were found in the *D. antiqua* mitochondrial genome, and 20 of them were identified by tRNAscane-SE [35]. Only the tRNA<sup>Arg</sup> and tRNA<sup>Ser(AGN)</sup> could not be detected by software, and they were determined through comparison with published dipteran mitochondrial genomes. All tRNAs were folded into the typical clover-leaf structure except for tRNA<sup>Ser(AGN)</sup> (Fig 3). All tRNAs ranged from 63 to 72 bp in length. The typical clover-leaf structure contains an amino acid arm (7 bp), T<sub>ΨC</sub> arm (3–5 bp), DHU arm (3–4 bp), anticodon arm (4–5 bp) and a variable extra arm. tRNA<sup>Ser(AGN)</sup> had a special clover-leaf structure without a DHU arm.

Based on the secondary structure of the tRNAs in the *D. antiqua* mitochondrial genome, there were 7 unmatched base pairs. All of them were U-U unmatched base pairs which were present in the amino acid arms, T<sub>ΨC</sub> arm and anticodon arm.

Ribosomal RNAs

The boundaries of rRNA genes were identified by sequence alignment with published dipteran sequences. There were two rRNA genes in the *D. antiqua* mitochondrial genome, 16S rRNA and 12S rRNA. The locations of the 16S rRNA and 12S rRNA genes were between tRNA<sup>Leu(CUN)</sup> and tRNA<sup>Val</sup> and between tRNA<sup>Val</sup> and the A+T-rich region, respectively. The 16S rRNA gene

| Feature | Percentage (%) |
|---------|----------------|
|         | Protein-coding genes | Protein-coding genes-J | Protein-coding genes-N |
| Ala     | 4.43             | 4.53               | 4.27               |
| Cys     | 0.99             | 0.52               | 1.75               |
| Asp     | 1.80             | 1.96               | 1.54               |
| Glu     | 2.04             | 1.96               | 2.17               |
| Phe     | 9.21             | 8.94               | 9.66               |
| Gly     | 5.80             | 5.71               | 5.95               |
| His     | 1.91             | 2.57               | 0.84               |
| Ile     | 9.70             | 10.55              | 8.33               |
| Lys     | 2.42             | 2.27               | 2.66               |
| Leu     | 16.20            | 14.82              | 18.40              |
| Met     | 6.31             | 5.67               | 7.35               |
| Asn     | 5.35             | 5.67               | 4.83               |
| Pro     | 3.65             | 4.45               | 2.38               |
| Gln     | 2.04             | 2.22               | 1.75               |
| Arg     | 1.50             | 1.61               | 1.33               |
| Ser     | 9.13             | 8.46               | 10.22              |
| Thr     | 5.10             | 6.32               | 3.15               |
| Val     | 5.26             | 4.97               | 5.74               |
| Trp     | 2.60             | 2.96               | 2.10               |
| Tyr     | 4.51             | 3.84               | 5.60               |

[34]. This bias is also reflected in the codon usage and amino acid composition. The protein-coding genes of the *D. antiqua* mitochondrial genome prefer to use codon UUA (Leu) and Leucine. This is expected because there are many transmembrane proteins in the mitochondrial genome and Leucine happens to be a kind of hydrophobic amino acid.
is 1,330 bp long, and the 12S rRNA is 784 bp long. Their A+T content was 82.26% and 78.32%, respectively. The location of the two rRNAs is same as in other dipteran mitochondrial genomes and they are very conservative.

The Control Region

The control region of the *Delia antiqua* mitochondrial genome is located between 12S rRNA and tRNAIle and is 1266 bp in length with the highest A+T content 93.7% of the whole genome. Three conserved structural elements have been identified in the control region of the *D. antiqua* mitochondrial genome. We found two poly-T stretches, one (TA)n stretch with 98 repeats and several G(A)nT structures by using the Tandem Repeats Finder [36]. One of the two poly-T stretches was found near the tRNAIle gene in the minority strand with 37 bp; the other was located close to the 12S rRNA which is in the majority strand and 27 bp in length. The (TA)n stretch was located in J-strand and the G(A)nT structures were on N-strand.

Five conserved special structures in the control region have been identified in insects: a poly-T stretch, a [TA(A)]n-like stretch, a highly conserved stem-and-loop structure, a G(A)nT structure, and a G+A-rich stretch [5]. But the five conserved structures are not all found in every insect [37–38]. In the control region of *D. antiqua*, three of these structures were found and they may be involved in the control of transcription or replication [39].

