Gene 569 (2015) 191–202

Contents lists available at ScienceDirect

Gene

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

Research paper

The mitochondrial genome of the peach fruit fly, *Bactrocera zonata* (Saunders) (Diptera: Tephritidae): Complete DNA sequence, genome organization, and phylogenetic analysis with other tephritids using next generation DNA sequencing

Jaipal S. Choudhary [a,*], Naiyar Naaz [a], Chandra S. Prabhakar [a,b], Mathukumalli Srinivasa Rao [c], Bikash Das [a]

[a] ICAR Research Complex for Eastern Region, Research Centre, Plandu, Ranchi 834 010, Jharkhand, India
[b] Department of Entomology, Bihar Agricultural University, Sabour, Bhagalpur 813 210, Bihar, India
[c] Central Research Institute for Dryland Agriculture (CRIDA), Santoshnagar, Saidabad PO, Hyderabad 500 059, India

ARTICLE INFO

Article history:
Received 6 April 2015
Received in revised form 13 May 2015
Accepted 27 May 2015
Available online 30 May 2015

Keywords:
Bactrocera zonata
Mitogenome
Tephritidae
Phylogenetic relationship
Genome organization

ABSTRACT

Mitochondrial genome can provide information for genomic structure as well as for phylogenetic analysis and evolutionary biology. The complete 15,935 bp mitochondrial genome of *Bactrocera zonata* (Diptera: Tephritidae), is assembled from Illumina MiSeq read data. The mitogenome information for *B. zonata* was compared to the homologous sequences of other tephritids. Annotation indicated that the structure and orientation of 13 protein coding genes (PCGs), 22 tRNA and 2 rRNA sequences were typical of, and similar to, the ten closely related tephritid species. The nucleotide composition shows heavily biased toward As and Ts accounting 73.34% and exhibits a slightly positive AT skew, which is similar to other known tephritid species. All PCGs are initiated by ATN codons, except for *cox1* with TCG and *atp8* with GTG. Nine PCGs use a common stop codon of TAA or TAG, whereas the remaining four use an incomplete termination codon T or TA likely to be completed by adenylation. All tRNAs have the typical clover-leaf structure, with an exception for $trnS^{(AGN)}$. Four short intergenic spacers showed high degree of conservation among *B. zonata* and other ten tephritids. A poly(T) stretch at the 5′ end followed by [TA(A)]n-like stretch and a tandem repeats of 39 bp has been observed in CR. The analysis of gene evolutionary rate revealed that the *cox1* and *atp6* exhibits lowest and highest gene substitution rates, respectively than other genes. The phylogenetic relationships based on Maximum Likelihood method using all protein-coding genes and two ribosomal RNA genes confirmed that *B. zonata* is closely related to *Bactrocera correcta*, *Bactrocera carambolae*, *Bactrocera papayae*, and *Bactrocera philippinensis* and *Bactrocera dorsalis* belonging to *B. dorsalis* species complex forms a monophyletic clade, which is in accordance with the traditional morphological classification and recent molecular works.

1. Introduction

The peach fruit fly, *Bactrocera zonata* (Saunders), a member of family Tephritidae is a serious pest of fruit crops in many parts of the world. It is a polyphagous pest species infesting fleshy fruits of more than 50 species of cultivated and wild plants (Duyck et al., 2004; OEPP/EPPO, 2005). Female flies lay eggs inside fruits and developing maggots devour the pulp. Subsequently, secondary infestation by bacterial and fungal pathogens on oviposition punctures causes dropping of infested rotten fruits (White and Elson-Harris, 1994). Owing to its enormous damage potential, it is considered as one of the most destructive fruit pests species of peach, guava, mango in temperate, tropical and subtropical climatic countries (Joomaye and Price, 2000; Hashem et al., 2001).

B. zonata is originally an Indian species and was first recorded from Bengal (Kapoor, 1993). It is, however at present distributed in many Asian countries viz. Bangladesh, Bhutan, India, Iran, Laos, Myanmar, Nepal, Oman, Pakistan, Saudi Arabia, Sri Lanka, Thailand, United Arab Emirates, Vietnam and Yemen (Kapoor, 1993) and in some parts of Africa (White and Elson-Harris, 1992). The pest species was recorded in Egypt as early as in 1924 (Efflatoun, 1924) and recently recognized attacking a range of fruit species including mango, guava, apricot, peach, apple, fig (El-Minshawy et al., 1999; Mosleh et al., 2011). It was

Abbreviations: Mitogenome, mitochondrial genome; mtDNA, mitochondrial DNA; PCGs, protein-coding genes; atp, F_0 ATPase; *cox3*, cytochrome oxidase subunits; *cob*/*cytb*, cytochrome B; *nad*, NADH dehydrogenase; *rrnL*, large subunit ribosomal RNA; *rrnS*, small subunit ribosomal RNA; tRNA, transfer RNA; *trn*, genes encoding transfer RNA molecules; CR, control region.

* Corresponding author.
E-mail addresses: choudhary.jaipal@gmail.com, jaipal.choudhary@icar.gov.in (J.S. Choudhary), naaz19biotech@gmail.com (N. Naaz), csprabhakar.ento@gmail.com (C.S. Prabhakar), msrao909@gmail.com (M.S. Rao), bikash41271@gmail.com (B. Das).

http://dx.doi.org/10.1016/j.gene.2015.05.066
0378-1119/

also trapped from California of North America but was eradicated later on (Spaugy, 1988).

The mitochondrial genome of insects typically encodes a set of 13 protein-coding genes (PCGs), 22 transfer RNAs (tRNAs) and 2 ribosomal RNAs (rRNAs) in a compact double-stranded closed circular DNA molecule conserved across bilateral metazoan that ranges in size from 14 to 40 kbp approximately (Wolstenholme, 1992; Boore, 1999). The mitogenome usually contains one A + T rich non-coding element called control region (CR) (Zhang et al., 1995; Lewis et al., 1995; Inohira et al., 1997; Shao et al., 2001) and the sites for genome replication and the initiation of gene transcription (Boore, 1999; Taanman, 1999). The mitochondrial genome content and organization has become one of the most useful markers for studies of comparative and evolutionary genomics, molecular evolution, phylogenetics, phylogeography and population genetics (Avise, 2000; Cameron et al., 2007) due to its simple maternal inheritance, the relatively high mutation rate, the availability of comparative data and conserved PCR primers that can amplify a number of fragments in most of hexapod taxa (Simon et al., 1994).

The documented information indicates that very limited studies have been done to date on *B. zonata* population genetic structure and geographical variability based on partial *mtCOI* gene sequences, barring few studies that have focused on very vague aspects of genetic distance between populations of *B. zonata* with many other pairs of distinctly different *Bactrocera* species (Muraji and Nakahara, 2001; Nakahara and Muraji, 2008; Zhang et al., 2010; Asokan et al., 2011). A promising, but still preliminary study on RT-PCR of partial *mtCOI* gene sequence from *B. zonata* in Egypt by Abd-El-Samie and El Fiky (2011) established mitochondrial sequence data. Tephritidae has been estimated to contain over 4257 species in 471 genera worldwide (Thompson, 1998); however, the mitogenome information is available for few species only. To date, only eleven mitogenomes of Tephritidae species are available in GenBank, including *Bactrocera papayae* (DQ917578), *Bactrocera correcta* (JX456552), *Bactrocera carambolae* (EF014414), *Bactrocera dorsalis* (DQ845759; Yu et al., 2007), *Bactrocera philippinensis* (DQ995281), *Bactrocera tryoni* (HQ130030), *Bactrocera oleae* (AY210702; Nardi et al., 2003), *Bactrocera minax* (HM776033; Zhang et al., 2014), *Bactrocera cucurbitae* (JN635562; Wu et al., 2012), *Ceratitis capitata* (AJ242872) and *B. zonata* (KP296150; submitted from the present study). All of these species belongs to the genus *Bactrocera*, except *C. capitata* which belongs to the genus *Ceratitis*. *B. carambolae*, *B. papaya*, *B. philippinensis* and *B. dorsalis* (Yu et al., 2007) belong to *B. dorsalis* species complex, *B. correcta*, *B. zonata* and *B. tryoni* belong to other species-group within the subgenus *Bactrocera*, and *B. oleae* (Nardi et al., 2003), *B. minax* (Zhang et al., 2014) and *B. cucurbitae* (Wu et al., 2012) belong to the subgenus *Daculus*, *Tetradacus* and *Zeugodacus*, respectively. Recently, *B. dorsalis* acquired the status as a senior synonym of *B. papaya*, *B. philippinensis* & *B. invadens* and these species are now known as *B. dorsalis* (Schutze et al., 2015).

In the present study, the complete mitogenome sequence of *B. zonata* selected from an Indian population has been described and sequence/genome organization has been compared with ten other available complete mitogenomes of tephritid fruit flies viz. *B. papayae*, *B. correcta*, *B. carambolae*, *B. dorsalis*, *B. philippinensis*, *B. tryoni*, *B. oleae*, *B. minax*, *B. cucurbitae* and *C. capitata*. Apart, the phylogenetic analysis based on the available complete mitogenome sequences of Tephritidae and other Diptera was performed to provide insight into the phylogenetic relationship of Tephritidae. The work provides the reference sequence of mitogenome for *B. zonata* that may be utilized for the determination of population genetic studies in the future.

2. Materials and methods

2.1. Sequencing and assembly of mitochondrial genome

The *B. zonata* culture was initiated under laboratory conditions from infested mango fruits collected from the experimental farm of ICAR Research Complex for Eastern Region, Research Centre, Ranchi, India (23° 45′ N latitude, 85° 30′ E longitude, elevation 620 m AMSL) (Prabhakar et al., 2009) and emerged adults were identified on the basis of morphological descriptions given by Drew and Raghu (2002), Madhura and Verghese (2004) and Prabhakar et al. (2012).

