The Complete Nucleotide Sequence of the Mitochondrial Genome of *Bactrocera minax* (Diptera: Tephritidae)

Bin Zhang¹, Francesco Nardi³, Helen Hull-Sanders⁵, Xuanwu Wan³, Yinghong Liu²

¹ Key Lab of Integrated Pest Management of Shandong Province, College of Agronomy and Plant Protection, Qingdao Agricultural University, Qingdao, China, ² Chongqing Key Laboratory of Entomology and Pest Control Engineering, College of Plant Protection, Southwest University, Chongqing, China, ³ Sichuan Plant Protection Station, Chengdu, China, ⁴ Department of Evolutionary Biology, University of Siena, Siena, Italy, ⁵ Department of Entomology, Pennsylvania State University, University Park, Pennsylvania, United States of America

**Abstract**

The complete 16,043 bp mitochondrial genome (mitogenome) of *Bactrocera minax* (Diptera: Tephritidae) has been sequenced. The genome encodes 37 genes usually found in insect mitogenomes. The mitogenome information for *B. minax* was compared to the homologous sequences of *Bactrocera oleae*, *Bactrocera tryoni*, *Bactrocera philippinensis*, *Bactrocera carambolae*, *Bactrocera papayae*, *Bactrocera dorsalis*, *Bactrocera correcta*, *Bactrocera cucurbitae* and *Ceratitis capitata*. The analysis indicated the structure and organization are typical of, and similar to, the nine closely related species mentioned above, although it contains the lowest genome-wide A+T content (67.3%). Four short intergenic spacers with a high degree of conservation among the ten tephritid species mentioned above and *B. minax* were observed, which also have clear counterparts in the control regions (CRs). Correlation analysis among these ten tephritid species revealed close positive correlation between the A+T content of zero-fold degenerate sites (P₀FD) and the ratio of nucleotide substitution frequency at P₀FD sites to all degenerate sites (zero-fold degenerate sites, two-fold degenerate sites and four-fold degenerate sites) and amino acid sequence distance (ASD) were found. Further, significant positive correlation was observed between the A+T content of four-fold degenerate sites (P₄FD) and the ratio of nucleotide substitution frequency at P₄FD sites to all degenerate sites; however, we found significant negative correlation between ASD and the A+T content of P₄FD and the ratio of nucleotide substitution frequency at P₄FD sites to all degenerate sites. A higher nucleotide substitution frequency at non-synonymous sites compared to synonymous sites was observed in *nad4*, the first time that has been observed in an insect mitogenome. A poly(T) stretch at the 5’ end of the CR followed by a [TA(A)]ₙ-like stretch was also found. In addition, a highly conserved G+A-rich sequence block was observed in front of the poly(T) stretch among the ten tephritid species and two tandem repeats were present in the CR.

**Citation:** Zhang B, Nardi F, Hull-Sanders H, Wan X, Liu Y (2014) The Complete Nucleotide Sequence of the Mitochondrial Genome of *Bactrocera minax* (Diptera: Tephritidae). PLOS ONE 9(6): e100558. doi:10.1371/journal.pone.0100558

**Editor:** Kostas Bourtzi, International Atomic Energy Agency, Austria

**Received:** November 16, 2013; **Accepted:** May 29, 2014; **Published:** June 25, 2014

**Copyright:** © 2014 Zhang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported in part by the Taishan Mountain Scholar Constructive Engineering Foundation of Shandong, China and the Startup Fund for Distinguished Scholars (No. 631316) supported by Qingdao Agricultural University, China. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

*Email: yhliu@swu.edu.cn

**Introduction**

The family Tephritidae, generally known as “true” fruit flies, includes 471 genera and 4257 species distributed throughout the temperate and tropical parts of the world. Many species are of critical importance to man either as pests of fruit and vegetable crops or as beneficial species for the control of weeds [1]. The family *Bactrocera* Enderlein (Diptera: Tephritidae), generally known as the Chinese citrus fruit fly, has been a serious pest of commercial citrus crops in China for more than half a century [2]. *B. minax* was first collected from India and Sikkim and assigned the species to the genus *Bactrocera* in 1920 and assigned the species to the *B. minax* in 1940, should be placed in synonymy with *B. minax*.

A wide variety of questions about the biology and phylogeny of *B. minax* have been addressed with the aid of molecular tools. These studies could have used two main sources of genetic data; namely, nuclear sequence data and, most frequently, mitochondrial sequence data. Insect mitochondrial DNA (mtDNA) usually occurs as a double-stranded closed circular molecule, ranging in size from 14–20 kb and generally encoding 13 protein-coding genes (PCGs), two ribosomal RNAs (rRNAs) and 22 transfer RNA (tRNAs), which is conserved across bilaterian metazoans with only a few exceptions (e.g. loss of a small number of genes in some derived groups) [8]. The molecule contains at least one sequence of variable length known as the A+T-rich region or control region...
Table 1. Summary of primers used for complete mtgenome of *B. minax* amplification.

| Fragment | Upper primer | Sequence | Location | Down primer | Sequence | Location | Fragment Length |
|----------|--------------|----------|----------|-------------|----------|----------|-----------------|
| F1       | F1-Ur        | GCTAATAGCTGGCTGGTTCA | 155–177  | F1-Dr       | TGCTCTACTAATTCCGGCTCA | 1539–1560 | 1406            |
| F2       | F2-Ur        | TACAATCTGCTGTAACCTCAGCA | 1442–1468 | F2-Dr       | TAGGACAGGATATCTACTATTTA | 2357–2379 | 938             |
| F3       | F3-Ur        | GATCTTTGGACACCCGAGAA | 2175–2194 | F3-Dr       | ATTCATACCTGACTACATTG | 3383–3406 | 1232            |
| F4       | F4-Ur        | ATGGCGAGTGATGCTATTTA | 3016–3035 | F4-Dr       | GCTAAAGGACCTGACTTGG | 3789–3810 | 795             |
| F5       | F5-Ur        | GAAATTGGCGGGCCTAATCTAGA | 3670–3692 | F5-Dr       | GAGGCTATAGCTCCAGTCAT | 5072–5096 | 1427            |
| F6       | F6-Ur        | ATGCAGCTGGGCTGCTATTAA | 4670–4690 | F6-Dr       | ACTTGAACAAATACCCCTTTG | 5223–5245 | 576             |
| F7       | F7-Ur        | GTAACATTAGGAAACGGTGGAGAAG | 4967–5091 | F7-Dr       | TGGCAATAACCTGCTGTATTTTCTT | 6027–6049 | 1083            |
| F8       | F8-Ur        | TATCGCCCTATACACCCGAGAAG | 5089–5112 | F8-Dr       | ATCAGGTTGGTCGAGAAGA | 6543–6560 | 653             |
| F9       | F9-Ur        | AATACCCTAACACCTTCAGTG | 6355–6376 | F9-Dr       | TATCTAATCGGATGGAATG | 7684–7705 | 1351            |
| F10      | F10-Ur       | GCTCTTCTTAGTATAGCTGAC | 7546–7565 | F10-Dr      | GCTAAACCTAGTTCGGTTT | 8783–8801 | 1256            |
| F11      | F11-Ur       | AAAAAAACACCTGACGAC | 8600–8619 | F11-Dr      | TAGGAAATGATCTTTTTTATA | 9215–9236 | 637             |
| F12      | F12-Ur       | GGGGCCTCAACAGTACGCC | 8913–8932 | F12-Dr      | ATCTCATTGGCTAGGCCTT | 10422–10441 | 1529          |
| F13      | F13-Ur       | AGGAAGTATAGTCTTCTAC | 10139–10161 | F13-Dr     | GCAAATAGGAAATCATTC | 11297–11314 | 1176          |
| F14      | F14-Ur       | AGCAACAGCATTACAGGC | 10858–10878 | F14-Dr     | CTTTACACTGTTCTGCATAT | 11802–11824 | 967              |
| F15      | F15-Ur       | ACAGTACTGAGGCCAGCAG | 11492–11511 | F15-Dr     | GTGGCTTTTTTTACTTTTTGGAACG | 12556–12579 | 1088          |
| F16      | F16-Ur       | TAGAATAGGAATGAGTGACG | 12254–12275 | F16-Dr     | ACTTACCGTAAACGATAGT | 12960–12979 | 726              |
| F17      | F17-Ur       | TTCTAAACTGCTCTTC | 12757–12776 | F17-Dr     | GTGGCTTTTGTGCTACGGTTT | 13713–13736 | 980             |
| F18      | F18-Ur       | ATGTTTTTTGGTAAACGGG | 13360–13379 | F18-Dr     | AGACTCAGTAGATACCCCTATTT | 14555–14574 | 1215            |
| F19      | F19-Ur       | TACAGGCGAGGGTGGATCTCCTG | 14585–14676 | F19-Dr     | GGCTGTATTGTCTTTTTACTTTA | 14826–14845 | 388             |
| F20      | F20-Ur       | AGGTATCATATTCTTATT | 14587–14597 | F20-Dr     | AGTCGATGAGGGCTGGTTTATTTA | 254–275 | 1762          |
| F21      | F21-Ur       | ACTCTAATCTGCTAGGCTT | 14618–14637 | F1-Dr      | TGCTCTACTATCCGCGCTCA | 1539–1560 | 2986            |