Phylogenetic Relationships

We performed phylogenetic analysis using the nucleotide sequences of 13 protein-coding genes of 25 species of complete dipteran mitochondrial genome sequences and the *D. antiqua* mitochondrial genome using *Bombyx mandarina* as outgroup. The topological structures of the 2 phylogenetic trees constructed separately by ML and BI analyses are very similar, with only 1 exception of the location of *Culicoides arakawae* in the family Ceratopogonidae of the superfamily Chironomoidaea (Figs 4 and 5). On the ML tree the species is located at the base of the Culicidae (Culicoidae) clade (Fig 4), whereas on the BI tree it is linked up the Culicidae clade (Fig 5). All but 3 clades are strongly supported with >80 bootstrap values. However, the clade
of Ephydroidea + Oestroidea + Muscoidea + Tephritoidea and the clade of Ba. carambolae + Ba. dorsalis in Tephritidae are with bootstrap values equal or less 66 on both trees. On the Ml tree the clade up *Culicoides arakawae* has only a bootstrap value of 4, and on the BI tree the

**Fig 3.** Inferred secondary structure of tRNAs in the *Delia antiqua* mitochondrial genome. The tRNAs are labeled with their corresponding amino acids.

doi:10.1371/journal.pone.0139736.g003
clade below *Culicoides arakawae* has a bootstrap value of 88, which indicates that the location of *Culicoides arakawae* is pending.

More importantly, *D. antiqua* of Anthomyiidae (Muscoidea) is nested inside Calliphoridae (Oestroidea) clade, and Muscidae (Muscoidea) is linked inside the Oestroidea. Kutty et al. (2008) constructed the phylogenetic trees of Muscoidea, Hippoboscoidea and Oestroidea using 4 mitochondrial genes 12S, 16S, COI, and Cytb, and 4 nuclear genes 18S, 28S, Efla and CAD [15]. The results showed that the Muscoidea is paraphyletic with a monophyletic Oestroidea nested within the Muscoidea as sister group to Anthomyiidae + Scathophagidae, the Anthomyiidae is possibly paraphyletic, and the Calliphoridae is paraphyletic. Marinho et al. (2012) inferred the phylogenetic relationship of families in the Oestroidea using ITS2, 28S, COI and 16S regions, and suggest that Calliphoridae is paraphyletic [16]. Nelson et al. (2012)
constructed the phylogenetic tree of 13 Calliphoridae species of whole mtgenome sequences using 13 protein-coding genes and 2 ribosomal RNA genes, and suggest that Calliphoridae is polyphyletic [17]. The present study suggest a closer relationship of Anthomyiidae with Calliphoridae, but more whole mtgenome sequences are necessary to elucidate its paraphyly and phylogenetic diversity inside the family. The study also suggest that Calliphoridae is a paraphyly, and further study might elucidate the tranditional taxonomy of Anthomyiidae and Calliphoridae. The study suggest that both Oestroidea and Muscoidea are polyphyletic, which are partially supported by Kutty et al. (2008) and Nelson et al. (2012) [15, 17].

Conclusions

This is the first report of complete mitochondrial genome of the family Anthomyiidae. Comparative analysis showed that the gene size, gene order, base content, and base composition are
comparatively conserved as with other dipteran mitochondrial genomes. All of the 13 protein-coding genes use ATN as the initiation codon except for the COI gene, which starts with ATCA. Most tRNAs have the typical clover-leaf structure, except tRNA\(^{Ser(AGN)}\), which does not contain the dihydrouridine (DHU) arm. The location and structure of the two rRNAs are conservative and comparable with Dipteran and other insects. The control region between 12S rRNA and tRNA\(^{Ile}\) has the highest A+T content 93.7% in the \(D.\) antiqua mitochondrial genome. There were three kinds of special structures found in the control region, poly-T stretches, a (TA)\(n\) stretch and G(A)\(n\)T structures, which are considered as important elements related to replication and transcription.

Both maximum likelihood and Bayesian inference analyses using nucleotide sequences of 13 protein-coding genes highly suggest a closer relationship of \(D.\) antiqua in Anthomyiidae has a closer with Calliphoridae, Calliphoridae is a paraphyly, and both Oestroidea and Muscoidea are polyphyletic. The whole mtgenome sequences have also been demonstrated as an effective method for resolving phylogenetic relationships [17, 40, 41].