Mitochondrial DNA extraction was carried out from ~30 mg of adult flies using BioVision Mitochondrial DNA Isolation Kit as per the instructions of manufacturer (BioVision, California, US). Illumina library preparation and sequencing was performed at the Xcelris Labs Limited, Ahmedabad, India. The paired-end sequencing library was prepared using Illumina Nextera DNA XT Library Preparation Kit (Illumina Inc.) according to the manufacturer's protocol. Library preparation was started with mtDNA fragmentation of 1 ng input, followed by paired-end adapter ligation. The ligated products were purified using 0.6× Ampure beads. The purified ligated products were amplified though PCR to achieve size range of 500–1000 bp. The amplified library was analyzed through Bioanalyzer 2100 (Agilent Technologies) using High Sensitivity (HS) DNA chip as per manufacturer's instructions. Sequencing was performed on Illumina MiSeq Sequencing System (Illumina Inc.) using 2 × 300 bp chemistry kit. After obtaining the Qubit concentration for the library and the mean peak size from Bioanalyser profile, the library was normalized to 4 nM concentration and further diluted to 20 pM concentration. The strength of final library loaded into MiSeq was 11 pM. Paired-end sequencing allows the template fragments to be sequenced in both forward and reverse directions on MiSeq. The kit reagents were used in binding of samples to complementary adapter oligos on paired-end flow cell. The adapters were designed to allow selective cleavage of the forward strands after re-synthesis of the reverse strand during sequencing. The copied reverse strand was then used to sequence from the opposite end of the fragment. Fastq file was preprocessed and trimmed using the FASTX-Tool Kit. A total of 40 million reads (~4.0 Gbp) from each of trimmed fastq sequence were loaded into the GS De novo Assembler program v.2.5.3 (Roche 454 Life Sciences, Branford, CT, USA) and assembled using the default parameters. Resulted contigs containing mitochondrial genome sequence were identified by querying the National Center for Biotechnology Information (NCBI) non-redundant (nr) nucleotide database with the blastn algorithm (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

2.2. Mitochondrial genome annotation

Protein coding, tRNA and rRNA gene coding regions were predicted using the MITOS web server (Bernt et al., 2012) using the invertebrate mitochondrial genetic code. Protein-coding regions and ribosomal RNA genes were also annotated manually and confirmed by comparison to the sequences of other 10 tephritid species mitogenome sequences available in Genbank viz. *B. papaya*, *B. correcta*, *B. carambolae*, *B. dorsalis*, *B. philippinensis*, *B. tryoni*, *B. oleae*, *B. minax*, *B. cucurbitae* and *C. capitata*. Two-dimensional tRNA structures that were downloaded in MITOS output were rechecked using tRNAscan-SEv1.21 (http://lowelab.ucsc.edu/cgi-bin/tRNAscan-SE.cgi) with a covariance cut-off value of 15.0 using the Mito/Chloroplast model (Lowe and Eddy, 1997). The circular map, GC content and GC skew of the mitochondrial genome were drawn with the CGView server V 1.0 (Grant and Stothard, 2008). Mitochondrial protein-coding regions were concatenated and aligned using ClustalW with default parameters implemented in MEGA 6.0 (Tamura et al., 2013). The A + T content, codon usage and relative synonymous codon usage (RSCU) analysis in PCGs were calculated using MEGA 6.0. Secondary structure folding of the repeated fragment and free energy values were predicted for the trnC-spacer-trnY region and in an equally long area in the CR using the DNA mfold web server with default parameters (http://mfold.bioinfo.rpi.edu/cgi-bin/dna-form1.cgi) (Zuker, 2003). The composition of skew analysis was carried out using formulae AT skew = [A − T] / [A + T] and GC skew = [G − C] / [G + C] (Perna and Kocher, 1995). The overlapping regions and intergenic spacers between genes were counted manually. The synonymous substitution

rate (Ks) and the non-synonymous substitution rate (Ka) with Jukes–Cantor adjusted Ka/Ks (JKa/JKs) for each PCG were calculated using software packages DnaSP 5.0 (Librado and Rozas, 2009). The complete assembled and annotated *B. zonata* mtDNA genome sequence was deposited in NCBI Genbank under accession number: KP296150.

2.3. Phylogenetic analysis

A total of 18 Diptera mitogenomes including eleven Tephritidae, two Muscidae, three Drosophilidae and two Culicidae were used to reconstruct the phylogenetic relationships. Beside this, mitogenomes of Lepidopteran moth, *Attacus atlas* (Chen et al., 2014) was used as outgroup. For phylogenetic analyses, datasets of concatenated nucleotide sequences of 13 PCGs plus 2 rRNAs were used. Nucleotide sequences of 13 PCGs and 2 rRNAs were concatenated into a single sequence in order *nad2*, *cox1*, *cox2*, *atp8*, *atp6*, *cox3*, *nad3*, *nad5*, *nad4*, *nad4L*, *nad6*, *cytb*, *nad1*, *16S rRNA* and *12S rRNA*. Concatenated nucleotide sequences were loaded into the MEGA 6.0 alignment utility (Tamura et al., 2013), and a multiple sequence alignment was constructed using ClustalW using default parameters. These concatenated sequences of the 13 PCGs and 2 rRNAs may give result in a more complete analysis (Hassanin, 2006). The analytical approach, Maximum Likelihood (ML) was used to infer phylogenetic trees with 1000 bootstrap replicates. Substitution model selection was also conducted based on the lowest BIC scores (Bayesian Information Criterion) using MEGA 6.0. The GTR + G + I model was found to be an appropriate model for the nucleotide sequence dataset according to the lowest Bayesian information criterion (BIC). At the same time, the phylogenetic analysis was also conducted without *atp6* with the above same procedure.

3. Result & discussion

3.1. Sequencing and assembly of mitochondrial genome

Next Generation Sequencing (NGS) such as 454 pyrosequencing, Solexa and SOLiD provided by Roche, Illumina and Applied Biosystems have the ability to generate a large number of sequences within a very short time when compared to Sanger's method of sequencing (Chilana et al., 2012). This has been applied for the rapid sequence data generation and successful de novo sequence assemblies for arthropods (Kirkness et al., 2010) and small low-repetitive bacterial genomes (Ribeiro et al., 2012), and more recently applied for the assembly of full mitochondrial genomes (Knaus et al., 2011; Ma et al., 2012; Coates, 2014) and complete genome sequencing of herbarium and insect museum specimens (Staats et al., 2013). In the present study, genomic DNA libraries were used to obtain mitochondrial DNA sequence data for *B. zonata*. Specifically, a total of 45.1 million reads were obtained from single end reads from a single Illumina MiSeq sequencing system run for library. Our assembly of 45 million genomic DNA reads from the library resulted in 18,950 contig (55,890 kb total contig length;

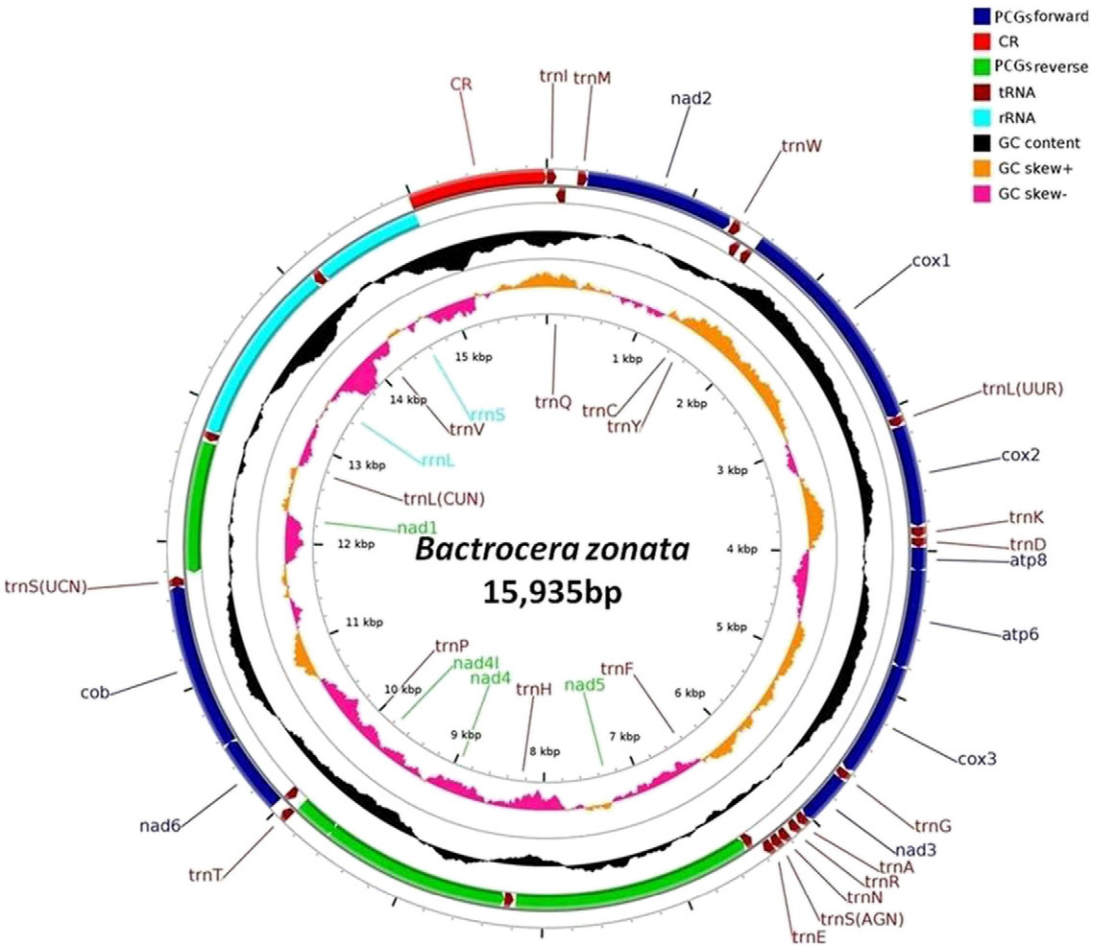

Fig. 1. Mitochondrial genome map of *B. zonata*. From outer to inner, the 1st circle shows the gene map (PCGs, rRNA, trn & CR) and tRNA genes are abbreviated by one letter symbols according to the IUPAC-IUB single-letter amino acid codes. The 2nd circle shows the GC content and the 3rd shows GC skew calculated as (G − C) / (G + C). GC content and GC skew are plotted as the deviation from the average value of the entire sequence.

range 1000 to 1, 26, 582 bp). The functional annotations for 18,950 contigs were performed by aligning them to non-redundant database of NCBI using blastn. Contig-1 was aligned to *B. correcta* isolate F11 mitochondrion, complete genome (JX456552) and the sequence analysis result of contig-1 of *B. zonata* against *B. correcta* gives E-value 0.0, total score 24,471, 31 gaps and 94% identities.