Note: Lowercase “r” behind some primer names represents these primers were designed on basis of Simon et al. (1994); Lowercase “r” behind some primer names represents these primers were designed by us.
doi:10.1371/journal.pone.0100558.t001
(CR), which contains initiation sites for transcription and replication [9] and ranges in size from tens to several thousands of base pairs [10–13]. As the results of highly conservative gene structures among phyla, maternal inheritance, high copy number and relatively fast evolution rates compared to nuclear DNA [14], mitochondrial genome (mitogenome) sequences have been regarded as useful molecular markers in studies focusing on comparative and evolutionary genomics, molecular evolution, phylogenetics, phylogeography and population genetics [15].

Many complete or nearly complete mitogenomes have been sequenced and comparative analyses at the genus or species level have used multiple complete mitochondrial genes instead of one or partial genes, including molecular systematics [16–20], population genetics/phylogeography [16], diagnostics [21], molecular evolutionary studies [13,22,23], the frequency and type of gene rearrangements [24,25] and the evolution of genome size [26]. To date, more than 500 insect mitogenomes have been sequenced from all orders, including 77 dipterans in 24 families, and are available in Genbank. In this study, we sequenced the complete sequence of the mitogenome of B. minax (Diptera: Tephritidae).

Genbank contains information for only ten Tephritidae species; Bactrocera oleae, Bactrocera tryoni, Bactrocera philippinensis, Bactrocera

Figure 1. Circular map of the mitogenome of B. minax. The genes located outside adjoined the bold line circle (J-strand) indicated that the direction of transcription is opposite to the genes located inside adjoined the bold line circle (N-strand). B. minax complete mitogenome was jointed using 21 (F1–F21) fragments shown as single lines within the bold line circle.

doi:10.1371/journal.pone.0100558.g001
carambolae, Bactrocera papayae, Bactrocera dorsalis, Bactrocera correcta, Bactrocera cucurbitae, Ceratitis capitata and B. minax. Nine of these species belong to the genus Bactrocera, including four species of the B. dorsalis species complex; the other species belongs to the genus Ceratitis. Within the nine Bactrocera species, B. philippinensis, B. carambolae, B. papayae and B. dorsalis belong to the B. dorsalis species complex, B. correcta, B. cucurbitae and B. tryoni belong to other species-groups within the subgenus Bactrocera, and B. oleae and B. minax belong to the subgenus Dacus and Tetradacus, respectively. Although recent molecular evidence suggests B. papaya, B. philippinensis and B. dorsalis likely represent one species [27–30], with anticipation of the analysis of the B. minax mitogenome, we compare the sequence and mitogenome origins to the tephritid species B. oleae, B. dorsalis, B. philippinensis, B. carambolae, B. papayae, B. correcta, B. cucurbitae, B. tryoni and C. capitata.

**Materials and Methods**

1. Insect and mtDNA extraction, protein-coding genes and sequencing

We collected B. minax adults from a citrus garden on private land at Xianli Zeng covering an area of 20 hectares in Wulong
(Chongqing Province, China). We confirm that Mr Zeng, the owner of this land, allowed us to conduct the study on this site. No specific permission was required for this location and our activity. We confirm the field studies did not involve endangered or protected species. *B. minax* adults were stored at 25°C in 99% (v/v) ethanol. Morphological identification was done according to White and Wang [7]. Total DNA was isolated from three adult specimens using the DNeasy Blood & Tissue kit (QIAGEN). The whole mitochondrial genome was amplified in 21 overlapping fragments, ranging in size from 388 bp to 1762 bp. PCR primers were designed as described [31] and by comparison to the available pieces, ranging in size from 388 bp to 1762 bp. PCR primers were amplified as described [31] and by comparison to the available sequences of *B. oleae*, *B. dorsalis*, *B. philippinensis*, *B. carambolae*, *B. papayae*, *B. correcta*, *B. curcubitae* and *C. capitata* (Table 1). Amplification was done in a thermocycler (Eppendorf Mastercycler 5333) in 50 μl reactions containing 5 μl of 25 mM MgCl₂, 5 μl of 10×PCR Buffer (Mg²⁺ free), 8 μl of a DNTP mixture (2.5 mM each), 5 μl of 10 μM each primer, 0.5 μl of 5 U/μl Taq polymerase (Takara Biomedical, Japan) and 2 μl of a 1/10 dilution of the DNA extract. Amplification conditions were: 5 min at 94°C followed by 34 cycles of 30 s at 94°C, 1 min at 40–58°C (depending on the primer pair) and 2 min at 72°C. The F21 fragment (Fig. 1) was amplified using LA Taq (Takara Biomedical, Japan) and a cycle consisting of a pre-PCR denaturation at 96°C for 2 min followed by 30 cycles of 10 s at 98°C and 2 min at 58°C with a final elongation step of 10 min at 72°C. PCR products were separated by electrophoresis and purified using a QiAquick Gel Extraction Kit (QIAGEN). PCR products were sequenced directly on both strands using amplification and additional ad hoc primers as needed. Individual sequences were combined in a consensus contig using DNAStar package software (DNAStar Inc.).

### 2. Sequence analysis and gene annotation

Genes encoded on the *B. minax* mitogenome were located initially by comparison to homologous full-length insect mitochondrial sequences using DNAStar. Nucleotide sequences of PCGs were translated using the invertebrate mtDNA genetic code. tRNA genes were identified initially using tRNAscan-SE Search Server version 1.21 (available online at http://lowelab.ucsc.edu/tRNAscan-SE/) [32] and refined using tRNAscan-SE and RNAshapes [33]. The presence and secondary structures of tRNA genes that could not be located by tRNAscan-SE owing to variant morphology were annotated manually by comparison to the sequences of other insect tRNAs [34–37]. Codon usage analysis and relative synonymous codon usage (RSCU) in PCGs were calculated using CodonW version 1.4.2 (John Peden, available at http://codonw.sourceforge.net/index.html) [38]. Potential secondary structure folds of non-coding sequences and sequences in the CR were calculated with the DNA mfold web server using default settings (http://mfold.bioinfo.rpi.edu/cgi-bin/dna-form1.cgi) [39]. The presence of tandem repeats in the CR was investigated using the Tandem Repeats Finder available online (http://tandem.bu.edu/trf/trf.shtml) [40]. The A+T content and nucleotide substitution frequency at synonymous sites and non-synonymous sites (the number of synonymous substitutions per site and the number of non-synonymous substitutions per site) were calculated on the basis of the data using MEGA 4.0 [41]. The correlation analysis was done by the bivariate method using SPSS version 13 (SPSS Inc., Chicago, IL). The overall average amino acid distance among each of the PCGs from ten tephritid species (*B. minax*, *B. oleae*, *B. tryoni* *B. philippinensis*, *B. carambolae*, *B. papayae*, *B. correcta*, *B. curcubitae* and *C. capitata*) were calculated by the method of Poisson distances by MEGA 4.0 [41]. The complete *B. minax* mtDNA sequence was deposited in Genbank under accession no. HM776033.