Supporting Information

S1 Table. The dipteran species used for this phylogenetic study with Bombyx mandarina of Lepidoptera as outgroup.

(DOC)

Author Contributions

Conceived and designed the experiments: BC NXZ. Performed the experiments: NXZ GY TJL QYH YZ FLS SR. Analyzed the data: NXZ BC. Wrote the paper: NXZ BC.

References

1. Wei SJ, Chen XX. Progress in research on the comparative mitogenomics of insects. Chinese Journal of Applied Entomology. 2011; 48(6): 1573–1585.
2. Zhai ZH, Wang XZ, Ding MX. Cell Biology. Third Edition. Beijing: Higher Education Press; 2007.
3. Fang SM, Zhang L, Lu C. Study and application in origin and evolution of mitochondrial DNA in the silk-moths. Chinese Bulletin of Entomology. 2010; 47(3): 439–445.
4. Boore JL. Animal mitochondrial genomes. Nucleic Acids Research. 1999; 27(8):1767–80. PMID: 10101183
5. Zhang DX, Hewitt GM. Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. Biochem Syst Ecol. 1997; 25: 99–120.
6. Macey JR, Schulte JA, Larson A. Evolution and phylogenetic information content of mitochondrial genomic structural features illustrated with acrodont lizards. Systematic Biology. 2000; 49(2):257–277. PMID: 12118408
7. Telford MJ, Hemioiu EA, Russell RB, Littlewood DT. Changes in mitochondrial genetic codes as phylogenetic characters: two examples from the flatworms. Proceedings of the National Academy of Sciences of the United States of America. 2000; 97(21): 11359–11364. PMID: 11027335
8. Castresana J, Feldmaier-Fuchs G, Psabo S. Codon reassignment and amino acid composition in hemichordate mitochondria. Proceedings of the National Academy of Sciences of the United States of America. 1998; 95(7): 3703–3707. PMID: 9520430
9. Ballard JW, Whitlock MC. The incomplete natural history of mitochondria. Molecular Ecology. 2004; 13(4): 729–244. PMID: 15012752
10. Brown WM, George M Jr, Wilson AC. Rapid evolution of animal mitochondrial DNA. Proceedings of the National Academy of Sciences of the United States of America. 1979; 76(4): 1967–1971. PMID: 109836
11. Blanchette M, Kunisawa T, Sankoff D. Gene order breakpoint evidence in animal mitochondrial phylogeny. Journal of Molecular Evolution. 1999; 49(2): 193–203. PMID: 10441671
1. Cameron SL. Insect mitochondrial genomics: implications for evolution and phylogeny. Annual Review of Entomology. 2014; 59: 95–117. doi: 10.1146/annurev-ento-011613-162007 PMID: 24160435

2. Chen B, Li WS, Feng GZ, He ZB and Li TJ. Laboratory rearing, biological characteristics and diapause induction of the onion maggot, Delia antiqua. Journal of Chongqing Normal University (Nature Science). 2010; 27(2): 9–13.

3. Chen B, Kayukawa T, Jiang H, Monteiro A, Hoshizaki S, Ishikawa Y. DaTrypsin, a novel clip-domain serine proteinase gene up-regulated during winter and summer diapases of the onion maggot, Delia antiqua. Gene. 2005; 347(1): 115–123. PMID: 15715962

4. Kutty SN, Pape T, Pont A, Wiegmann BM, Meier R. The Muscoidea (Diptera: Calyptratae) are paraphyletic: Evidence from four mitochondrial and four nuclear genes. Molecular Phylogenetics and Evolution. 2008; 49(2): 639–652. doi: 10.1016/j.ympev.2008.08.012 PMID: 18793735

5. Marinho MA, Junqueira AC, Paulo DF, Esposito MC, Villet MH, Azeredo-Espin AM. (2012) Molecular phylogenetics of Oestroidea (Diptera: Calyptratae) with emphasis on Calliphoridae: insights into the inter-familial relationships and additional evidence for paraphyly among blowflies. Molecular Phylogenetics and Evolution. 2012; 65(3): 840–854. doi: 10.1016/j.ympev.2012.08.007 PMID: 22926310

6. Nelson LA, Lambkin CL, Batterham P, Wallman JF, Dowton M, Whiting MF, et al. Beyond barcoding: a mitochondrial genomics approach to molecular phylogenetics and diagnostics of blowflies (Diptera: Calliphoridae). Gene. 2012; 511(2): 131–142. doi: 10.1016/j.gene.2012.09.103 PMID: 23043935

7. Zhang NX, Zhang YJ, Yu G, Chen B. Structure characteristics of the mitochondrial genomes of Diptera and design and application of universal primers for their sequencing. Acta Entomologica Sinica. 2013; 56(4): 398–407.

8. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research. 1997; 25(24): 4876–82. PMID: 9396791

9. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution. 2007; 24(8): 1596–1599. PMID: 17488738

10. Cannone JJ, Subramanian S, Schnare MN, Collett JR, D’Souza LM, Du Y, et al. The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. BMC Bioinformatics. 2002; 3:2. PMID: 11869452

11. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Research. 1997; 25(5): 955–964. PMID: 9023104

12. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Research. 2003; 31(13): 3406–3415. PMID: 12824337

13. Perna NT, Kocher TD. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. Journal of Molecular Evolution. 1995; 41(3): 353–358. PMID: 7563121

14. Nylander JAA. MrModeltest v2. Program distributed by the author. Uppsala University: Evolutionary Biology Centre. 2004.

15. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. Bioinformatics. 1998; 14(9): 817–818. PMID: 9918953

16. Hulsenbeek JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 2001; 17(8): 754–755. PMID: 11524383

17. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology. 2003; 52(5): 696–704. PMID: 14530136

18. Lewis DL, Farr CL, Kaguni LS. Drosophila melanogaster mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons. Insect Molecular Biology. 1995; 4(4): 263–278. PMID: 8825764

19. Lindahl T. Instability and decay of the primary structure of DNA. Nature. 1993; 362(6422): 709–715. PMID: 8469282

20. Oliveira MT, Barau JG, Junqueira AC, Feijao PC, Rosa AC, Abreu CF, et al. Structure and evolution of the mitochondrial genomes of Haematobia irritans and Stomoxys calcitrans: the Muscidae (Diptera: Calyptratae) perspective. Molecular Phylogenetics and Evolution. 2008; 48(3): 850–857. doi: 10.1016/j.ympev.2008.05.022 PMID: 18621550

21. de Oliveira MT, de Azaredo-Espin AM, Lessinger AC. Evolutionary and structural analysis of the cytochrome c oxidase subunit I (COI) gene from Haematobia irritans, Stomoxys calcitrans and Musca domestica (Diptera: Muscidae) mitochondrial DNA. DNA Sequence: the Journal of DNA Sequencing and Mapping. 2005; 16(2): 156–160. PMID: 16147969

22. Yu DJ, Xu L, Nardi F, Li JG, Zhang RJ. The complete nucleotide sequence of the mitochondrial genome of the oriental fruit fly, Bactrocera dorsalis (Diptera: Tephritidae). Gene. 2007; 396(1): 66–74. PMID: 17433576
34. Ojala D, Montoya J, Attardi G. tRNA punctuation model of RNA processing in human mitochondria. Nature. 1981; 290(5806):470–474. PMID: 7219536

35. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Research. 1997; 25(5): 955–964. PMID: 9023104

36. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Research. 2003; 31(13): 3406–3415. PMID: 12824337

37. Zhou Z, Huang Y, Shi F, Ye H. The complete mitochondrial genome of Deracantha onos (Orthoptera: Brachyptera). Molecular Biology Reports. 2009; 36(1): 7–12. PMID: 17891510

38. Spanos L, Koutroumbas G, Kotsyfakis M, Louis C. The mitochondrial genome of the mediterranean fruit fly, Ceratitis capitata. Insect Molecular Biology. 2000; 9(2): 139–144. PMID: 10762421

39. Duarte GT, De Azeredo-Espin AM, Junqueira AC. The mitochondrial control region of blowflies (Diptera: Calliphoridae): a hot spot for mitochondrial genome rearrangements. Journal of Medical Entomology. 2008; 45(4): 667–676. PMID: 18714866

40. Zhao Z, Su TJ, Chesters D, Wang SD, Ho SY, Zhu CD, et al. The mitochondrial genome of Elodia flavipalpis Aldrich (Diptera: Tachinidae) and the evolutionary timescale of Tachinid flies. PLoS One. 2013; 8(4): e61814. doi: 10.1371/journal.pone.0061814 PMID: 23626734

41. Wiegmann BM, Trautwein MD, Winkler IS, Barr NB, Kim JW, Lambkin C, et al. Episodic radiations in the fly tree of life. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108(14): 5690–5695. doi: 10.1073/pnas.1012675108 PMID: 21402926