3.2. Genome organization

The complete mitogenome of *B. zonata* (Genbank accession: KP296150) is a closed-circular molecule of 15,935 bp in length, which is well within the range observed in the completely sequenced other ten tephritid mitogenomes available with the size ranging from 15,815 bp in *B. oleae* to 16,043 bp in *B. minax*. It presents the typical set of 37 genes observed in metazoan mitogenomes, including 13 PCGs (*cox1-3*, *cob*, *nad1-6*, *nad4l*, *atp6* and *atp8*), 22 tRNA genes (one for each amino acid, two each for Leucine and Serine), 2 genes for ribosomal RNA subunits (*rrnS* and *rrnL*) (Fig. 1, Supp. Table 1). The gene order in the *B. zonata* mitogenome follows the typical and plesiomorphic state hypothesized for the Pancrustacea arrangement (Crease, 1999). Only a long uninterrupted non-coding region of 950 bp, likely homologous to the insect A + T rich region by positional homology, general structure and base content, is present between *rrnS* and *trnI*, corresponding to position 14,986 to 15,935 in the annotated sequence (Fig. 1). Genome organization of *B. zonata* is very compact, with only 239 nucleotides dispersed in 15 intergenic spacers from 2 to 78 bp and contiguous genes overlapping at 10 boundaries by a total of 33 bases from 1 to 8 bp.

3.3. Nucleotide composition and skewness

Base composition shows both genome wide and strand specific compositional biases. Similar to other insect sequences, the *B. zonata* mitogenome nucleotide composition is biased toward adenine and thymine (accounting for 73.3%: A = 39.1%, T = 34.2%, G = 10.4% and C = 16.3%). This bias is well within the range observed for the sequenced Tephritidae species, from 67.28% in *B. minax* to 77.48% in *C. capitata*. The A + T content of separately analyzed all protein coding genes (70.9%), tRNAs (74.9%), rRNAs (78.1%) and the CR (84.4%) is also well within the range observed in the completely sequenced mitogenomes of 10 tephritid species used in the study (Table 1). When two strands were analyzed separately, the PCGs on the J-strand (majority-strand with 9 PCGs) have a lower A + T content (68.9%) compared to 74.0% in N-strand (minority-strand with other four PCGs). Furthermore, PCGs encoded on the J-strand and N-strand show a strong bias for T content (37.0% and 47.6%) compared to A content (32.0% and 26.6%), respectively. Similar observations have been recorded in case of other tephritid mitogenomes by Nardi et al. (2003, 2010), Yu et al. (2007) and Zhang et al. (2014) and in other insects (Stewart and Beckenbach, 2003, 2005, 2006; Hong et al., 2009; Song and Liang, 2009; Friedrich and Muqim, 2003; Fenn et al., 2007; Cha et al., 2007; Yang et al., 2009).

The nucleotide bias was also reflected in the codon usage. Base composition at each codon position of concatenated 13 PCGs showed that the A + T content of the third codon positions were significantly higher than the first and second positions owning to genetic code degeneracies. In particular, T in each codon position of PCGs is over-represented. Among all the three codon positions, the second codon position has over-representing T, which is similar to many insect mitogenomes (Stewart and Beckenbach, 2003, 2005, 2006; Hong et al., 2009; Song and Liang, 2009; Friedrich and Muqim, 2003; Fenn et al., 2007; Cha et al., 2007; Yang et al., 2009). The G content is higher compared to C at all the three codon positions, in PCGs on the N-strand, while G is under-represented compared to C in PCGs on the J-strand. This base compositional bias is parallel with the common trend in the mitogenome toward a lower G content (Lessinger et al., 2000).

Table 1
Length, base composition and skewness of different genomic regions in 11 tephritid species, *B. zonata*, *B. papayae*, *B. correcta*, *B. carambolae*, *B. dorsalis*, *B. philippinensis*, *B. tryoni*, *B. oleae*, *B. minax*, *B. cucurbitae* and *C. capitata*.

Accession no. and species	Whole mtDNA				PCGs				tRNAs				rRNAs				CR			
	Size	(A + T)%	AT skew	GC skew	Size	(A + T)%	AT skew	GC skew	Size	(A + T)%	AT skew	GC skew	Size	(A + T)%	AT skew	GC skew	Size	(A + T)%	AT skew	GC skew
KP296150 *Bactrocera zonata*	15,935	73.34	0.065	−0.223	11,190	70.99	0.070	−0.224	1469	74.95	−0.017	0.109	2120	78.07	0.077	−0.308	950	84.42	0.040	−0.216
DQ917578 *Bactrocera papayae*	15,915	73.52	0.066	−0.226	11,190	71.04	0.073	−0.234	1466	75.38	0.024	−0.086	2114	77.72	0.069	−0.287	950	88.21	0.064	−0.161
JX456552 *Bactrocera correcta*	15,936	73.17	0.063	−0.222	11,192	71.22	0.070	−0.221	1470	75.31	0.024	−0.113	2117	77.85	0.080	−0.322	949	78.61	0.019	−0.163
ER014414 *Bactrocera carambolae*	15,915	73.55	0.066	−0.224	11,192	71.10	0.072	−0.228	1466	75.10	0.019	−0.096	2113	77.57	0.079	−0.304	950	87.89	0.049	−0.130
DQ845759 *Bactrocera dorsalis*	15,915	73.58	0.068	−0.228	11,185	71.12	0.073	−0.235	1467	75.19	0.017	−0.093	2123	77.81	0.076	−0.287	949	88.09	0.062	−0.186
DQ995281 *Bactrocera philippinensis*	15,915	73.63	0.066	−0.224	11,192	71.18	0.071	−0.229	1466	75.31	0.020	−0.099	2114	77.67	0.074	−0.292	949	88.20	0.061	−0.179
HQ130030 *Bactrocera tryoni*	15,925	72.42	0.070	−0.227	11,187	69.61	0.074	−0.236	1467	74.98	0.022	−0.090	2115	77.73	0.073	−0.295	951	86.96	0.083	−0.129
AY210702 *Bactrocera oleae*	15,815	72.63	0.088	−0.280	11,189	70.18	0.091	−0.288	1466	74.76	0.040	−0.130	2116	77.13	0.099	−0.347	949	86.93	0.091	−0.177
HM776033 *Bactrocera minax*	16,043	67.28	0.131	−0.316	11,183	64.30	0.148	−0.319	1466	72.31	0.055	−0.182	2115	73.71	0.121	−0.356	1141	77.65	0.081	−0.333
JN635562 *Bactrocera cucurbitae*	15,825	72.89	0.047	−0.213	11,190	70.71	0.049	−0.226	1467	75.12	0.005	−0.074	2110	77.82	0.080	−0.303	946	82.35	0.014	0.042
AJ242872 *Ceratitis capitata*	15,980	77.48	0.021	−0.185	11,272	75.59	0.019	−0.178	1472	76.77	0.021	−0.094	2123	80.22	0.058	−0.267	1004	91.14	−0.010	−0.258

Table 2
Codon usage and relative synonymous codon usage within *B. zonata* mitochondrial genome.

	U				C				A				G			
	Codon		Count	RSCU	Codon		Count	RSCU	Codon		Count	RSCU	Codon		Count	RSCU	
U	UUU	Phe	143	1.19	UCU	Ser	60	1.48	UAU	Tyr	119	1.34	UGU	Cys	24	1.45	U
	UUC	Phe	98	0.81	UCC	Ser	26	0.64	UAC	Tyr	58	0.66	UGC	Cys	9	0.55	C
	UUA	Leu	219	2.75	UCA	Ser	98	2.41	UAA	Stop	105	1.49	UGA	Trp	67	1.7	A
	UUG	Leu	24	0.3	UCG	Ser	11	0.27	UAG	Stop	36	0.51	UGG	Trp	12	0.3	G
C	CUU	Leu	68	0.86	CCU	Pro	62	1.47	CAU	His	49	1.01	CGU	Arg	13	0.91	U
	CUC	Leu	33	0.42	CCC	Pro	40	0.95	CAC	His	48	0.99	CGC	Arg	8	0.56	C
	CUA	Leu	117	1.47	CCA	Pro	64	1.51	CAA	Gln	89	1.58	CGA	Arg	28	1.96	A
	CUG	Leu	16	0.2	CCG	Pro	3	0.07	CAG	Gln	24	0.42	CGG	Arg	8	0.56	G
A	AUU	Ile	211	1.57	ACU	Thr	69	1.22	AAU	Asn	196	1.41	AGU	Ser	36	0.89	U
	AUC	Ile	57	0.43	ACC	Thr	43	0.76	AAC	Asn	83	0.59	AGC	Ser	36	0.89	C
	AUA	Met	178	1.66	ACA	Thr	105	1.85	AAA	Lys	228	1.55	AGA	Ser	36	0.89	A
	AUG	Met	36	0.34	ACG	Thr	10	0.18	AAG	Lys	67	0.45	AGG	Ser	22	0.54	G
G	GUU	Val	54	1.43	GCU	Ala	52	1.63	GAU	Asp	42	1.45	GGU	Gly	26	0.85	U
	GUC	Val	11	0.29	GCC	Ala	23	0.72	GAC	Asp	16	0.55	GGC	Gly	7	0.23	C
	GUA	Val	78	2.07	GCA	Ala	50	1.56	GAA	Glu	59	1.62	GGA	Gly	77	2.52	A
	GUG	Val	8	0.21	GCG	Ala	3	0.09	GAG	Glu	14	0.38	GGG	Gly	12	0.39	G

Additionally, the relative synonymous codon usage (RSCU) of NNU and NNA codons greater than 1 indicated that the third positions of the U/A have high frequency of codon usage in *B. zonata* mitogenome (Table 2). There was a strong bias toward A + T-rich codons with the five most prevalent codons in *B. zonata*, being in order, 2.75 (TTA-Leu), 1.66 (ATA-Met), 1.57 (ATT-Ile), 1.55 (AAA-Lys) and 1.41 (AAT-Asn) (Table 2).

Several previous 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). In the present study, the A + T content of zero-fold sites (P_{0FD}), two-fold (P_{2FD}) and four-fold degenerate sites (P_{4FD}) was determined for each of the PCGs from eleven tephritid species (*B. zonata*, *B. papaya*, *B. correcta*, *B. carambolae*, *B. dorsalis*, *B. philippinensis*, *B. tryoni*, *B. oleae*, *B. minax*, *B. cucurbitae* and *C. capitata*) (Fig. 2). We observed that, with the exception of *nad4*, all PCGs had significantly lower variation of A + T content among the eleven tephritid species at P_{0FD} compared to both P_{2FD} and P_{4FD} sites. The P_{0FD} of nad4 had a higher A + T content (76.62 ± 3.51) as compared to both P_{2FD} (77.10 ± 1.71) and P_{4FD} (41.69 ± 0.70). Here phylogenetic inference is known to be affected by base compositional heterogeneity and among-site rate variation (ASRV), leading to the identification of incorrect phylogenetic relationships.

