Molecular Biology Reports (2021) 48:6047–6056
https://doi.org/10.1007/s11033-021-06608-2

ORIGINAL ARTICLE

Complete mitochondrial genome of *Dacus vijaysegarani* and phylogenetic relationships with congeners and other tephritid fruit flies (Insecta: Diptera)

Hoi-Sen Yong[1] · Kah-Ooi Chua[1,2] · Sze-Looi Song[3,4] · Yvonne Jing-Mei Liew[1,5] · Praphathip Eamsobhana[6] · Kok-Gan Chan[1,7]

Received: 4 March 2021 / Accepted: 29 July 2021 / Published online: 6 August 2021

Abstract

Background Tephritid fruit flies of the genus *Dacus* are members of the tribe Dacini, subfamily Dacinae. There are some 274 species worldwide, distributed in Africa and the Asia-Pacific. To date, only five complete mitochondrial genomes (mitogenomes) of *Dacus* fruit flies have been published and are available in the GenBank.

Methods and results In view of the lack of study on their mitogenome, we sequenced (by next generation sequencing) and annotated the complete mitogenome of *D. vijaysegarani* from Malaysia to determine its features and phylogenetic relationship. The whole mitogenome of *D. vijaysegarani* has identical gene order with the published mitogenomes of the genus *Dacus*, with 13 protein-coding genes, two rRNA genes, 22 tRNAs, a non-coding A + T rich control region, and intergenic spacer and overlap sequences. Phylogenetic analysis based on 15 mitochondrial genes (13 PCGs and two rRNA genes), reveals *Dacus*, *Zeugodacus* and *Bactrocera* forming a distinct clade. The genus *Dacus* forms a monophyletic group in the subclade containing also the *Zeugodacus* group; this *Dacus-Zeugodacus* subclade is distinct from the *Bactrocera* subclade. *D.* (*Mellesis*) *vijaysegarani* forms a lineage with *D.* (*Mellesis*) *trimacula* in the subcluster containing also the lineage of *D.* (*Mellesis*) *conopsoides* and *D.* (*Callantra*) *longicornis*. *D.* (*Dacus*) *bivittatus* and *D.* (*Didacus*) *ciliatus* form a distinct subcluster. Based on *cox1* sequences, the Malaysia and Vietnam taxa of *D. vijaysegarani* may not be conspecific.

Conclusions Overall, the mitochondrial genome of *D. vijaysegarani* provided essential molecular data that could be useful for further studies for species diagnosis, evolution and phylogeny research of other tephritid fruit flies in the future.

Keywords Mitogenome · Tribe Dacini · Subfamily Dacinae · Tephritoidea · Systematics

Introduction

Tephritid fruit flies of the genus *Dacus* Fabricius are members of the tribe Dacini, subfamily Dacinae. There are some 274 species worldwide, distributed in Africa and the Asia-Pacific [1, 2]. The genus is divided into 10 subgenera: *Callantra*, *Dacus*, *Didacus*, *Leptoxyda*, *Lophodacus*, *Mellesis*, *Metidacus*, *Mictodacus*, *Neodacus*, and *Psilodacus* [3–5].

✉ Kah-Ooi Chua
ckahooi@gmail.com

✉ Sze-Looi Song
szelooi@um.edu.my

1 Faculty of Science, Institute of Biological Sciences, Universiti Malaya, Kuala Lumpur, Malaysia

2 Centre for Research in Biotechnology for Agriculture (CEBAR), Universiti Malaya, Kuala Lumpur, Malaysia

3 Institute for Advanced Studies, Universiti Malaya, Kuala Lumpur, Malaysia

4 Institute of Ocean and Earth Sciences, Universiti Malaya, Kuala Lumpur, Malaysia

5 Department of Medical Microbiology, Faculty of Medicine, Universiti Malaya, Kuala Lumpur, Malaysia

6 Department of Parasitology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand

7 International Genome Centre, Jiangsu University, Zhenjiang, Jiangsu, China

Dacus vijaysegarani Drew & Hancock is a member of the *siamensis* group of subgenus *Mellesis* [3]. It has been found in Malaysia, Thailand and Vietnam [2, 6]. The male flies are attracted to cue lure and zingerone [1, 6]. There is no report of known hosts.