### Results and Discussion

#### 1. Genome organization

The mitochondrial genome of *B. minax* is a closed circular molecule of 16043 bp; hence, it is longer than the other nine tephritid mitogenomes available (range 15,815 bp in *B. oleae* to 15,980 bp in *C. capitata* but is still well within the range of other insect mitogenomes [14,503 bp in *Rhopalomyia pomum* [42] to 19517 bp in *Drosophila melanogaster* [11]). The gene content is typical of metazoan mitogenomes, with 13 PCGs (cox1-3, cox, nad1-6, nad4l, atp6 and atp8), 22 tRNAs and two genes for ribosomal RNA subunits (rrnS and rrnL). A long uninterrupted non-coding region of 1141 bp, likely homologous to the insect A+T-rich region, is present between rrnS and rrnL, corresponding to position 14,903 to 16,043 in the annotated sequence. The gene order in the
# Table 4.

Relative synonymous codon usage of 10 tephritid species, B. minax, B. correcta and B. curcurbitae.

| Codon   | B. minax | B. dorsalis | B. carambolae | B. papaya | B. correcta | B. curcurbitae |
|---------|----------|-------------|---------------|-----------|-------------|---------------|
| **A**   | **A**    | **A**       | **A**         | **A**     | **A**       | **A**         |
| **B**   | **B**    | **B**       | **B**         | **B**     | **B**       | **B**         |
| **C**   | **C**    | **C**       | **C**         | **C**     | **C**       | **C**         |
| **D**   | **D**    | **D**       | **D**         | **D**     | **D**       | **D**         |
| **E**   | **E**    | **E**       | **E**         | **E**     | **E**       | **E**         |
| **F**   | **F**    | **F**       | **F**         | **F**     | **F**       | **F**         |
| **G**   | **G**    | **G**       | **G**         | **G**     | **G**       | **G**         |
| **H**   | **H**    | **H**       | **H**         | **H**     | **H**       | **H**         |
| **I**   | **I**    | **I**       | **I**         | **I**     | **I**       | **I**         |
| **J**   | **J**    | **J**       | **J**         | **J**     | **J**       | **J**         |
| **K**   | **K**    | **K**       | **K**         | **K**     | **K**       | **K**         |
| **L**   | **L**    | **L**       | **L**         | **L**     | **L**       | **L**         |
| **M**   | **M**    | **M**       | **M**         | **M**     | **M**       | **M**         |
| **N**   | **N**    | **N**       | **N**         | **N**     | **N**       | **N**         |
| **O**   | **O**    | **O**       | **O**         | **O**     | **O**       | **O**         |
| **P**   | **P**    | **P**       | **P**         | **P**     | **P**       | **P**         |
| **Q**   | **Q**    | **Q**       | **Q**         | **Q**     | **Q**       | **Q**         |
| **R**   | **R**    | **R**       | **R**         | **R**     | **R**       | **R**         |
| **S**   | **S**    | **S**       | **S**         | **S**     | **S**       | **S**         |
| **T**   | **T**    | **T**       | **T**         | **T**     | **T**       | **T**         |
| **U**   | **U**    | **U**       | **U**         | **U**     | **U**       | **U**         |
| **V**   | **V**    | **V**       | **V**         | **V**     | **V**       | **V**         |
| **W**   | **W**    | **W**       | **W**         | **W**     | **W**       | **W**         |
| **X**   | **X**    | **X**       | **X**         | **X**     | **X**       | **X**         |
| **Y**   | **Y**    | **Y**       | **Y**         | **Y**     | **Y**       | **Y**         |
| **Z**   | **Z**    | **Z**       | **Z**         | **Z**     | **Z**       | **Z**         |
| Anticodon | R. donaldsi | R. okara | R. tropaeoli | C. capitata | B. philippinensis | B. carambolae | B. papaya | B. correcta | B. curcubitae |
|-----------|-------------|----------|-------------|------------|-----------------|---------------|-----------|------------|-------------|
| UAU       | 0.99        | 0.97     | 0.98        | 0.89       | 0.95            | 0.92          | 0.97      | 0.95       | 0.98        |
| UAC       | 0.99        | 0.97     | 0.98        | 0.89       | 0.95            | 0.92          | 0.97      | 0.95       | 0.98        |
| UGC       | 0.99        | 0.97     | 0.98        | 0.89       | 0.95            | 0.92          | 0.97      | 0.95       | 0.98        |
| CGG       | 0.99        | 0.97     | 0.98        | 0.89       | 0.95            | 0.92          | 0.97      | 0.95       | 0.98        |
| AGG       | 0.99        | 0.97     | 0.98        | 0.89       | 0.95            | 0.92          | 0.97      | 0.95       | 0.98        |
| GGG       | 0.99        | 0.97     | 0.98        | 0.89       | 0.95            | 0.92          | 0.97      | 0.95       | 0.98        |

**Note:** All rates represent the relative synonymous codon usage for each of the FGs. The highest relative synonymous codon usage for the FGs on doi:10.1371/journal.pone.0100558.t004
B. minax mitogenome corresponds to the typical and plesiomorphic state hypothesized for the Pancrustacea, and is shared with all tephritids analyzed to date (Fig. 1).

Genes in the B. minax mitogenome overlap by a total of 43 bp, distributed in 12 segments from 1 to 17 bp long and are separated by a total of 178 bp dispersed in 16 intergenic spacers from 2 to 42 bp (without taking the tRNA-like sequence into account; Table 2). Despite its relatively large size, the B. minax mitogenome has more overlapping sequences between genes compared to those of other tephritids; genes overlap by a total of 35 bp at 11 boundaries in B. oleae, 29 bp in seven locations in B. tryoni, 27 bp in five locations in B. dorsalis, 34 bp in ten locations in B. philippinensis, 32 bp in nine locations in B. carambolae, 34 bp in ten locations in B. papaya, 35 bp in 11 locations in B. correcta, 32 bp in nine locations in B. carambolae and only 3 bp at three boundaries in C. capitata.

2. Nucleotide composition

The overall base composition of B. minax is 38.0% A, 11.2% G, 29.3% T and 21.5% C. Similar to other insect sequences, the B. minax mitogenome nucleotide composition is biased toward adenine and thymine (67.3% A+T), which is the lowest value among the tephritid mitogenomes available. Analyzed separately, all PCGs (64.3%), tRNAs (72.2%), sRNAs (73.7%) and CR (77.6%) have the lowest A+T content compared to the other known tephritid mitogenomes (Table 3).

Considering the two strands separately, the PCGs on the Majority strand (J-strand, nine PCGs are located on this strand) (61.5%) have a lower A+T content compared to the Minority strand (N-strand, the other four PCGs are located on this strand) (68.9%). Furthermore, PCGs encoded on the J-strand have a comparable content of A (51.0%) and T (30.5%), whereas PCGs on the N-strand show a strong bias for T content (46.3%) compared to A content (22.6%). The above situation has been observed in the other tephritid mitogenomes available (data not shown) and in other insects [34–37,43–50]. However, tRNAs on the two opposite strands have nearly equal A+T contents, which has been found in the other nine tephritid species. For three PCG codon positions, the third codon positions have significantly higher A+T content than the first and second codon positions owing to genetic code degeneracies. In particular, T in each codon position of PCGs on the N-strand is over-represented. With exception of the second codon position over-representing T, however, the first and third codon positions of PCGs show a preponderance of A on
the J-strand on the N-strand, which is similar to many insect mitogenomes [34–37,43–50] (Table 3).