The AT skew and GC skew (Perna and Kocher, 1995) calculated for all available complete mitogenome of Tephritidae species (Table 1) revealed that the AT skewness of the *B. zonata* mitogenomes was slightly positive (0.065), indicating the occurrence of more As than Ts. Similar results were found in other tephritid species, with the value ranging from 0.021 in *C. capitata* to 0.131 in *B. minax*. Although the GC skew values are negative in all analyzed Tephritidae mitogenomes (−0.185 to −0.361), it is worth mentioning that the GC skew of *B. zonata* exhibits in range skew value (−0.223), meaning more Cs than Gs, similar to the skewness values for other dipteran, lepidopteran and animal mitogenomes (Junqueira et al., 2004; Hou et al., 2007; Jiang et al., 2009; Chen et al., 2014). When considering the 13 PCGs, the bias toward the use of As over Ts is more obvious in the *B. zonata* mitogenome, in which the AT skewness is 0.070. In *B. zonata* tRNAs, the AT skewness is slightly negative (−0.017), and GC skewness is positive (0.109). It differs from those found in previously sequenced Tephritidae species, with the value

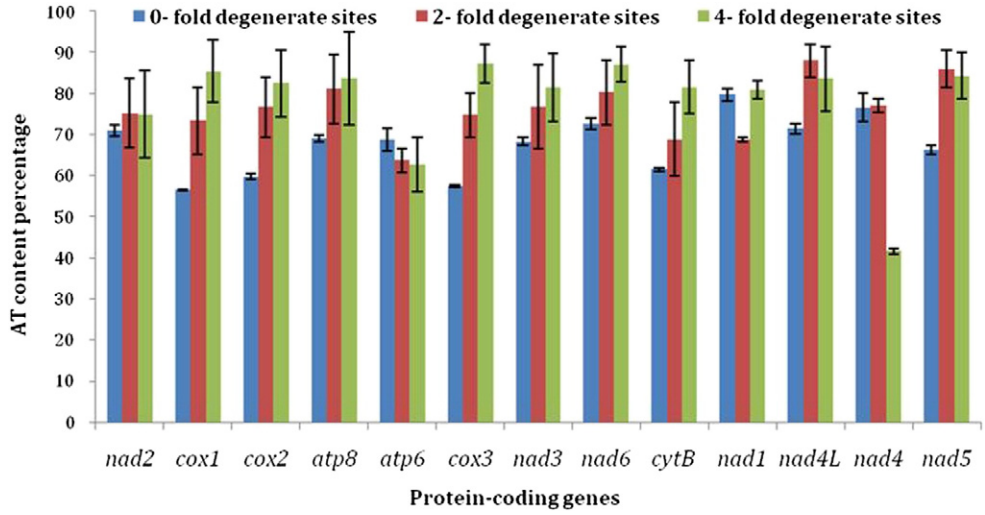

Fig. 2. The AT content percentage of 0-fold degenerate sites, 2-fold degenerate sites and 4-fold degenerate sites in each protein coding gene of mitochondrial genomes of 11 tephritid species, *B. zonata*, *B. papayae*, *B. correcta*, *B. carambolae*, *B. dorsalis*, *B. philippinensis*, *B. tryoni*, *B. oleae*, *B. minax*, *B. cucurbitae* and *C. capitata*. The black line with short line on the top of each bar represents the standard deviation value (SD).

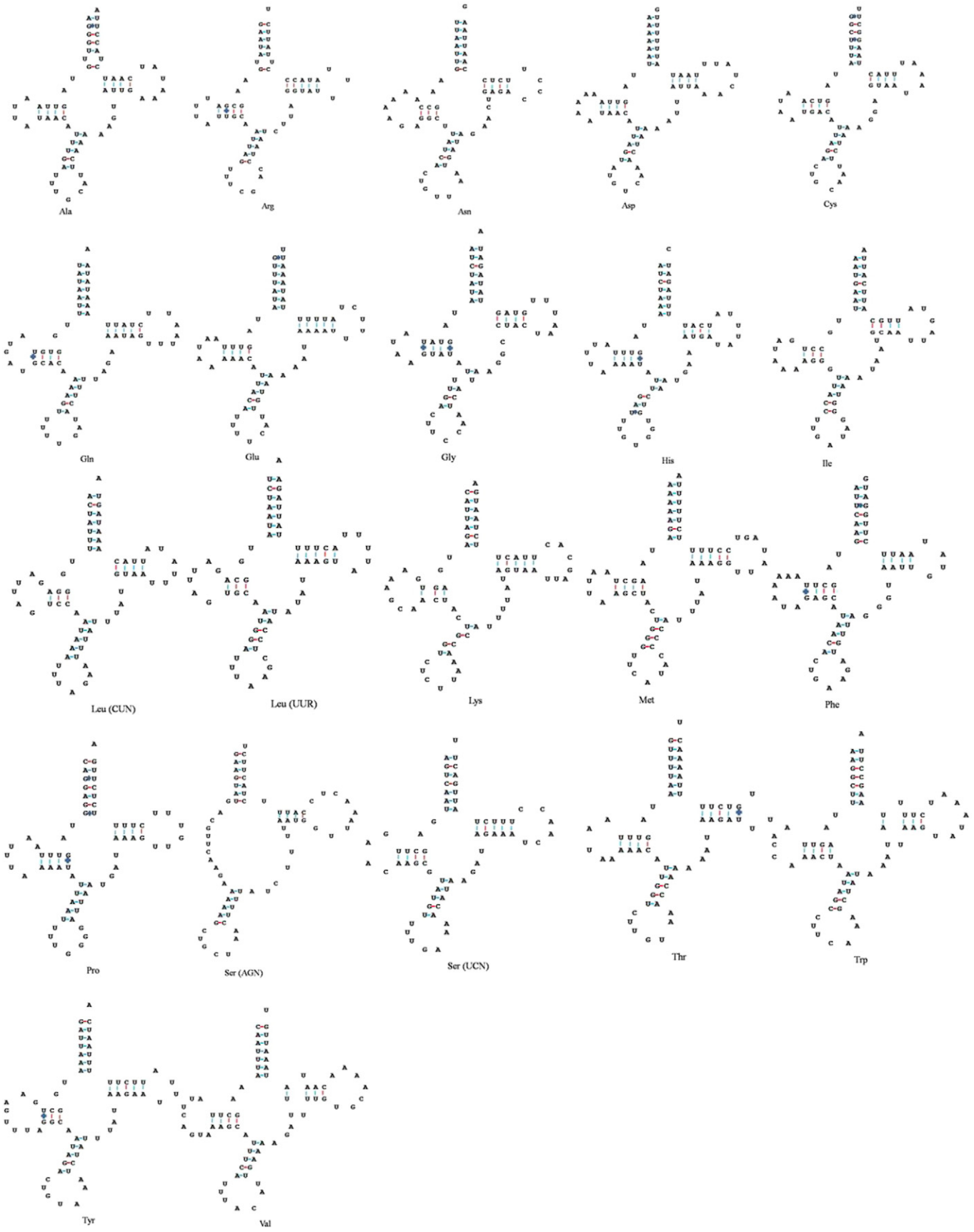

Fig. 3. Predicated secondary clover-leaf structures for the 22 tRNA genes of *B. zonata*. The tRNAs are labled with abbreviation of their corresponding amino acids below each tRNA gene structure. Arms of tRNAs (clockwise from top) are the amino acid acceptor arm, TYC arm, the anticodon arm, and dihydrouridine (DHU) arm.

ranging from 0.005 in *B. cucurbitae* to 0.055 in *B. minax* and −0.074 in *B. cucurbitae* to −0.182 in *B. minax* for AT and GC skewness, respectively.

3.4. Protein-coding genes

A total of 13 protein-coding genes encode in the mitochondrial genome of *B. zonata* (Supp. Table 1). With the exception of *cox1* and *atp8*, all protein coding genes start with a typical ATN initiation codon, i.e., ATG used in *cox2*, *atp6*, *cox3*, *nad4*, *nad4L* and *cytb*, ATT in *nad2*, *nad3*, *nad5* and *nad6* and ATA in *nad1*. ATG or ATA initiation codons, encoding for Methionine, are the most typical among insects (Jiang et al., 2009; Chen et al., 2014); ATT codon is rare but often observed among tephritids (Yu et al., 2007; Zhang et al., 2014). Genes for *cox1* and *atp8* used TCG and GTG as initiation codons, respectively. Some unusual codons for *cox1* have been proposed in insect mtDNAs. The tetranucleotide, TTAG, has been designated as a *cox1* start codon in *Bombyx mori* (Yukuhiro et al., 2002), *Bombyx mandarina* (Yukuhiro et al., 2002; Pan et al., 2008), *Antheraea pernyi* (Liu et al., 2008), *Antheraea yamamai* (Kim et al., 2009) and *Coreana raphaelis* (Kim et al., 2006); whereas in case of *Ostrina* the hexanucleotide ATTTAG has been proposed (Coates et al., 2005). The initiation codon for *cox1* was TCG(s) in *B. zonata*, which was observed in other tephritid species (Nardi et al., 2003; Yu et al., 2007; Wu et al., 2012; Zhang et al., 2014). GTG being the initiation codon for *atp8* was only observed in *B. dorsalis* (Yu et al., 2007) but not in other tephritids species.

Standard TAA and TAG termination codons are found in seven (*nad2*, *cox2*, *atp8*, *atp6*, *cox3*, *nad4L* and *nad6*) and two (*nad4* and *cytb*) PCGs, respectively. The remaining genes show an incomplete termination codon (TA in *cox1*; T in *nad3*, *nad5* and *nad1*) (Supp. Table 1), which may likely extend to TAA during the maturation of transcript, a phenomenon commonly observed in metazoan mitochondrial genes (Clary and Wolstenholme, 1985; Yu et al., 2007).

3.5. Ribosomal RNA genes, transfer RNA genes and tRNA like structure

Like other insect mitogenomes, two ribosomal RNA genes (*rrnL* and *rrnS*) are also present in the mitogenome of *B. zonata*, which are located between $trnL^{(CUN)}$ and *trnV*, and between *trnV* and the CR region, respectively. The length of *B. zonata rrnS* and *rrnL* was 790 bp and 1333 bp, respectively, well within the range observed in the completely sequenced other tephritid species (Yu et al., 2007; Zhang et al., 2014).

The typical set of 22 tRNA genes was identified in the mitogenome of *B. zonata*, and their predicted secondary clover-leaf structures (Fig. 3), these tRNA genes are scattered throughout the mitogenome, and vary from 64 bp for *trnC* and *trnR* to 72 bp for *trnV*. All the predicted tRNAs display a typical clover-leaf secondary structure, except for $trnS^{(AGN)}$, where the DHU arm appears to be replaced by eight unpaired nucleotides, as observed in *B. dorsalis* and *B. minax* mitogenomes and a feature typical of other animal mitochondria (Wolstenholme, 1992). The tertiary structure unusual nematode $trnS^{(AGN)}$ can fit the ribosome by altering its structural conformation and function in a similar way to that of usual tRNAs' works in the ribosome proposed by nuclear magnetic resonance analysis (Ohtsuki et al., 2002).