The molecular phylogeny of 32 species of African *Dacus* has been studied based on two mitochondrial (*cox1*, *rrnL*) and one nuclear (*per*) gene fragments [7]. More recently, the phylogeny of 167 species of Dacini fruit flies (24 species of *Dacus*, 100 species of *Bactrocera* and 43 species of *Zeugodacus*) has been studied based on partial sequences of one mitochondrial gene (cytochrome *c* oxidase I) and six nuclear genes (CAD – CAD1 and CAD5, wingless, white-eye, phosphogluconate dehydrogenase, elongation factor 1 alpha, and period) [8]. This study confirms the monophyly of the genera *Dacus*, *Bactrocera* and *Zeugodacus*, but most groups below the genus level are not monophyletic [8]. Within the genus *Dacus*, only the subgenus *Neodacus* is monophyletic; all the other subgenera are either para- or polyphyletic [8].

Despite the large number of species, to date, only five complete mitochondrial genomes (mitogenomes) of *Dacus* fruit flies have been published and are available in the GenBank: *D.* (*Callantra*) *longicornis* [9], *D.* (*Dacus*) *bivittatus* [10], *D.* (*Didacus*) *ciliatus* [10], *D.* (*Mellesis*) *conopsoides* [11], and *D.* (*Mellesis*) *trimacula* [12].

In view of the lack of study on the mitogenome of the genus *Dacus* and the unresolved systematic status of some taxa, we sequenced and annotated the complete mitogenome of *D.* (*Mellesis*) *vijaysegarani* to determine its genomic features and phylogenetic relationship with members of the Dacinae and other subfamilies of Tephritidae. We also included two other families (Platystomatidae and Lonchaeidae) of Tephritoidea for comparison. In addition, although Peninsular Malaysia is the type locality of *D. vijaysegarani*, the species was also described in other localities including Thailand and Vietnam [2, 6]. Hence, we also determined the phylogenetic relationships between *D. vijaysegarani* from different localities and closely related Tephritid fruit flies, with the *cox1* gene sequences obtained from the NCBI database.

Materials and methods

Specimen collection and mitochondrial DNA extraction

The male fruit fly of *D. vijaysegarani* was collected in a wayside forest of Wang Kelian, Perlis, Peninsular Malaysia (6° 40′ 44″ N, 100° 11′ 3.12″ E) on 1 February 2011. It was attracted to cue lure applied on the surface of a green leaf. The specimen was collected by means of a specimen tube, preserved in absolute ethanol and stored in − 20 °C deep freezer until use for DNA extraction. David Hancock helped with the taxonomic identification according to Drew et al. [6]. The isolation of mitochondria and extraction of mtDNA were according to the method of Yong et al. [13].

Mitogenomes from GenBank, library preparation and genome sequencing

The complete mitogenomes of the genera *Dacus* (n = 5), *Bactrocera* (n = 21), *Zeugodacus* (n = 9), *Ceratitis* (n = 4), and other tephritid taxa (n = 8) available from the GenBank (Table S1) were used for phylogenetic comparison. Three other tephritoid mitogenomes (*Prosthiochaeta* sp. MT528242 and *Rivellia syngenesiae* MT410799 of the family Platystomatidae, and *Silba* sp. MK913844 of the family Lonchaeidae) available from the GenBank were also included for comparison. *Drosophila melanogaster* NC_024511 and *Drosophila yakuba* NC_001322 were used as outgroup taxa.

Sample and library preparation (using Nextera DNA Sample Preparation Kit) and genome sequencing using the Illumina MiSeq Desktop Sequencer (2 × 150 bp paired-end reads) (Illumina, USA) were as described in Song et al. [14]. The mitogenome sequence has been deposited in the GenBank, under the accession number MW429439.

Analysis of mitogenome

The overall quality of the raw sequence reads, obtained from the MiSeq system in FASTQ format, was assessed from their phred scores using FastQC software [15]. Sequences with lower than Q20 phred score and ambiguous nucleotides were trimmed and removed using CLC genomic workbench v.7.0.4 (Qiagen, Germany). Quality-filtered DNA sequences were mapped against the reference *D. conopsoides* mitogenome (NC_043843); a *de novo* assembly was then performed on the mapped DNA sequences. Contigs larger than 13 kbp were extracted for a BLAST search against NCBI nucleotide database to identify the mitochondrial genome. In addition, NOVO Plasty was used for *de novo* assembly of demultiplexed raw sequence reads, with different lengths of k-mer and the mitogenome of *D. conopsoides* (NC_043843) as the seed sequence [16]. The assembled genomes were aligned and examined for terminal repeats to evaluate their circularity and completeness.