The base compositional bias for A+T in PCGs is reflected in the relative synonymous codon usage statistics of the B. minax mitogenome (Table 4). With the exception of amino acid His, codons with A or T in the third codon position are generally more strongly over-represented compared to codons terminating with A or T in the third codon position (Fig. B). In this study, the A+T content and nucleotide substitution frequency at zero-fold degenerate sites (P0FD), two-fold degenerate sites (P2FD) and four-fold degenerate sites (P4FD) (the number of substitutions per P0FD, P2FD and P4FD site) in each PCG of each mitogenome of 10 tephritid species, B. oleae, B. tryoni, B. philippinensis, B. carambolae, B. papaya, B. dorsalis, C. capitata, B. minax, B. correcta and B. curcurbitae.

Table 5. A+T content percentage and nucleotide substitution frequency at 0-fold degenerate sites (P0FD), 2-fold degenerate sites (P2FD) and 4-fold degenerate sites (P4FD) (the number of substitutions per P0FD, P2FD and P4FD site) in each PCG of each mitogenome of 10 tephritid species, B. oleae, B. tryoni, B. philippinensis, B. carambolae, B. papaya, B. dorsalis, C. capitata, B. minax, B. correcta and B. curcurbitae.

| Protein-coding genes | P0FD | P2FD | P4FD |
|----------------------|------|------|------|
|                      | A+T percentage (%) | nucleotide substitution frequency | A+T percentage (%) | nucleotide substitution frequency | A+T percentage (%) | nucleotide substitution frequency |
| nad2                 | 70.96±1.41 | 0.198 | 74.61±8.66 | 0.811 | 74.33±10.99 | 1.689 |
| cox1                 | 56.53±0.21 | 0.025 | 73.31±8.55 | 0.743 | 84.93±7.82 | 1.409 |
| cox2                 | 59.71±0.74 | 0.074 | 76.92±7.67 | 0.624 | 82.47±8.64 | 1.306 |
| atp8                 | 68.92±0.81 | 0.235 | 81.54±8.66 | 0.577 | 83.33±11.86 | 1.400 |
| atp6                 | 68.76±2.92 | 0.459 | 63.36±2.91 | 0.227 | 62.46±8.84 | 0.590 |
| cox3                 | 57.46±0.36 | 0.034 | 75.10±5.61 | 0.655 | 87.34±4.95 | 1.266 |
| nad3                 | 68.38±1.05 | 0.192 | 76.61±10.74 | 0.831 | 82.12±8.50 | 1.485 |
| nad6                 | 72.59±1.44 | 0.265 | 80.12±8.60 | 0.655 | 87.68±4.11 | 1.326 |
| cob                  | 61.62±0.42 | 0.057 | 68.66±9.26 | 0.741 | 81.14±6.72 | 1.417 |
| nad1                 | 65.54±0.53 | 0.094 | 88.33±2.51 | 0.381 | 73.72±1.09 | 1.090 |
| nad4l                | 71.44±1.44 | 0.107 | 87.94±4.12 | 0.397 | 83.46±8.32 | 1.154 |
| nad4                 | 76.42±3.72 | 0.783 | 77.49±1.86 | 0.172 | 41.72±0.73 | 0.004 |
| nad5                 | 66.33±1.23 | 0.172 | 85.77±4.83 | 0.313 | 85.18±5.45 | 1.234 |

Correlation coefficient (r) = 0.735, Confidence probability (P) = 0.004.<0.01, 0.477<0.05, 0.000<0.01

Note: the correlation analysis used Pearson coefficient under two-tailed test of significance.

doi:10.1371/journal.pone.0100558.t005

Mitochondrial Genome of Bactrocera minax

The general trend in the mitogenome toward a lower G content higher compared to C in coding genes on the N-strand (PCGs, tRNAs, CR and intergenic nucleotides), while the G content is higher compared to T in PCGs is reflected in the base compositional bias for A+T in PCGs. The T-content percentage and nucleotide substitution frequency at 0-fold degenerate sites (P0FD), 2-fold degenerate sites (P2FD) and 4-fold degenerate sites (P4FD) (the number of substitutions per P0FD, P2FD and P4FD site) in each PCG of each mitogenome of 10 tephritid species, B. oleae, B. tryoni, B. philippinensis, B. carambolae, B. papaya, B. dorsalis, C. capitata, B. minax, B. correcta and B. curcurbitae.

The base compositional bias for A+T in PCGs is reflected in the relative synonymous codon usage statistics of the B. minax mitogenome (Table 4). With the exception of amino acid His, codons with A or T in the third codon position are generally strongly over-represented compared to codons terminating with either G or C. The ratio of Gr+G-rich (Pro, Ala, Arg and Gly) codons to A+T-rich codons (Phe, Ile, Met, Tyr, Asn and Lys) in B. minax PCGs was 0.44, which is higher compared to the other nine tephritid species. B. dorsalis (0.29), B. philippinensis (0.29), B. carambolae (0.30), B. papaya (0.29), B. correcta (0.30), B. cucurbitae (0.32), B. oleae (0.31), B. tryoni (0.32) and C. capitata (0.23). This demonstrates the amino acid composition is affected by the lower A+T mutational bias in B. minax (67.3%) and the stronger A+T mutational bias in B. dorsalis (73.6%), B. philippinensis (73.6%), B. carambolae (73.6%), B. papaya (73.5%), B. correcta (73.2%), B. cucurbitae (72.8%), B. oleae (72.6%), B. tryoni (72.4%) and C. capitata (77.5%).

With the exception of first codon positions, G is under-represented compared to C in coding genes on the J-strand (PCGs, tRNAs, CR and intergenic nucleotides), while the G content is higher compared to C in coding genes on the N-strand (PCGs, tRNAs and rRNAs). This base compositional bias is in line with the general trend in the mitogenome toward a lower G content [51].

Base compositional heterogeneity and among-site rate variation (ASRV) are known to affect phylogenetic inference, resulting in the identification of incorrect phylogenetic relationships [52]. The easiest solution is simply to avoid non-stationary genes [53] but most earlier studies used relatively intuitive mitogenome data partitioning schemes, including by gene type (PCG, rRNA and tRNA), by gene, by codon position, by codon and gene, or by the strand on which the coding gene is located [15]. Inevitably, different intuitive partitioning schemes can each result in strong conflicting topologies, especially at deeper phylogenetic levels [25,54,55]. Therefore, selection of stationary, reversible compositional homogeneous is vital for reliable phylogenetic inference [52,56].

Many earlier studies were focused on the A+T content of different genes or regions to investigate the base compositional heterogeneity and among-site rate variation ASRV [57]. For mitogenomes, composition bias of A+T content was verified in most earlier studies; e.g. A+T content was usually over-represented in non-coding regions [58] and the third codon position generally had stronger A+T composition bias compared to the other two codon positions [59] etc.. We asked how variability between PCGs is related to underlying A+T content and its distribution across synonymous and non-synonymous sites.