The anticodons are identical to those observed in other tephritid species. The length of the anticodon loop, the acceptor stem and the anticodon stem of all *B. zonata* is 7 bp, 7 bp and 5 bp except in trnM and trnK, respectively. Size variability was observed mostly in DHU arms (loop size 4–8 bp, stem size 3–4 bp) and the TΨC arms (loop size 3–9 bp, stem size 4–5 bp). In addition, a total of 20 mismatched base pairs and G–U wobble pairs were observed in 22 tRNA genes. Of these, there were 17 G–U wobble pairs that form a weak bond and 3 unmatched U–U base pairs were observed in trnG (anticodon arm), trn and trnV (TΨC arm).

3.6. Intergenic spacers

In *B. zonata*, the two longest intergenic spacers were of 78 bp and 32 bp observed between *trnI* and *trnQ* and between *trnC* and *trnY*, respectively (Supp. Table 1). The longer intergenic spacer (78 bp) shows no significant similarity to the other regions in the *B. zonata* mitogenomes. On the other hand, the smaller spacer sequence (32 bp) has a clear counterpart in the control region (CR), where the first 21 bases out of 32 bases of the spacer are found exactly repeated. Specifically, the first 21/32 bp of the spacers from 1440–1460 bases correspond to bases 15,342–15,362 in the CR in inverted orientation, while the remaining 11 bp of the spacer (bases 1461–1471) have no resemblance with the collinear bases 15,331–15,341 in the CR. Zhang et al. (2014) reported that the 42 bp intergenic spacer located between *trnC* and *trnY* in *B. minax* had a clear counterpart in the CR with the first 15 of 42 bp matching. Similarly, in *B. dorsalis* 45 bp intergenic spacer located between *trnC* and *trnY* had a clear counterpart in the CR bases with the first 33 of 45 bp reported by Yu et al. (2007). Interestingly, the 32 bp intergenic spacer located between *trnC* and *trnY* was predicted to form one stem-loop secondary structure with 5 bp loop with 7 bp stem ($\Delta G = -3.70$). The counterpart in the CR bases 15,331–15,362 are predicted to form a long stem structure together with a partially complementary neighboring sequence ($\Delta G = -8.27$) (Fig. 4). The secondary structures of intergenic spacer located between *trnC* and *trnY* and their counterpart in CR were also observed in *B. dorsalis* and *B. minax* (Yu et al., 2007; Zhang et al., 2014). Therefore, the results obtained from the present and previous studies support the hypothesis that the secondary structures formed by the intergenic spacer and their counterparts in CR might play a major role as hotspot in recombination (Stanton et al., 1994; Yu et al., 2007; Zhang et al., 2014).

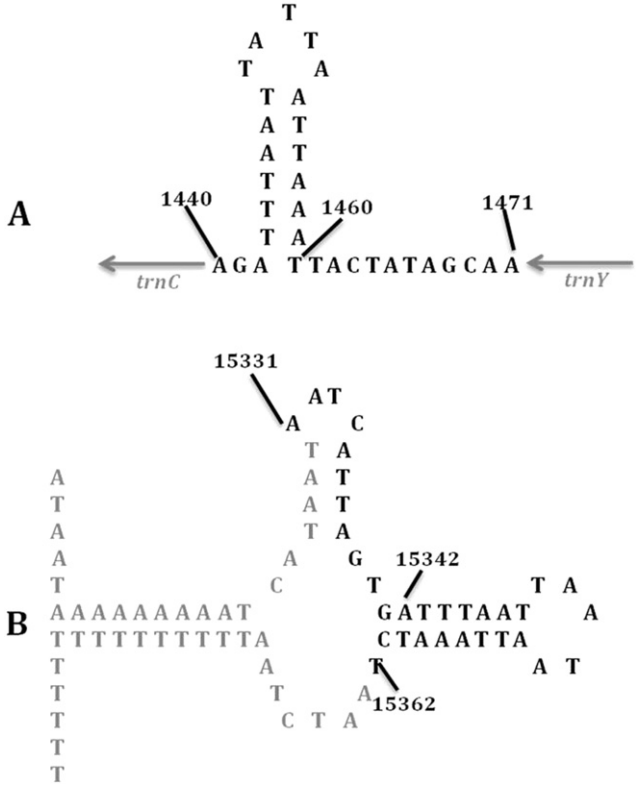

Fig. 4. Hypothetical secondary structures in the regions surrounding the repeated sequence (in bold) between trnC and trnY (A) and in the CR (B).

The four intergenic spacers in *B. zonata*, IGS-1 (18 bp between *trnE* and *trnF*), IGS-2 (15 bp between *nad5* and *trnH*), IGS-3 (15 bp between $trnS^{(UCN)}$ and *nad1*) and IGS-4 (10 bp between *nad1* and $trnL^{(CUN)}$), were observed to be conserved in size across other tephritids, *B. papayae*, *B. correcta*, *B. carambolae*, *B. dorsalis*, *B. philippinensis*, *B. tryoni* (18 bp, 15 bp, 15 bp and 10 bp), *B. oleae*, *B. minax* (18 bp, 14 bp, 16 bp and

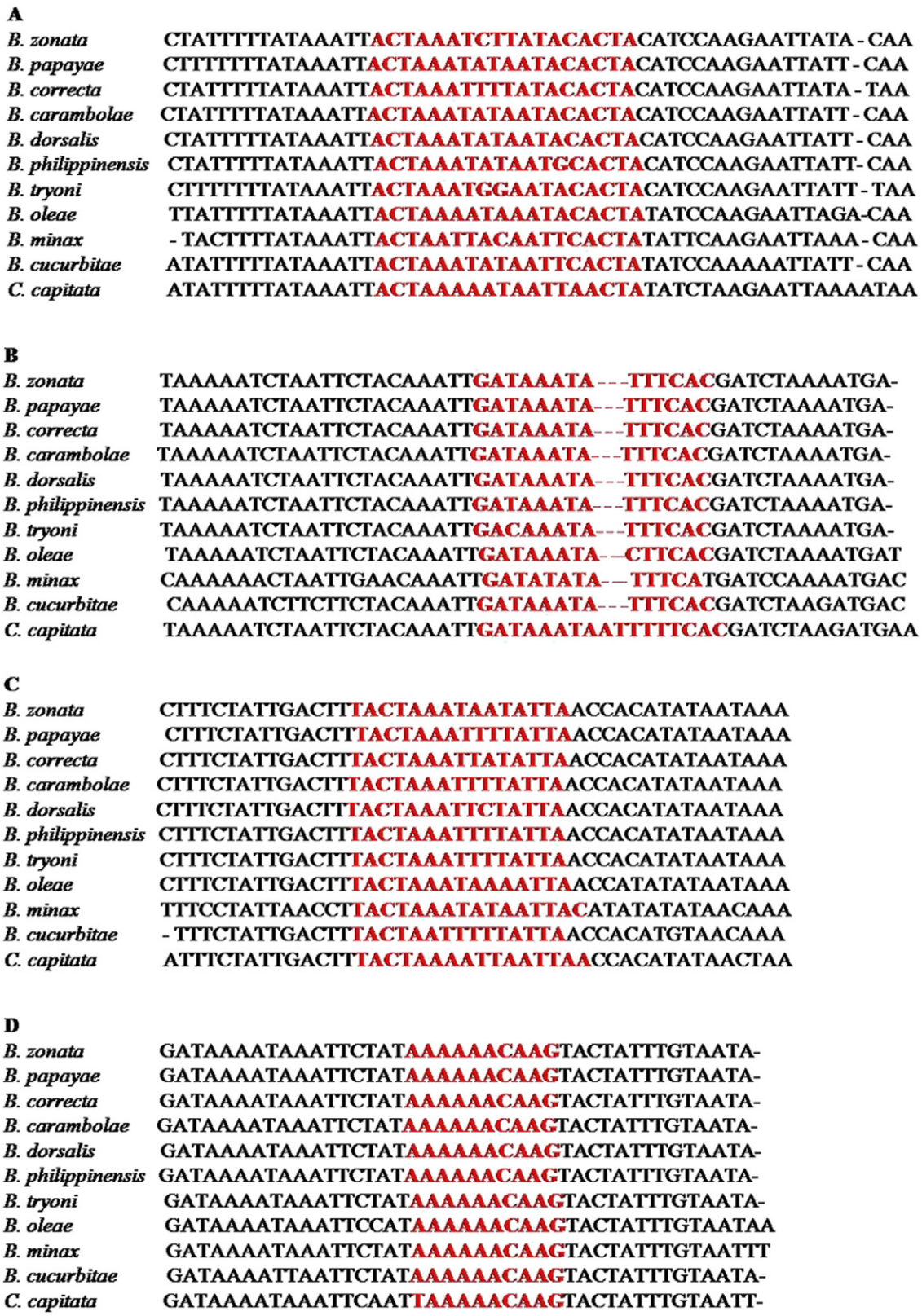

Fig. 5. Alignment of the intergenic spacer regions (in red) between $tRNA^{Gln}$ and $tRNA^{Phe}$ (A), between ND5 and $tRNA^{His}$ (B), between $tRNA^{Ser(UCN)}$ and ND1 (C), and between ND1 and $tRNA^{Leu(CUN)}$ (D) in 11 tephritid species, *B. zonata*, *B. papayae*, *B. correcta*, *B. carambolae*, *B. dorsalis*, *B. philippinensis*, *B. tryoni*, *B. oleae*, *B. minax*, *B. cucurbitae* and *C. capitata*.

10 bp), *B. cucurbitae* and *C. capitata* (18 bp, 18 bp, 16 bp and 10 bp) at the same locations. All intergenic spacers were found at the same locations and have highly significant similarity in percentage identity (71.4–100%) (Fig. 5). These intergenic spacers have clear counterpart in the CR of their respective species but cannot be able to form secondary structure studied by Zhang et al. (2014). A previous study has been done to find out original longer intergenic spacer sequences and structures in the CR (Yu et al., 2007), but the longer intergenic spacers had significantly different features, including sequence, length and location even among the tephritid species. Some earlier studies focused on the short intergenic spacers described the no gene rearrangement in tephritid species supported by nonforming stem-loop structures of small intergenic spacers located at 3′ end of PCGs (Zhang et al., 2014).