Gene annotation, visualization and comparative analysis

Gene annotation of the assembled mitogenome was first carried out at MITOS web-server (http://mitos.bioinf.uni-leipzig.de/index.py) [17]. The reference mitogenomes of all available *Dacus* species were used to validate the

coding regions using nucleotide-nucleotide BLAST (BLASTn) and protein-protein BLAST (BLASTp) [18]. ClustalW [19] was used to align the 13 PCGs in order to curate the gene boundaries, the start and stop codons of PCGs as well as the overlapping and intergenic spacer regions. The overlapping and intergenic spacer regions were curated manually. MEGA X [20] was used to calculate the nucleotide composition, amino acid frequency and relative synonymous codon usage (RSCU). DnaSP6.0 [21] was used to estimate the ratios of non-synonymous substitutions (Ka) and synonymous (Ks) substitutions for the PCGs. The AT and GC skewness were determined according to Perna and Kocher [22]. Tandem Repeats Finder (http://tandem.bu.edu/trf/trf.html) [23] was used to check for inverted repeats (palindromes) in the control region. The circular map of the mitogenome was created using blast ring image generator (BRIG) [24].

Phylogenetic analysis

The nucleotide sequences of 13 PCGs and two rRNA genes of all mitogenomes were aligned using MAFFT version 7 [25]. Using MAFFT, the rRNA genes were treated as highly divergent data, aligned by an auto algorithm that selected an appropriate strategy instead of the default FFT-NS-2, and adjusted for their direction according to the first sequence. The aligned sequences of individual genes were concatenated into a dataset of 15 mt-genes (13 PCGs, 2 rRNA genes).

The concatenated dataset was imported into PhyloSuite [26] for maximum likelihood (ML) phylogenetic analysis based on IQ-tree [27] under ultrafast bootstrap algorithm with 10,000 replicates. ModelFinder [28] based on the Bayesian information criterion [29] was used to determine the best-fit nucleotide substitution models. The phylogenetic trees were exported in newick format and visualized in MEGA X [20].

Bayesian analysis was conducted using the Markov chain Monte Carlo (MCMC) method via Mr. Bayes v.3.1.2 [30], with two independent runs of 2×10^6 generations with four chains, and with trees sampled every 200th generation. The best-fit nucleotide substitution models were determined by Kakusan v.3 [31], using the Bayesian Information Criterion [29]. Likelihood values for all post-analysis trees and parameters were evaluated for convergence and burn-in using the "sump" command in MrBayes and the computer program Tracer v.1.5 (http://tree.bio.ed.ac.uk/software/tracer/). The first 200 trees from each run were discarded as burn-in (where the likelihood values were stabilized prior to the burn-in), and the remaining trees were used for the construction of a 50% majority-rule consensus tree. The phylogenetic tree was viewed and edited by FigTree v.1.4 [32].

Phylogenetic analysis based on *cox1* gene

The *cox1* gene sequences (complete/near complete and COI-3P as well as COI-5P partial sequences) of *D. vijaysegarani* and other *Dacus* species available in the GenBank were used for phylogenetic comparison, with *Zeugodacus caudatus* Malaysia and *Z. caudatus* Indonesia as outgroup taxa. The *cox1* gene sequences were aligned using MAFFT [25] and the 5' and 3' end of the alignment were manually trimmed using MEGA X [20]. The best fit nucleotide substitution model for maximum likelihood (ML) analysis was determined using ModelFinder [28]. A ML analysis was performed using the IQ-TREE [27] under ultrafast bootstrap algorithm with 10,000 replicates. The phylogenetic tree was visualized in MEGA X [20]. The genetic distances among the *cox1* gene sequences of different *Dacus* species were calculated in MEGA X using the Kimura two-parameter (K2P) substitution model [20]. The BI analysis was conducted as for the 15mt-genes.

Results

Mitogenome features

The total length of the complete mitogenome of *D. vijaysegarani* was 15,886 bp (Table 1, Fig. S1). Its length was slightly longer than those of *D. bivittatus* (15,833 bp), *D. ciliatus* (15,808 bp), *D. conopsoides* (15,852 bp), and *D. trimacula* (15,851 bp) but shorter than that of *D. longicornis* (16,253 bp). It was AT rich (73.0%), as is also in the mitogenomes of the other five *Dacus* species (Table S2). All the six *Dacus* whole mitogenomes had positive values for AT skewness and negative values for GC skewness, indicating the bias toward the use of Cs over Gs (Table S2).