In this study, the A+T content of zero-fold sites (P0FD), two-fold (P2FD) and four-fold degenerate sites (P4FD) was determined for each of the PCGs from ten tephritid species (B. minax, B. oleae, B. tryoni, B. dorsalis, B. philippinensis, B. carambolae, B. papaya, B. correcta, B. cucurbitae and C. capitata). (Fig. 2). Nucleotide substitution frequency was calculated in P0FD, P2FD and P4FD for each of the PCGs among five tephritid species (Fig. 3). After analyzing the correlation between A+T content and nucleotide substitution frequency for each of the PCGs, we found a significant positive correlation between A+T content percentage of zero-fold degenerate sites (AT0F) and nucleotide substitution frequency at P0FD (r=0.735, P=0.004) as well as between A+T content percentage of four-fold degenerate sites (AT4F) and nucleotide substitution frequency at P4FD (r=0.864, P=0.000) (Table 5). Correlation analysis indicated there is a significant positive correlation between AT0F and ASD (r=0.792, P=0.003), ASD and the nucleotide...
Table 6. The A+T content percentage of 0-fold degenerate sites (AT0F), the nucleotide substitution number of 0-fold degenerate sites/the nucleotide substitution number of all degenerate sites (R0F/all) and the mean genetic distance based on amino acid sequence (ASD) in each protein-coding gene of mitochondrial genomes of 10 tephritid species, *B. oleae*, *B. tryoni*, *B. philippinensis*, *B. carambolae*, *B. papaya*, *B. dorsalis*, *C. capitata*, *B. minax*, *B. correcta* and *B. curcubitae*, and the correlation coefficient between them (AT0F, R0F/all and ASD).

| Protein-coding genes | AT0F | AT4F | R0F/all | R4F/all | ASD | AT0F & R0F/all | AT0F & ASD | R0F/all & ASD | AT4F & R0F/all | AT4F & ASD | R4F/all & ASD | R0F/all & ASD | AT4F & ASD |
|----------------------|------|------|---------|---------|-----|--------------|------------|--------------|--------------|------------|--------------|-------------|------------|
| nad2                 | 70.96 ± 1.41 | 74.33 ± 10.99 | 0.290 | 0.362 | 0.117 | *r* = 0.760, *P* = 0.003 < 0.01 | *r* = 0.752, *P* = 0.003 < 0.01 | *r* = 0.983, *P* = 0.000 < 0.01 | *r* = 0.809, *P* = 0.000 < 0.01 | *r* = 0.828, *P* = 0.000 < 0.01 | *r* = 0.970, *P* = 0.000 < 0.01 |
| cox1                 | 56.53 ± 0.21 | 84.93 ± 7.82 | 0.046 | 0.606 | 0.014 |              |            |              |              |              |              |              |            |
| cox2                 | 59.71 ± 0.74 | 82.47 ± 8.64 | 0.148 | 0.514 | 0.036 |              |            |              |              |              |              |              |            |
| atp5                 | 68.92 ± 0.81 | 83.33 ± 11.86 | 0.400 | 0.350 | 0.164 |              |            |              |              |              |              |              |            |
| atp6                 | 68.76 ± 2.92 | 62.46 ± 6.84 | 0.780 | 0.141 | 0.280 |              |            |              |              |              |              |              |            |
| cox3                 | 57.46 ± 0.36 | 87.34 ± 4.95 | 0.068 | 0.552 | 0.014 |              |            |              |              |              |              |              |            |
| nad3                 | 68.38 ± 1.05 | 82.12 ± 8.50 | 0.300 | 0.350 | 0.105 |              |            |              |              |              |              |              |            |
| nad6                 | 72.59 ± 1.44 | 87.68 ± 4.11 | 0.417 | 0.297 | 0.168 |              |            |              |              |              |              |              |            |
| cob                  | 61.62 ± 0.42 | 81.14 ± 6.72 | 0.106 | 0.497 | 0.029 |              |            |              |              |              |              |              |            |
| nad1                 | 65.54 ± 0.53 | 73.72 ± 1.09 | 0.222 | 0.449 | 0.052 |              |            |              |              |              |              |              |            |
| nad4                 | 71.44 ± 1.44 | 83.46 ± 8.32 | 0.244 | 0.385 | 0.053 |              |            |              |              |              |              |              |            |
| nad5                 | 76.42 ± 3.72 | 41.72 ± 0.73 | 0.936 | 0.004 | 0.382 |              |            |              |              |              |              |              |            |
|                     | 66.33 ± 1.23 | 85.18 ± 5.45 | 0.357 | 0.431 | 0.083 |              |            |              |              |              |              |              |            |

Note: the correlation analysis used Pearson coefficient under two-tailed test of significance.

The A+T content percentage of 4-fold degenerate sites (AT4F), the nucleotide substitution number of 0-fold degenerate sites/the nucleotide substitution number of all degenerate sites (R0F/all) and the mean genetic distance based on amino acid sequence (ASD) in each protein-coding gene of mitochondrial genomes of 10 tephritid species above, and the correlation coefficient between them (AT4F, R0F/all and ASD).

doi:10.1371/journal.pone.0100558.t006
Mitochondrial Genome of *Bactrocera minax*
P.082, P4FD of PCGs could be useful to judge the homogeneity of PCGs. This is because PCGs are used to analyze phylogenetic relationships for frequency compared to both P2FD and P4FD sites (Fig. 3). The P0FD conserved PCGs is due to higher A\textsubscript{h} hypothesize divergence at the amino acid level of less well or lower A\textsubscript{h} rate sites (R4F/all)(degenerate sites/the nucleotide substitution number of all degenerate sites/the nucleotide substitution number of zero-fold degenerate sites/the nucleotide substitution number of all degenerate sites (R\textsubscript{4F/a}d) (r = 0.983, P = 0.000), AT\textsubscript{a}p and R\textsubscript{4F/a}d (r = 0.760, P = 0.005) (Table 6). Interestingly, the significant positive correlation was observed between AT\textsubscript{a}p and the nucleotide substitution number of four-fold degenerate sites/the nucleotide substitution number of all degenerate sites (R\textsubscript{4F/a}d) (r = 0.809, P = 0.001); however, there was significant negative correlation between AT\textsubscript{a}p and ASD (r = -0.828, P = 0.000), between R\textsubscript{4F/a}d and ASD (r = -0.970, P = 0.000) (Table 6). On the basis of the above results, we can hypothesize divergence at the amino acid level of less well conserved PCGs is due to higher A+T at P\textsubscript{a}p in those genes and/or lower A+T at P\textsubscript{a}T. On the basis of this result, when we choose which PCGs are used to analyze phylogenetic relationships for different evolutionary time scales, the A+T content of P\textsubscript{a}p and/or P\textsubscript{a}T of PCGs could be useful to judge the homogeneity of PCGs.

Nucleotide substitution is considered to be a reflection of evolution at the molecular level. Many earlier studies indicated the substitution was directional bias across different genes in the mitogenome [15]. Some researchers have proposed variation of A+T% among taxa is associated with directional mutation pressure and has a phylogenetic component [57,60,61]. In this study, with the exception of nad\textsubscript{4}, all PCGs had significantly lower variation of A+T content among the ten tephritid species at P\textsubscript{a}p compared to both P\textsubscript{a}T and P\textsubscript{a}T sites. We observed that, with the exception of nad\textsubscript{4}, P\textsubscript{a}T sites had lower nucleotide substitution frequency compared to both P\textsubscript{a}T and P\textsubscript{a}T sites (Fig. 3). The P\textsubscript{a}T of nad\textsubscript{4} had a higher nucleotide substitution frequency (0.783) compared to both P\textsubscript{a}T (0.172) and P\textsubscript{a}T sites (0.004), and the R\textsubscript{a}d was 0.936. As a result of functional constraints, the number of nucleotide substitution per non-synonymous site is usually lower than that per synonymous site [62]. In this study, a higher nucleotide substitution frequency at P\textsubscript{a}T of nad\textsubscript{4} indicates the non-synonymous nucleotide substitution frequency was higher compared to the synonymous sites for this gene. Higher number of nucleotide substitution per non-synonymous site has been observed at the variable-region genes of immunoglobulins [63] and some genes of the histocompatibility complex [64] but this is the first reported occurrence in the mitogenome.

3. Protein-coding genes

With the exception of cox1 and nad3, all protein coding genes start with an ATN codon, with ATG used in cox2, atp6, cox3, nad4, nad4l, nad6 and cob, ATT in nad2, atp8 and nad5 and ATA in nad1. Genes for cox1 and nad3 used TCG and GTC as initiation codons, respectively. The initiation codon for cox1 was TCG(S) in B. minax, which was observed in other Diptera species [54], GTC being the initiation codon for nad3 was a new observation in tephritids, but it is common in other insects [65].