3.7. Control region (CR)

The control regions of tephritid mitogenomes are located at the conserved position between *rrnS* and *trnI-trnQ-trnM* gene cluster (Fig. 1). The *B. zonata* CR spans 950 bp, with high A + T content (84.42%) in the mitogenomes, which is well within range as observed in the completely sequenced tephritid species with size ranging from 946 bp in *B. cucurbitae* to 1141 bp in *B. minax* with A + T content varying from 77.65% in *B. minax* to 91.14% in *C. capitata*. The mitochondrial CR of both vertebrates and invertebrates has a high A + T content and the initiation of replication feature (Boore, 1999).

In the *B. zonata* CR region, a prominent poly-T stretch has been observed at the 5′ end followed by a $[TA(A)]_n$ similar common feature observed in tephritid mitogenomes (Zhang et al., 2014) and in other dipteran and orthopteran (Zhang and Hewitt, 1997). This poly-T stretch has been found conserved within dipteran, hymenopteran and lepidopteran mitogenomes (Zhang et al., 2014; Cha et al., 2007; Jiang et al., 2009) and is involved in controlling transcription and/or replication initiation or may have some other unknown functions (Zhang et al., 1995; Zhang and Hewitt, 1997). In addition, the CR region of *B. zonata* does not contain any tRNA-like sequence, but contains tandem repeats of 39 bp size. The sequence TTTATAAATACTTAAATTTA occurred twice located at positions from to 15,401 to 15,439. Other than *B. zonata*, variable number tandem repeats in the CR region has been also observed in tephritid species (Zhang et al., 2014).

3.8. Gene evolutionary rate

In order to analyze the gene evolutionary rate of Tephritidae mitogenomes, the rate of non-synonymous substitutions (Ka, pi modified) and synonymous substitutions (Ks, pi modified) as well as the Ka/Ks ratio were calculated for all the PCGs (Table 3). The *atp6* showed the highest evolutionary rates, while *cox1* with the lowest rate, which coincides with the fact that *coxI* can be used as a common barcoding marker in tephritids (Wu et al., 2012; Jiang et al., 2009). Analogously, *cox3* and *cytb* with relatively slow rates may also be candidate for barcoding markers. Jukes–Cantor adjusted Ka/Ks (JKa/JKs) were also calculated. Notably, the ratios of Ka/Ks and JKa/JKs for all PCGs except *atp6* were below 0.5, indicating that these genes were evolving under the purifying selection indicating that all of them except *atp6* could be combined to analyze the phylogeny of tephritids.

3.9. Phylogenetic relationships

The use of protein-coding gene sequences of the mitogenomes has become an informative strategy for inferring phylogenetic relationships (Boore et al., 2005). Although the phylogenetic relationships of *Bactrocera* species has been inferred up to great extent using partial mitochondrial gene sequences (*cox1*, *nad1*, *16S rRNA* genes), whereas present analysis is based on Maximum Likelihood method, a model-based evolutionary method using sequence dataset containing nucleotide sequences of 13 PCGs plus 2 rRNAs (Fig. 6). The optimal cladograms inferred by this dataset indicate the presence of four distinct families viz. Tephritidae, Muscidae, Drosophilidae and Culicidae in Order Diptera. Among Tephritidae, *B. carambolae*, *B. papayae*, *B. philippinensis* and *B. dorsalis* belong to *B. dorsalis* species complex which forms a monophyletic clade. It is clear that *B. zonata* is closely related to *B. correcta* in lineage clade (Muraji and Nakahara, 2001; Nakahara and Muraji, 2008), with a very strong support by ML analysis. We also inferred phylogenetic relationship with 12 PCGs (except *atp6*) plus 2 rRNAs, showing the similar relationship as containing whole set of PCGs. Testing was done without *atp6* due to high rate of evolution in study (>0.5). Even the recent molecular evidence also suggests that *B. papayae*, *B. philippinensis*, *B. invadens* and *B. dorsalis* represent one species (Krosch et al., 2013; Schutze et al., 2012, 2015; Jose et al., 2013; Boykin et al., 2014) which has also been confirmed by the phylogeny based on PCGs and rRNAs of mitogenomes in the present study.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gene.2015.05.066.

Acknowledgments

This work was supported by the Ministry of Agriculture, Government of India through the National Initiative on Climate Resilient Agriculture (NICRA) project under the Indian Council of Agricultural Research (ICAR) (ICAR-RCER/RC R/E.F./2011/29). We are grateful to Dr. B.P. Bhatt (Director of institute) and Dr. A.K. Singh (Head of centre) for giving valuable suggestions and providing laboratory facilities. We

Table 3
Rates of non-synonymous substitutions (Ka, pi modified), synonymous substitutions (Ks, pi modified) and the Ka/Ks ratio as well as Jukes–Cantor adjusted Ka/Ks (JKa/JKs) ratio in each PCG of mitogenome of 11 tephritid species, *B. zonata*, *B. papayae*, *B. correcta*, *B. carambolae*, *B. dorsalis*, *B. philippinensis*, *B. tryoni*, *B. oleae*, *B. minax*, *B. cucurbitae* and *C. capitata*.

Protein-coding genes	Rates of non-synonymous substitutions (Ka)	Rates of synonymous substitutions (Ks)	Ka/Ks ratio	Rates of non-synonymous substitutions Jukes–Cantor adjusted J (Ka)	Rates of synonymous substitutions Jukes–Cantor adjusted J (Ks)	JKa/JKs ratio
nad2	0.05850	0.53587	0.10917	0.05980	0.59910	0.09982
cox1	0.01046	0.45762	0.02286	0.01170	0.89380	0.01309
cox2	0.02199	0.42945	0.05121	0.02255	0.79971	0.02820
atp8	0.07257	0.39027	0.18595	0.07812	0.42540	0.18364
atp6	0.11775	0.15588	0.75539	0.13087	0.17972	0.72819
cox3	0.00965	0.41640	0.02317	0.00976	0.69202	0.01410
nad3	0.05227	0.51680	0.10114	0.05670	0.58910	0.09625
nad6	0.08708	0.43212	0.20152	0.09473	0.76322	0.12412
cytb	0.01832	0.50719	0.03612	0.01902	0.59140	0.03216
nad1	0.02490	0.29687	0.08388	0.02554	0.39961	0.06391
nad4L	0.02573	0.32631	0.07885	0.02675	0.46818	0.05714
nad4	0.05233	0.37953	0.13788	0.05663	0.57966	0.09770
nad5	0.04417	0.40062	0.11025	0.04628	0.63263	0.07315

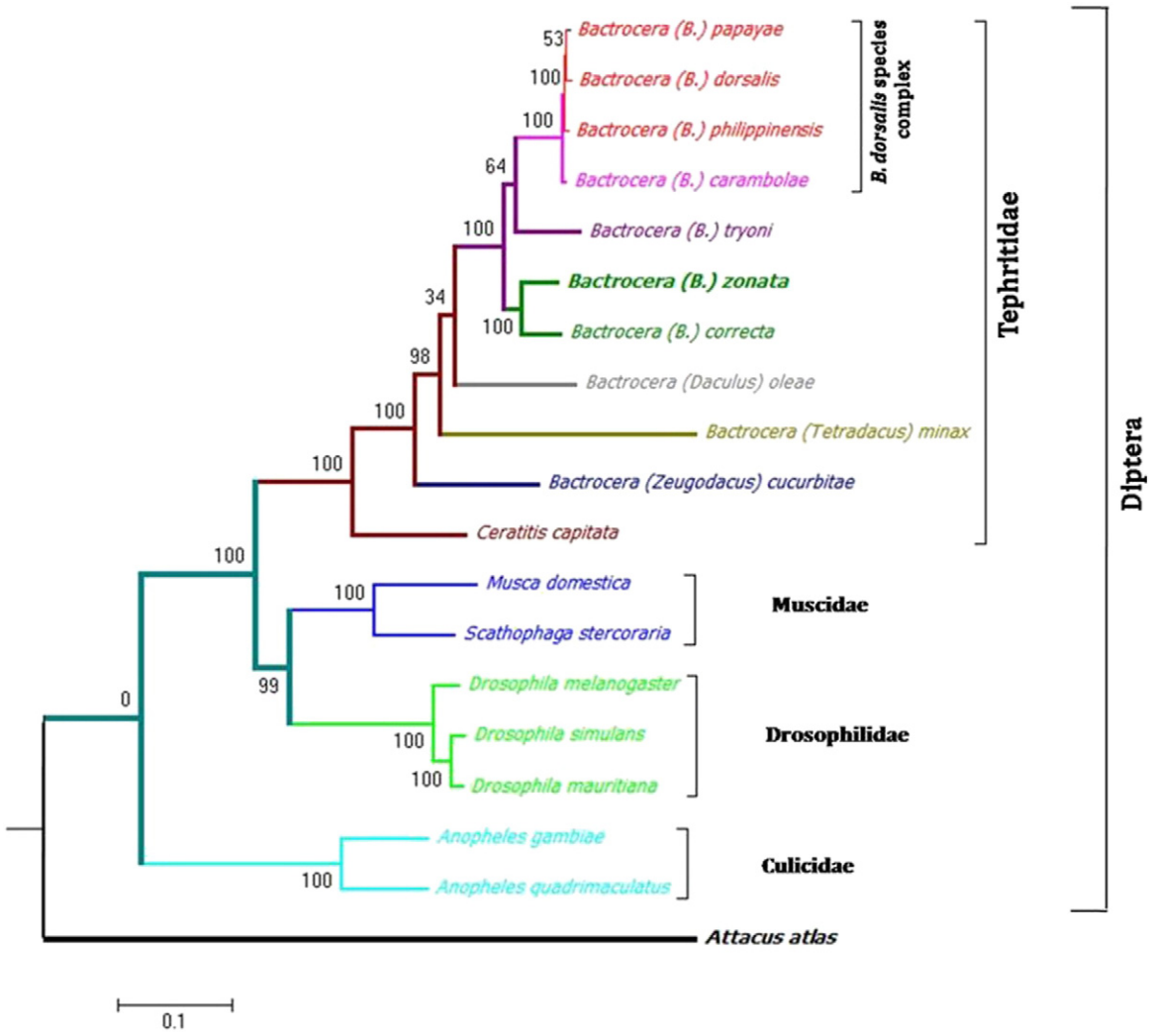

Fig. 6. Maximum Likelihood estimation of the phylogenetic relationship among species from the insect order Diptera. Tree based on 13 complete protein coding gene and 2 rRNAs sequence from complete mitochondrial genome sequences. Clustering based upon close relation and families are indicated by bars along the right hand side of the phylogenetic tree. *Attacus atlas* (Insecta: Lepidoptera) was used out group to root the tree.

are also thankful to the four anonymous reviewers for their constructive comments on an earlier version of the manuscript.