Both the J and N strands of the *Dacus* mitogenomes were AT rich, with the A + T content of the N strand slightly higher than that of the J strand (Table S2), and with positive skewness value for the N strand in all the six *Dacus* mitogenomes but variable skewness value for the J strand (Table S2). The GC skewness value was negative for both the J and N strands (Table S2).

The mitogenome of *D. vijaysegarani* had identical gene order with the published mitogenomes of the genus *Dacus* [9–12], with 13 protein-coding genes (PCGs), two rRNA genes, 22 tRNAs, a non-coding A + T rich control region, and a large number of intergenic sequences (spacers and overlaps) (Table 1; Fig. S1). There were 16 intergenic overlaps and 13 spacers; the longest spacer had 58 bp between *trnQ* and *trnM*, and the longest overlap with − 42 bp was between *trnL1* and *rrnL* (Table 1).

Table 1 Gene order and features of the mitochondrial genome of *Dacus vijaysegarani*

	Location	Strand	Length (bp)	Spacer(+)/ Overlap(−)*	Start codon	Stop codon	A + T%
trnI(atc)	1–66	J	66	− 3			
trnQ(caa)	64–132	N	69	58			
trnM(atg)	191–259	J	69	0			
nad2	260–1282	J	1023	9	ATT	TAA	73.7
trnW(tga)	1292–1359	J	68	− 8			
trnC(tgc)	1352–1414	N	63	1			
trnY(tac)	1416–1482	N	67	− 2			
cox1	1481–3019	J	1539	− 5	TCG	TAA	64.7
trnL2(tta)	3015–3080	J	66	4			
cox2	3085–3774	J	690	5	ATG	TAA	68.5
trnK(aag)	3780–3850	J	71	1			
trnD(gac)	3852–3919	J	68	0			
atp8	3920–4081	J	162	− 7	ATT	TAA	71.7
atp6	4075–4752	J	678	− 1	ATG	TAA	67.9
cox3	4752–5540	J	789	6	ATG	TAA	65.4
trnG(gga)	5547–5612	J	66	0			
nad3	5613–5966	J	354	− 2	ATA	TAG	72.4
trnA(gca)	5965–6030	J	66	10			
trnR(cga)	6041–6104	J	64	37			
trnN(aac)	6142–6206	J	65	0			
trnS1(agc)	6207–6274	J	68	0			
trnE(gaa)	6275–6339	J	65	18			
trnF(ttc)	6358–6422	N	65	0			
nad5	6423–8141	N	1719	16	ATT	T	73.4
trnH(cac)	8158–8221	N	64	3			
nad4	8225–9565	N	1341	− 7	ATG	TAA	73.7
nad4L	9559–9855	N	297	2	ATG	TAA	75.8
trnT(aca)	9858–9922	J	65	0			
trnP(cca)	9923–9988	N	66	2			
nad6	9991–10,515	J	525	− 1	ATT	TAA	77.7
cob	10,515–11,651	J	1137	− 2	ATG	TAG	68.1
trnS2(tca)	11,650–11,716	J	67	16			
nad1	11,733–12,671	N	939	10	ATA	TAG	72.6
trnL1(cta)	12,682–12,746	N	65	− 42			
rrnL	12,705–14,078	N	1374	− 2			79.6
trnV(gta)	14,077–14,148	N	72	− 1			
rrnS	14,148–14,939	N	792	0			75.0
control region	14,940–15,886		947				80.4

*Minus sign indicates overlap

Protein-coding genes and codon usage

The A + T content for the 13 PCGs of *D. vijaysegarani* mitogenome was 70.6 %, with negative AT skewness value of − 0.147; the GC skewness value was − 0.014 (Table S2). For the individual PCGs, it ranged from 64.7 % for *atp8* to 77.7 % for *nad6* (Table S3). The PCGs of *D. vijaysegarani* mitogenome were characterized by four start codons (Table 1, Table S4): ATG, ATT, ATA and TCG. There were three stop codons for the PCGs (Table 1): TAA (9 PCGs); TAG (3 PCGs), and truncated incomplete T (1 PCG). The incomplete stop codon was presumed to be completed by post-translational polyadenylation [33].