With the exception of nad3, nad5 and nad1, all PCGs are terminated by complete stop codons: TAG is used for nad2, atp6 and cob, TAA is used for cox2, atp8, cox3, nad4, nad4l and nad6 and TGA is used for cox1. The remaining genes, nad3, nad5 and nad4, are terminated by incomplete stop codons “T”.

4. Transfer RNA genes, ribosomal RNA genes and tRNA-like structure

All of 22 tRNA genes typical of metazoan mitogenomes were identified in the B. minax mitogenome, and the predicted structures are shown in Fig. 4. All tRNAs display a typical clover-leaf secondary structure, except for trn\textsubscript{D}tRN\textsubscript{H}(ACN), where the DHU arm appears to be replaced by seven unpaired nucleotides, a feature typical of other animal mitochondria [66]. Nuclear magnetic resonance analysis of the tertiary structure of nematode trn\textsubscript{D}tRN\textsubscript{H}(ACN) suggested such aberrant tRNA can fit the ribosome by adjusting its structural conformation and function in a way similar to that of usual tRNAs in the ribosome [67].

Like most insect tRNAs, all B. minax tRNAs have a length of 7 bp for the anticodon loop, 7 bp for the acceptor stem and 5 bp for anticodon stem. Most of the size variability in the B. minax tRNA genes originated from length variation in the DHU arms (loop size 4–9 bp, stem size 3–4 bp) and the T\textsuperscript{V}C arm (loop size 2–9 bp, stem size 3–5 bp); in addition, trn\textsubscript{M} and trn\textsubscript{H} contained U-U mismatches. trn\textsubscript{V}tRN\textsubscript{Y}(ACN) encodes an A-C mismatch, trn\textsubscript{H} encodes

Table 7. Locations, length and sequences of four shorter intergenic spacers in 10 tephritid species, B. oleae, B. tryoni, B. philippinensis, B. carambolae, B. papaya, B. dorsalis, C. capitata, B. minax, B. correcta and B. curcubitae.

| Species         | tRNA\textsuperscript{D}, tRNA\textsuperscript{H} | NDS - tRNA\textsuperscript{H} | tRNA\textsuperscript{D}(ACN), ND1 | ND1 - tRNA\textsuperscript{D}(ACN) |
|-----------------|---------------------------------|-------------------------------|---------------------------------|---------------------------------|
|                 | Sequence | Size (bp) | Sequence | Size (bp) | Sequence | Size (bp) | Sequence | Size (bp) |
| B. minax        | ACTAATTACAATTACACTA 18         | 18                            | TGATATAATTCTCA 14             | 14                | TACTAAATATATTAC 16          | 16                       | AAAAAACAG 10 |
| B. oleae        | ACTAAATAATAACTACACTA 18       | 18                            | TGATAATATTTTCA 15            | 15                | TACTAATAATATTAC 16          | 16                       | AAAAAACAG 10 |
| B. tryoni       | ACTAAATGAAATACACTA 18         | 18                            | TGACAATTCCAC 15             | 15                | TACTAAATTATATTAC 16         | 16                       | AAAAAACAG 10 |
| B. dorsalis     | ACTAAATATAATACACTA 18         | 18                            | TGAAAAATTTTCA 15            | 15                | TACTAAATTTATTAC 16          | 16                       | AAAAAACAG 10 |
| B. philippinensis | ACTAAATATAATGACACTA 18      | 18                            | TGAAAAATTTTCA 15            | 15                | TACTAAATTATATTAC 16         | 16                       | AAAAAACAG 10 |
| B. carambolae   | ACTAAATATAATACACTA 18         | 18                            | TGAAAAATTTTCA 15            | 15                | TACTAAATTATATTAC 16         | 16                       | AAAAAACAG 10 |
| B. papaya       | ACTAAATATAATACACTA 18         | 18                            | TGAAAAATTTTCA 15            | 15                | TACTAAATTATATTAC 16         | 16                       | AAAAAACAG 10 |
| B. correcta     | ACTAAATATAATACACTA 18         | 18                            | TGAAAAATTTTCA 15            | 15                | TACTAAATTATATTAC 16         | 16                       | AAAAAACAG 10 |
| B. curcubitae   | ACTAAATATAATACACTA 18         | 18                            | TGAAAAATTTTCA 15            | 15                | TACTAAATTATATTAC 16         | 16                       | AAAAAACAG 10 |
| C. capitata     | ACTAATAATAATAATACACTA 18      | 18                            | TGAAAAATTTTCA 15            | 15                | TACTAAATTATATTAC 16         | 16                       | AAAAAACAG 10 |

doi:10.1371/journal.pone.0100558.t007
an A-G mismatch and trnRtmR has a U-C mismatch in the acceptor stem. Additionally, trnV contains a U-U mismatch in the TPVC stem.

Anticodon sequences were the same as in B. dorsalis, B. oleae, B. tryoni and C. capitata, which are considered common for other insects, including Glycydidae orientalis [68], Philaenus spumarius [35], Phthornandria atrilineata [30] and Artogeia melite [36].

On the basis of the sequence similarity of B. dorsalis, the two genes coding for the small and the large ribosomal subunits were located in the B. minax mitogenome between trnV(UGN) and trnM and between trnV and the CR region. The length of B. minax rns and rmL was 782 bp and 1333 bp, respectively, similar to B. dorsalis, B. oleae and C. capitata.

5. Intergenic spacers

In B. minax, the two longest intergenic spacers were 42 bp between trnC and trnT and 28 bp between trnR and trnN. In B. dorsalis, the second longest intergenic spacer was 45 bp between trnC and trnT. In B. tryoni, the second longest intergenic spacer was 33 bp between trnR and trnN and the third longest intergenic spacer was 30 bp between trnC and trnT. In B. oleae, the longest intergenic spacer was 28 bp between trnR and trnN. In B. minax, however, only a 10 bp intergenic spacer was observed between trnQ and trnM, which is shorter compared to 66 bp in B. dorsalis, 71 bp in B. tryoni and 47 bp in C. capitata at the same location. Yu et al. [48] reported the 45 bp intergenic spacer located between trnC and trnT in B. dorsalis had a clear counterpart in the CR with the first 33 of 45 bp matching. These counterparts were predicted to form a small internal stem and a long stem-structure pairing with the partially complementary sequence in the CR. A similar phenomenon was observed in the B. tryoni mitogenome, where both the second longest (33 bp between trnR and trnN) and the third longest intergenic spacer (30 bp between trnC and trnT) have clear counterparts (32 out of 33 bases and 25 out of 30 bases, respectively) on the N-strand of the CR. Two of these intergenic spacers have highly significant similarity and their counterparts were located in the same position of the CR. We asked whether the 42 bp intergenic spacer located between trnC and trnT in B. minax had these features. The first 15/42 bp of the spacer have a clear counterpart in the CR at positions 13,670–13,684. The 42 bp of intergenic spacer was predicted to form two stem-loop secondary structures with 4 bp loops and one with a 3 bp stem and the other with a 4 bp stem. The first 15 of the 42 bp formed one of the two structures; a 4 bp stem with a 4 bp loop and a 3 bp flanking sequence. The counterpart in the CR also formed a long stem structure with the neighboring sequence. Yu et al. [48] compared the 33 bp counterpart in the CR from B. dorsalis with the B. oleae CR and found 25 of the 33 bp were identical. Surprisingly, of the original 33 bases present in the B. minax CR, 23 were identical. Therefore, the results obtained in this study support the hypothesis that the secondary structures of the counterparts in both the intergenic spacer and the CR might have a major role in recombination [48,69].