References

Abd-El-Samie, E.M., El Fiky, Z.A., 2011. Molecular phylogeny and identification of the peach fruit fly, *Bactrocera zonata*, established in Egypt. J. Insect Sci. 11 (177), 1–11.

Asokan, R., Rebijith, K.B., Singh, S.K., Sidhu, A.S., Siddharthan, S., Karanth, P.K., Ellango, R., Ramamurthy, V.V., 2011. Molecular identification and phylogeny of *Bactrocera* species (Diptera: Tephritidae). Fla. Entomol. 94 (4), 1026–1035.

Avise, J.C., 2000. Phylogeography, the History and Formation of Species. Harvard University Press, Cambridge, Massachusetts, USA.

Bernt, M., Donath, A., Juhling, F., Externbrink, F., Florentz, C., Fritzsch, G., Putz, J., Middendorf, M., Stadler, P.F., 2012. MITOS: improved de novo metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. 69, 313–319.

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

Boore, J.L., Macey, J.R., Medina, M., 2005. Sequencing and comparing whole mitochondrial genomes of animals. Methods Enzymol. 395, 311–348.

Boykin, L.M., Schutze, M.K., Krosch, M.N., Chomicˇ, A., Chapman, T.A., Englezou, A., Armstrong, K.F., Clarke, A.R., Hailstones, D., Cameron, S.L., 2014. Multigene phylogenetic analysis of south-east Asian pest members of the *Bactrocera dorsalis* species complex (Diptera: Tephritidae) does not support current taxonomy. J. Appl. Entomol. 138 (4), 235–253.

Cameron, S.L., Lambkin, C., Barker, S.C., Whiting, M.F., 2007. A mitochondrial genome phylogeny of Diptera: whole genome sequence data accurately resolve relationships over broad timescales with high precision. Syst. Entomol. 32, 40–59.

Cha, S.Y., Yoon, H.J., Lee, E.M., Yoon, M.H., Hwang, J.S., Jin, B.R., Han, Y.S., Kim, I., 2007. The complete nucleotide sequence and gene organization of the mitochondrial genome of the bumblebee, *Bombus ignites* (Hymenoptera: Apidae). Gene 392, 206–220.

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

Chilana, P., Sharma, A., Rai, A., 2012. Insect genomic resources: status, availability and future. Curr. Sci. 102, 571–580.

Clary, D.O., Wolstenholme, D.R., 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. J. Mol. Evol. 22, 252–271.

Coates, B.S., 2014. Assembly and annotation of full mitochondrial genomes for the corn rootworm species, *Diabrotica virgifera virgifera* and *Diabrotica barberi* (Insecta: Coleoptera: Chrysomelidae), using next generation sequence data. Gene 542, 190–197.

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

Crease, T.J., 1999. The complete sequence of the mitochondrial genome of *Daphnia pulex* (Cladocera: Crustacea). Gene 233, 89–99.

Drew, R.A.I., Raghu, S., 2002. The fruit fly fauna (Diptera: Tephritidae: Dacinae) of the rainforest habitat of the Western Ghats, India. Raffles Bull. Zool. 50 (2), 327–352.

Duyck, P.F., Sterlin, J.F., Quilici, S., 2004. Survival and development of different life stages of Bactrocera zonata (Diptera: Tephritidae) reared at five constant temperatures compared to other fruit fly species. Bull. Entomol. Res. 94, 89–93.

Efflatoun, H.C., 1924. A monograph of Egyptian Diptera. Part II, family Trypaneidae. Mémoires de la société entomologique de l'Egypte. Fortin Masson & Ci, Paris.

El-Minshawy, A.M., El-Eryan, M.A., Awad, A.I., 1999. Biological and morphological studies on the guava fruit fly *Bactrocera zonata* (Saunders) (Diptera: Tephritidae) found recently in Egypt. 8th Nat. Conf. of Pests & Dis. of Veg. & Fruits in Ismailia, Egypt, pp. 71–82.

Fenn, J.D., Cameron, S.L., Whiting, M.F., 2007. The complete mitochondrial genome sequence of the Mormon cricket (*Anabrus simplex*: Tettigoniidae: Orthoptera) and an analysis of control region variability. Insect Mol. Biol. 16, 239–252.

Friedrich, M., Muqim, N., 2003. Sequence and phylogenetic analysis of the complete mitochondrial genome of the flour beetle *Tribolium castanaeum*. Mol. Phylogenet. Evol. 26, 502–512.

Grant, J.R., Stothard, P., 2008. The CGView Server: a comparative genomics tool for circular genomes. Nucleic Acids Res. 36, W181–W184.

Hashem, A.G., Mohamed, S.M.A., EI-Wakkad, M.F., 2001. Diversity and abundance of Mediterranean and peach fruit flies (Diptera: Tephritidae) in different horticultural orchards. Egypt. J. Appl. Sci. 16, 303–314.

Hassanin, A., 2006. Phylogeny of Arthropoda inferred from mitochondrial sequences: strategies for limiting the misleading effects of multiple changes in pattern and rates of substitution. Mol. Phylogenet. Evol. 38, 100–116.

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

Hou, W.R., Chen, Y., Wu, X., Hu, J.C., Peng, Z.S., Yang, J., Tang, Z.X., Zhou, C.Q., Li, Y.M., Yang, S.K., Du, Y.J., Kong, L.L., Ren, Z.L., Zhang, H.Y., Shuai, S.R., 2007. A complete mitochondrial genome sequence of Asian black bear Sichuan subspecies (*Ursus thibetanus mupinensis*). Int. J. Biol. Sci. 3, 85–90.

Inohira, K., Hara, T., Matsuura, E.T., 1997. Nucleotide sequence divergence in the A + T-rich region of mitochondrial DNA in *Drosophila simulans* and *Drosophila mauritiana*. Mol. Biol. Evol. 14, 814–822.

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

Joomaye, A., Price, N.S., 2000. Pest risk analysis and quarantine of fruit flies. Proceedings of the Indian Ocean Commission Regional Fruit Fly Symposium, June 5–9, 2000, Mauritius, pp. 3–16.

Jose, M.S., Leblanc, L., Geib, S.M., Rubinoff, D., 2013. An evaluation of the species status of *Bactrocera invadens* and the systematics of the *Bactrocera dorsalis* (Diptera: Tephritidae) complex. Ann. Entomol. Soc. Am. 106, 684–694.

Junqueira, A.C., Lessinger, A.C., Torres, T.T., Da Silva, F.R., Vettore, A.L., Arruda, P., Azeredo, E.A.M., 2004. The mitochondrial genome of the blowfly *Chrysomya chloropyga* (Diptera: Calliphoridae). Gene 339, 7–15.

Kapoor, V.C., 1993. Indian Fruit Flies (Insecta: Diptera: Tephritidae). International Sciences Publisher, New York (NY), p. 228.

Kim, I., Lee, E.M., Seol, K.Y., Yun, E.Y., Lee, Y.B., Hwang, J.S., Jin, B.R., 2006. The mitochondrial genome of the Korean hairstreak, *Coreana raphaelis* (Lepidoptera: Lycaenidae). Insect Mol. Biol. 15, 217–225.

Kim, S.R., Kim, M.I., Hong, M.Y., Kim, Y.Y., Kang, P.D., Hwang, J.S., Han, Y.S., Jin, B.R., Kim, I., 2009. The complete mitogenomes sequence of the Japanese oak silkmoth, *Antheraea yamamai* (Lepidoptera: Saturniidae). Mol. Biol. Rep. 36 (7), 1871–1880. http://dx.doi.org/10.1007/s11033-008-9393-2 (Epub 2008 Nov 2).

Kirkness, E.F., Haas, B.J., Sun, W., Briag, H.R., Perotti, M.A., Clark, J.M., Lee, S.H., Robertson, H.M., Kennedy, R.C., Elhaik, E., Gerlach, D., Kriventseva, E.V., Elsik, C.G., Graur, D., Hill, C.A., Veenstra, J.A., Walenz, B., Tubío, J.M., Ribeiro, J.M., Rozas, J., Johnston, J.S., Reese, J.T., Popadic, A., Tojo, M., Raoult, D., Reed, D.L., Tomoyasu, Y., Kraus, E., Mittapalli, O., Margam, V.M., Li, H.M., Meyer, J.M., Johnson, R.M., Romero-Severson, J., Vanzee, J.P., Alvarez-Ponce, D., Vieira, F.G., Aguadé, M., Guirao-Rico, S., Anzola, J.M., Yoon, K.S., Strycharz, J.P., Unger, M.F., Christley, S., Lobo, N.F., Seufferheld, M.J., Wang, N., Dasch, G.A., Struchiner, C.J., Madey, G., Hannick, L.I., Bidwell, S., Joardar, V., Caler, E., Shao, R., Barker, S.C., Cameron, S., Bruggner, R.V., Regier, A., Johnson, J., Viswanathan, L., Utterback, T.R., Sutton, G.G., Lawson, D., Waterhouse, R.M., Venter, J.C., Strausberg, R.L., Berenbaum, M.R., Collins, F.H., Zdobnov, E.M., Pittendrigh, B.R., 2010. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. PNAS USA 207, 12168–12173.

Knaus, B.J., Cronn, R., Liston, A., Pilgrim, K., Schwartz, M.K., 2011. Mitochondrial genome sequences illuminate maternal lineages of conservation concern in a rare carnivore. BMC Ecol. 11, 10.

Krosch, M.N., Schutze, M.K., Armstrong, K.F., Boontop, Y., Boykin, L.M., Chapman, T.A., Englezou, A., Cameron, S.L., Clarke, A.R., 2013. Piecing together an integrative taxonomic puzzle: microsatellite, wing shape and aedeagus length analyses of *Bactrocera dorsalis s.s.* (Diptera: Tephritidae) find no evidence of multiple lineages in a proposed contact zone along the Thai/Malay Peninsula. Syst. Entomol. 38, 2–13.

Lessinger, A.C., Martins, Junqueira, A.C., Lemos, T.A., Kemper, E.L., da Silva, F.R., Vettore, A.L., Arruda, P., Azeredo-Espin, A.M., 2000. The mitochondrial genome of primary screwworm fly *Cochliomyia hominivorax* (Diptera: Calliphoridae). Insect Mol. Biol. 9, 521–529.

Lewis, J.A., Huq, A., Liu, W., Jacob, A., 1995. Induction of antiviral responses and gene activation by intracellular interferon. Virology 212, 438–450.

Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452.

Liu, Y.Q., Li, Y.P., Pan, M.H., Dai, F., Zhu, X., Lu, C., Xiang, Z., 2008. The complete mitochondrial genome of the Chinese oak silkmoth, *Antheraea pernyi* (Lepidoptera: Saturniidae). Acta Biochim. Biophys. Sin. 40, 694–703.

Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 955–964.

Ma, P.F., Guo, Z.H., Li, D.Z., 2012. Rapid sequencing of the bamboo mitochondrial genome using Illumina technology and parallel episodic evolution of organelle genomes in grasses. PLoS One 7, e30297.

Madhura, H.S., Verghese, A., 2004. A guide to identification of some common fruit flies (*Bactrocera* spp.) (Diptera: Tephritidae: Dacinae). Pest Manag. Hortic. Ecosyst. 10 (1), 87–96.

Mosleh, Y.Y., Moussa, S.F.M., Mohamed, L.H.Y., 2011. Comparative toxicity of certain pesticides to Peach Fruit Fly, *Bactrocera zonata* Saunders (Diptera: Tephritidae) under laboratory conditions. Plant Prot. Sci. 47, 5–120.

Muraji, M., Nakahara, S., 2001. Phylogenetic relationships among fruit flies, *Bactrocera* (Diptera, Tephritidae), based on the mitochondrial rDNA sequences. Insect Mol. Biol. 10, 549–559.

Nakahara, S., Muraji, M., 2008. Phylogenetic analyses of *Bactrocera* fruit flies (Diptera: Tephritidae) based on nucleotide sequences of the mitochondrial COI and COII Genes. Res. Bull. Plant Prot. Jpn. 44, 1–12.

Nardi, F., Carapelli, A., Dallai, R., Frati, F., 2003. The mitochondrial genome of the olive fly *Bactrocera oleae*: two haplotypes from distant geographical locations. Insect Mol. Biol. 12, 605–611.

Nardi, F., Carapelli, A., Boore, J.L., Roderick, G.K., Dallai, R., Frati, F., 2010. Domestication of olive fly through a multi-regional host shift to cultivated olives: comparative dating using complete mitochondrial genomes. Mol. Phylogenet. Evol. 57, 678–686.

OEPP/EPPO, 2005. Data Sheets on Quarantine Pests OEPP/EPPO Bulletin. 35, 371–373.

Ohtsuki, T., Kawai, G., Watanabe, K., 2002. The minimal tRNA: unique structure of *Ascaris suum* mitochondrial tRNAser-UCU having a short T arm and lacking the entire D arm. FEBS Lett. 514, 37–43.

Pan, M., Yu, Q., Xia, Y., Dai, F., Liu, Y., Lu, C., Zhang, Z., Xiang, Z., 2008. Characterization of mitochondrial genome of Chinese wild mulberry silkworm, *Bombyx mandarina* (Lepidoptera: Bombycidae). Sci. China Ser. C Life Sci. 51, 693–701.

Perna, N.T., Kocher, T.D., 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J. Mol. Evol. 41, 353–358.

Prabhakar, C.S., Sood, P., Kapoor, V., Kanwar, S.S., Mehta, P.K., Sharma, P.N., 2009. Molecular and biochemical characterization of three bacterial symbionts of fruit fly, *Bactrocera tau* (Tephritidae: Diptera). J. Gen. Appl. Microbiol. 55, 213–220.

Prabhakar, C.S., Sood, P., Mehta, P.K., 2012. Pictorial keys for predominant *Bactrocera* and *Dacus* fruit flies (Diptera: Tephritidae) of north western Himalaya. Arthropods 1 (3), 101–111.

Ribeiro, F.J., Przybylski, D., Yin, S., Sharpe, T., Gnerre, S., Abouelleil, A., Berlin, A.M., Montmayeur, A., Shea, T.P., Walker, B.J., Young, S.Y., Russ, C., Nusbaum, C., MacCallum, I., Jaffe, D.B., 2012. Finishing bacterial genomes from shotgun sequence data. Genome Res. 22, 2270–2277.

Schutze, M.K., Krosch, M.N., Armstrong, K.F., Chapman, T.A., Englezou, A., Chomic, A., Cameron, S.L., Hailstones, D., Clarke, A.R., 2012. Population structure of *Bactrocera dorsalis s.s.*, *B. papayae* and *B. philippinensis* (Diptera: Tephritidae) in southeast Asia: evidence for a single species hypothesis using mitochondrial DNA and wing-shape data. BMC Evol. Biol. 12, 130 (http://www.biomedcentral.com/1471-2148/12/130).

Schutze, M.K., Aketarawong, N., Amornsak, W., Armstrong, K.F., Augustinos, A.A., Barr, N.B., Bo, W., Bourtzis, K., Boykin, L.M., Cáceres, C.E., Cameron, S.L., Chapman, T.A., Chinvinijkul, S., Chomic, A., De Meyer, M., Drosopoulou, E., Englezou, A., Ekesi, S., Gariou-Papalexiou, A., Geib, S.M., Hailstones, D.L., Hasanuzzaman, M., Haymer, D.S., Hee, A.K.W., Hendrichs, J., Jessup, A.J., Ji, Q., Khamis, F.M., Krosch, M.N., Leblanc, L., Mahmood, K., Malacrida, A.R., Mavragani-Tsipidou, P., Mwatawala, M.W., Nishida, R., Ono, H., Reyes, J., Rubinoff, D.Z., San Jose, M., Shelly, T.E., Srikachar, S., Tan, K., Thanaphum, S., Haq, I.U., Vijaysegaran, S., Wee, S., Yesmin, F., Zacharopoulou, A., Clarke, A.R., 2015. Synonymization of key pest species within the *Bactrocera dorsalis* species complex (Diptera: Tephritidae): taxonomic changes based on a review of 20 years of integrative morphological, molecular, cytogenetic, behavioural and chemoecological data. Syst. Entomol. 40, 456–471.

Shao, R., Campbell, N.J., Barker, S.C., 2001. Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, *Heterodoxus macropus* (Phthiraptera). Mol. Biol. Evol. 18, 858–865.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., Flook, P., 1994. Evolution, weighing, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann. Entomol. Soc. Am. 87, 651–701.

Song, N., Liang, A.P., 2009. The complete mitochondrial genome sequence of *Geisha distinctissima* (Hemiptera: Flatidae) and comparison with other Hemipteran insects. Acta Biochim. Biophys. Sin. 41, 206–216.

Spaugy, L., 1988. Fruit flies. Two more eradication projects over. Citrograph 73, 168.

Staats, M., Erkens, R.H.J., van de Vossenberg, B., Wieringa, J.J., Kraaijeveld, K., Stielow, B., Geml, J., Richardson, J.E., Bakker, F.T., 2013. Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. PLoS ONE 8 (7), e69189. http://dx.doi.org/10.1371/journal.pone.0069189.

Stanton, D.J., Daehler, L.L., Moritz, C.C., Brown, W.M., 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.

Stewart, J.B., Beckenbach, A.T., 2003. Phylogenetic and genomic analysis of the complete mitochondrial DNA sequence of the spotted asparagus beetle *Crioceris duodecimpunctata*. Mol. Phylogenet. Evol. 26, 513–526.

Stewart, J.B., Beckenbach, A.T., 2005. Insect mitochondrial genomics: the complete mitochondrial genome sequence of the meadow spittlebug *Philaenus spumarius* (Hemiptera: Auchenorrhyncha: Cercopoidae). Genome 48, 46–54.

Stewart, J.B., Beckenbach, A.T., 2006. Insect mitochondrial genomics 2: the complete mitochondrial genome sequence of a giant stonefly, *Pteronarcys princeps*, asymmetric directional mutation bias, and conserved plecopteran A + T-region elements. Genome 49, 815–824.

Taanman, J.W., 1999. The mitochondrial genome: structure, transcription, translation and replication. Biochim. Biophys. Acta Bioenerg. 1410, 103–123.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729.

Thompson, F.C., 1998. Fruit Fly Expert Identification System and Systematic Information Database. The International Journal of the North American Dipterists' Society (524 pp.).

White, I.M., Elson-Harris, M., 1992. Fruit Flies of Economic Significance: Their Identification and Bionomics. International Institute of Entomology. CAB International, Wallingford, Oxon, UK.

White, I.M., Elson-Harris, M.M., 1994. Fruit Flies of Economic Significant; Their Identification and Bionomics. CAB International and ACIAR, Wallingford.
Wolstenholme, D.R., 1992. Animal mitochondrial DNA: structure and evolution. Int. Rev. Cytol. 141, 173–216.
Wu, P., Roques, A., Xiong, Z., Xu, L., Huang, Y.Y., Pan, Y.Z., 2012. The complete mitochondrial genome of the melon fly *Bactrocera cucurbitae* (Diptera: Tephritidae). Mitochondrial DNA 1–2. http://dx.doi.org/10.3109/19401736.2012.710224.
Yang, L., Wei, Z.J., Hong, G.Y., Jiang, S.T., Wen, L.P., 2009. The complete nucleotide sequence of mitochondrial genome of *Phthonandria atrilineata* (Lepidoptera: Geometridae). Mol. Biol. Rep. 36, 1441–1449.
Yu, D.J., Xu, L., Nardi, F., Li, J.G., Zhang, R.J., 2007. The complete nucleotide sequence of the mitochondrial genome of the oriental fruit fly, *Bactrocera dorsalis* (Diptera: Tephritidae). Gene 396, 66–74.
Yukuhiro, K., Sezutsu, H., Itoh, M., Shimizu, K., Banno, Y., 2002. Significant levels of sequence divergence and gene rearrangements have occurred between the mitochondrial genomes of the wild mulberry silkmoth, *Bombyx mandarina*, and its close relative, the domesticated silkmoth, *Bombyx mori*. Mol. Biol. Evol. 19, 1385–1389.
Zhang, D.X., Hewitt, G.M., 1997. Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. Biochem. Syst. Ecol. 25, 99–120.
Zhang, D.X., Szymura, J.M., Hewitt, G.M., 1995. Evolution and structural conservation of the control region of insect mitochondrial DNA. J. Mol. Evol. 40, 382–391.
Zhang, B., Liu, Y.H., Wu, W.X., Wang, Z.L., 2010. Molecular phylogeny of *Bactrocera* species (Diptera: Tephritidae: Dacini) inferred from mitochondrial sequences of 16S rDNA and *COI* sequences. Fla. Entomol. 93 (3), 369–377.
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. http://dx.doi.org/10.1371/journal.pone.0100558.
Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31 (13), 3406–3415. http://dx.doi.org/10.1093/nar/gkg595.