The four intergenic spacers in B. minax, ISS-1 (18 bp between trnE and trnF), ISS-2 (14 bp between nad5 and tRNAPro), ISS-3 (16 bp between trnG(GCU) and nad1) and ISS-4 (10 bp between nad1 and trnL(UUA)), were observed to be of similar size in the tephritids B. dorsalis, B. philippinensis, B. carambolae, B. papayae, B. dorsalis, C. capitata, B. minax, B. correcta and B. curcubitae [68,69]. The poly-T stretch runs from nucleotide positions from 15974 to 15997 with respect to the B. minax mitogenome in the direction of 5’-3’.

doi:10.1371/journal.pone.0100558.g005

Figure 5. Alignment of the poly-thymidine stretch at the 5’ end of the control region described by Zhang et al. (1997) among 10 tephritid species, B. oleae, B. tryoni, B. philippinensis, B. carambolae, B. papayae, B. dorsalis, C. capitata, B. minax, B. correcta and B. curcubitae. The poly-T stretch runs from nucleotide positions from 15974 to 15997 with respect to the B. minax mitogenome in the direction of 5’-3’.
stem-loop structures with conserved flanking sequences were not found in the CR of these ten tephritid species. In addition, the B. minax CR does not contain any tRNA-like sequence, but contains two tandem repeats ranging in size from 33 to 45 bp. The sequence TATTAATTaaaT occurred twice and the sequence CcTtTTAaATTCc occurred three times. The two repeats were located at positions from 15,325 to 15,357 and from 15,030 to 15,093, respectively. For other tephritid species, we found one tandem repeat in the CR of B. doraslis, B. correcta, B. curculioidea, and B. capitata, two in B. philippinensis and B. carabamble, three in B. olene and B. papaya but none in B. tryoni.

Acknowledgments

We thank Kostas Bourtis, Jane Skeltonand the anonymous reviewers whose invaluable critical comments helped greatly to improve this article.

Author Contributions

Conceived and designed the experiments: YL. Performed the experiments: BZ. Analyzed the data: BZ. Contributed reagents/materials/analysis tools: FN HH-S XW. Wrote the paper: BZ.

References

1. Thompson FC (1990) Fruit Fly Expert Identification System and Systematic Information Database. The International Journal of the North American Dipterists’ Society. 524p.
2. Wang XJ, Luo LY (1995) Research progress in the Chinese citrus fruit fly. Entomological Knowledge 32: 310–315.
3. Drew RA (1979) The genus Drosophila (Diptera: Tephritidae) two new species from northern Australia and a discussion of some subgenera. Australian Journal of Entomology 18: 71–80.
4. Allwood AJ, Chinajaiyawong A, Drew RA, Hamecke EL, Hancock DL, et al. (1995) Host plants for fruit flies (Diptera: Tephritidae) in South East Asia. The Raffles Bulletin of Zoological Supplement 7: 1–92.
5. Rieger M (2006) Citrus: lemon, lime, orange, tangerine, grapefruit-Citrus spp. Available: http://www.uga.edu/fruit/citrus/html. Accessed 9 February 2006.
6. White IM, Elson-Harris M (1992) Fruit flies of economic significance: their identification and biometrics. Wallingford, USA.
7. White IM, Wang XJ (1992) Taxonomic notes on some Dacine (Diptera: Tephritidae) fruit flies associated with citrus, olives, and cucurbits. Bulletin of Entomological Research 82: 275–279.
8. Boore JL (1999) Animal mitochondrial genomes. Nucleic Acids Research 27: 1767–1780.
9. Taaman JW (1999) The mitochondrial genome: structure, transcription, translation and replication. Biochimica et Biophysica Acta (BBA)-Bioenergetics 1410: 103–123.
10. Zhang DX, Szymura JM, Hewitt GM (1995) Evolution and structural conservation of the control region of insect mitochondrial DNA. Journal of Molecular Evolution 40: 382–391.
11. Lewis DL, Farr CL, Kaguni LS (1995) Drosophila melanogaster mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons. Insect Molecular Biology 4: 263–278.
12. Inohira K, Hara T, Matsura ET (1997) Nucleotide sequence divergence in the Av-T-rich region of mitochondrial DNA in Drosophila simulans and Drosophila mauritiana. Molecular Biology and Evolution 14: 814–822.
13. Shao R, Campbell NJ, Barker SC (2001) Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, Heterodoxus macropus (Phthiraptera). Molecular Biology and Evolution 18: 858–865.
14. Morizama EN, Powell JR (1997) Synonymy substitution rates in Drosophila: mitochondrial versus nuclear genes. Journal of Molecular Evolution 45: 376–384.
15. Cameron SL (2014) Insect mitochondrial genomics implications for evolution and phylogeny. Annual Review of Entomology 59: 95–117.
16. Ma C, Yang P, Jiang F, Chapuis MP, Shali Y, et al. (2012) Mitochondrial genases reveal the global phylography and dispersal routes of the migratory locust. Molecular Ecology 21: 4344–4358.
17. Beckenbach AT (2012) Mitochondrial genome sequences of Nematocera (Lower Diptera): evidence of rearrangement following a complete genome duplication in a winter crane fly. Genome Biology and Evolution 4: 89–101.
18. Cameron SL, Miller KB, D’Haese CA, Whiting MF, Barker SC (2004) Mitochondrial genome data alone are not enough to unambiguously resolve the relationships of Entognatha, Insecta and Crustacea sensu lato (Arthropoda). Cladistics 20: 534–557.
19. Wan X, Kim MI, Kim MJ, Kim I (2012) Complete mitochondrial genome of the free-living earwig, Challa fletcheri (Dermaptera: Pygidicranidae) and phylogeny of Polynoptera. PLoS ONE 7(8): e42056. doi:10.1371/journal.pone.0042056.
20. Nardi F, Spinsanti G, Boveri JL, Caragelli A, Dalila R, et al. (2003a) Hexapod origins: monophyletic or paraphyletic? Science 299: 1817–1819.
21. Nelson LA, Lambkin CL, Batterham P, Wallman JF, Dowton M, et al. (2012) Beyond barcoding: A mitochondrial genomics approach to molecular phylogenetics and diagnostics of blowflies (Diptera: Calliphoridae). Gene 311: 131–142.
22. Castro LR, Austin AD, Dowton M (2009) Contrasting rates of mitochondrial molecular evolution in parasitic Diptera and Hymenoptera. Molecular Biology and Evolution 19: 1100–1113.
23. Salvato P, Simonato M, Battini A, Negriolo E (2008) The complete mitochondrial genome of the bag-shelter moth Ophiusa baauj (Lepidoptera: Notodontidae). BMC Genomics 9: 331.
24. Cameron SL, Yoshizawa K, Mizukoshi A, Whiting MF, Johnson KP (2011) Mitochondrial genome deletions and minicircles are common in lice (Insecta: Phthiraptera). BMC Genomics 12: 394.
25. Dowton M, Cameron SL, Doscovic J, Austin AD, Whiting MF (2009) Characterization of 67 mitochondrial tRNA gene rearrangements in the Hymenoptera suggests that mitochondrial tRNA gene position is selectively neutral. Molecular Biology and Evolution 26: 1607–1617.
26. Shao R, Kirkness EF, Barker SC (2009) The single mitochondrial chromosome typical of animals has evolved into 18 minichromosomes in the human body, Pellicula humana. Genome Research 19: 904–912.
27. Krosch MN, Schutze MK, Armstrong KP, Boetnig Y, Boykin LM, et al. (2013) Piecing together an integrative taxonomic puzzle: microsatellite, wing shape and aedoeagus length analyses of Bactrocera durius s.l. (Diptera: Tephritidae) find no evidence of multiple lineages in a proposed contact zone along the Thai/Malay Peninsula. Systematic Entomology 38: 2–13.
28. Jose MS, Leblanc L, Geb SM, Rubinoff D (2013) An evaluation of the species status of Bactrocera invadens and the systematics of the Bactrocera durius (Diptera: Tephritidae) complex. Annals of the Entomological Society of America 106: 684–694.
29. Boykin LM, Schutze MK, Krosch MN, Chomait A, Chapman TA, et al. (2013) Multi-gene phylogenetic analysis of south-east Asian pest members of the Bactrocera durius species complex (Diptera: Tephritidae) does not support current taxonomy. Journal of Applied Entomology in press.
30. Simon C, Frati F, Beckenbach A, Crespi B, Liu H, et al. (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Annals of the Entomological Society of America 87: 651–701.
31. Lowe TM, Eddy SR (1997) tRNA-scan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Research 25: 955–964.
32. Stetten P, Voje B, Rehmbecker M, Reeder J, Giegerich R (2006) RNAshapes: an integrated RNA analysis package based on abstract shapes. Bioinformatics 22: 500–503.
33. Stewart JB, Beckenbach AT (2005) Phylogenetic and genomic analysis of the complete mitochondrial DNA sequence of the spotted asparagus beetle Craspedosia annulata (Lepidoptera: Notodontidae). BMC Evolutionary Biology 12: 130.
34. Stewart JB, Beckenbach AT (2003) Phylogenetic and genomic analysis of the complete mitochondrial genome of the bag-shelter moth Ophiusa baauj (Lepidoptera: Notodontidae). BMC Genomics 9: 331.
35. Stewart JB, Beckenbach AT (2005) Insect mitochondrial genomics: the complete mitochondrial genome sequence of the meadow spittlebug Philaenus spumarius (Hemiptera: Auchenorrhyncha: Cercopidae). Genome 48: 46–54.
36. Hong GY, Jiang ST, Yu M, Yang Y, Li F, et al. (2009) The complete nucleotide sequence of the mitochondrial genome of the cabbage butterfly, *Argea inodora* (Lepidoptera: Pieridae). Acta Biochimica et Biophysica Sinica 41: 446–455.
37. Song N, Liang AP (2009) The complete mitochondrial genome sequence of *Geikea distinctivina* (Hemiptera: Flatidae) and comparison with other hemipteran insects. Acta Biochimica Et Biophysica Sinica 41: 206–216.
38. Wang M, Liu YS, Zhou JH, Chen HT, Ma LN, et al. (2011) Analysis of codon usage in newcastle disease virus. Virus Genes 42: 245–253.
39. Zuber M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Research 31: 3406–3415.
40. Benson G (1999) Tandem repeat finder: a program to analyze DNA sequences. Nucleic Acids Research 27: 573–580.
41. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular evolutionary genetic analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24: 1596–1599.
42. Beckenbach AT, Joy JB (2009) Evolution of the mitochondrial genomes of gall midges (Diptera: Cecidomyiidae): rearrangement and severe truncation of tRNA genes. Genome Biology and Evolution 1: 278–287.
43. Friedrich M, Muqim N (2003) Sequence and phylogenetic analysis of the complete mitochondrial genome of the flour beetle *Tribolium castaneum*. Molecular Phylogenetics and Evolution 26: 502–512.
44. Nardi F, Carapelli A, Boore JL, Roderick GK, Dallai R, et al. (2003b) The mitochondrial genome of the Mormon cricket (Anabrus simplex: Tettigoniidae) and an analysis of control region variability. Insect Molecular Biology 12: 605–611.
45. Steward JB, Beckenbach AT (2006) Insect mitochondrial genomics 2: the complete mitochondrial genome sequence of a giant fly, *Hybomitra nitens*, asymmetric directional mutation bias, and conserved plecopteran A+T-region elements. Genome 49: 815–824.
46. Penn JD, Cameron SL, Whiting MF (2007) The complete mitochondrial genome sequence of the Mormon cricket (Anabrus simplex: Tettigoniidae: Orthoptera) and an analysis of control region variability. Insect Molecular Biology 16: 239–252.
47. Yu DJ, Xu L, Nardi F, Li JG, Zhang RQ (2007) The complete nucleotide sequence of the mitochondrial genome of the oriental fruit fly, *Bactrocera dorsalis* (Diptera: Tephritidae). Gene 396: 66–74.
48. Cha SY, Yoon HJ, Lee EM, Yoon MH, Hvang JS, et al. (2007) The complete nucleotide sequence and gene organization of the mitochondrial genome of the bumblebee, *Bombus ignitus* (Hymenoptera: Apidae). Gene 392: 206–220.
49. Yang L, Wei ZJ, Hong GY, Jiang ST, Wen LP (2009) The complete nucleotide sequence of mitochondrial genome of *Pithanodra attiliana* (Lepidoptera: Geometridae). Molecular Biology Reports 36: 1441–1449.
50. Lessinger AC, Martins-Junqueira AC, Lemos TA, Kemper EL, da Silva FR, et al. (2000) The mitochondrial genome of primary screwworm fly *Goniopoda hominicornis* (Diptera: Calliphoridae). Insect Molecular Biology 9: 521–529.
51. Song H, Sheffield NC, Cameron SL, Miller KB, Whiting MF (2010) When phylogenetic assumptions are violated: base compositional heterogeneity and among-site rate variation in beetle mitochondrial phylogenomics. Systematic Entomology 35: 493–504.
52. Collins TM, Fedrigo O, Naylor GJ (2005) Choosing the best genes for the job: the case for stationary genes in genome-scale phylogenetics. Systematic Biology 54: 493–500.
53. Cameron SL, Lembäti CL, Barker SC, Whiting MF (2007a) A mitochondrial genome phylogeny of Diptera: whole genome sequence data accurately resolve relationships over broad timescales with high precision. Systematic Entomology 32: 40–59.
54. Yoon MJ, Song H, Cameron SL, Whiting MF (2008) A preliminary mitochondrial genome phylogeny of Orthoptera (Insecta) and approaches to maximizing phylogenetic signal found within mitochondrial genome data. Molecular Phylogenetics and Evolution 49: 59–68.
55. Sheffield NC, Song H, Cameron SL, Whiting MF (2009) Nonstationary evolution and compositional heterogeneity in beetle mitochondrial phylogenomics. Systematic Biology 58: 381–394.
56. Jermini LS, Crozier RH (1994) The cytochrome b region in the mitochondrial DNA of the ant *Takepanotoma rhytida*: sequence divergence in hymenoptera may be associated with nucleotide content. Journal of Molecular Evolution 38: 282–294.
57. Nei M (1987) Positive darwinian selection observed at the variable region genes of immunoglobulins. Molecular Biology and Evolution 5: 447–459.
58. Hughes AL, Nei M (1989) Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proceedings of the National Academy of Sciences 86: 958–962.
59. Cameron SL, Lo N, Bourguignon T, Svenson GJ, Evans TA (2012) A mitochondrial genome phylogeny of termites (*Blattodea*: Termitoidea): Robust support for interfamilial relationships and molecular synapomorphies define major clades. Molecular Phylogenetics and Evolution 65: 163–173.
60. Wolstenholme DR (1992) Animal mitochondrial DNA: structure and evolution. International Review of Cytology 141: 173–216.
61. Ohnuki T, Kawai G, Watanabe K (2008) The minimal tRNA: unique structure, evolution and usefulness in evolutionary studies. Biochemical and Biophysical Research Communications 373: 37–43.
62. Kim I, Cha SY, Yoon MH, Hvang JS, Lee SM, et al. (2005) The complete nucleotide sequence and gene organization of the mitochondrial genome of the oriental mole cricket, *Gryllotalpa orientalis* (Orthoptera: Gryllotalpidae). Gene 355: 155–168.
63. Stanton DJ, Daehler LL, Moritz CC, Brown WM (1994) Sequences with the potential to form stem-and-loop structures are associated with coding region duplications in animal mitochondrial DNA. Genetics 137: 233–241.
64. Cameron SL, Johnson KP, Whiting MF (2007b) The mitochondrial genome of the screamer *Loise Bechsteinotus* (*Phalaripatra*: Ichnocoera): effects of extensive gene rearrangements on the evolution of the genome. Journal of Molecular Evolution 65: 589–604.
65. Zhang DX, Hewitt GM (1997) Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. Biochemical Systematics and Ecology 25: 99–120.