The frequency of individual amino acids was highly similar in the six congeners of *Dacus* (Fig. 1). However, the frequency of the codons varied among these mitogenomes. The predominant amino acids (with frequency above 300) in all the six *Dacus* mitogenomes were

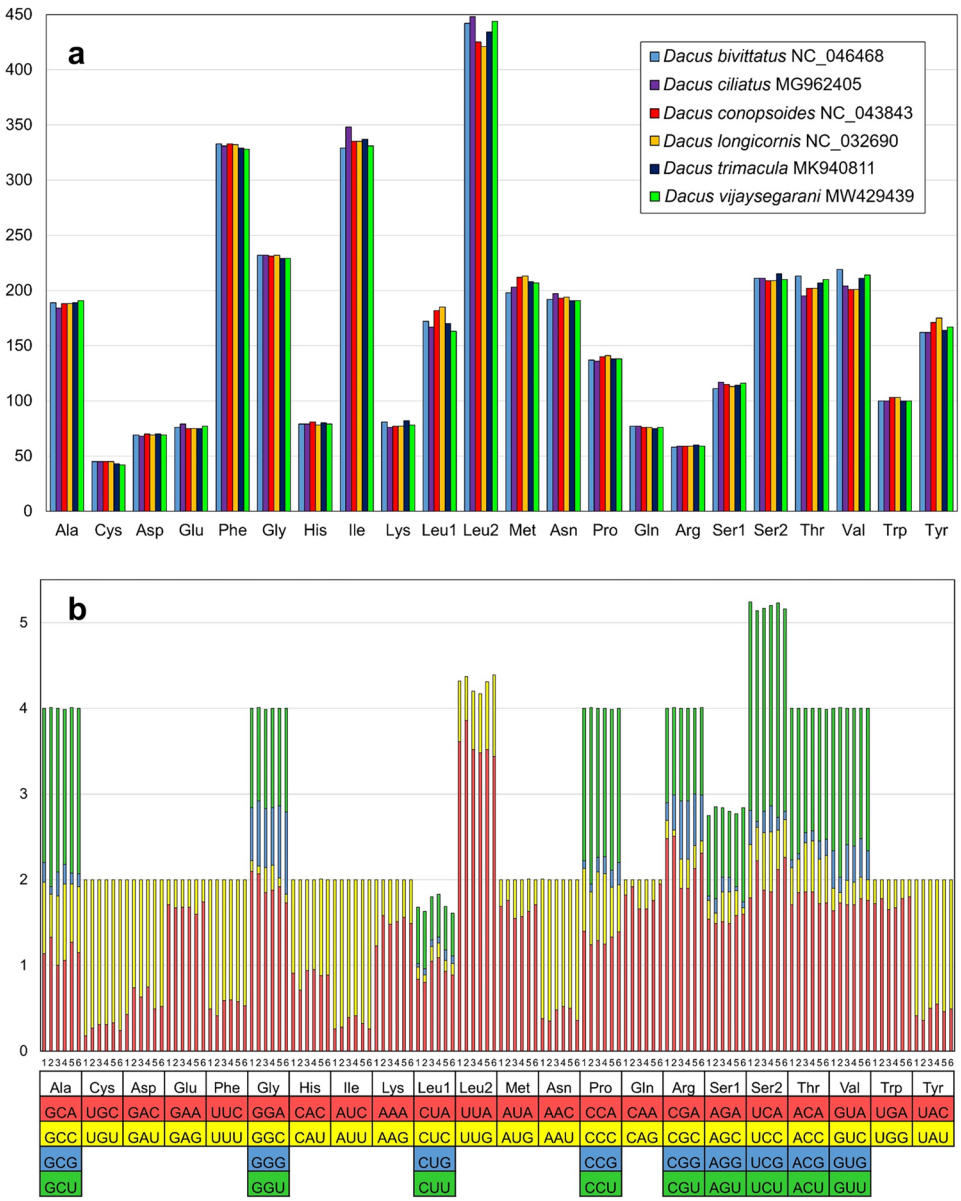

Fig. 1 Amino acid frequency (**a**) and relative synonymous codon usage (**b**) of the protein-coding genes in *Dacus* mitogenomes. 1, *Dacus bivittatus* NC_046468; 2, *Dacus ciliatus* MG962405; 3, *Dacus conopsoides* NC_043843; 4, *Dacus longicornis* NC_032690; 5, *Dacus trimacula* MK940811; 6, *Dacus vijaysegarani* MW429439

phenylalanine, isoleucine and leucine2 (Table S5; Fig. 1). Analysis of the relative synonymous codon usage (RSCU) revealed the biased usage of A/T than G/C at the third codon position (Fig. 1). The most commonly used codon was UUA (TTA) encoding for leucine2 (Fig. 1).

Excepting the *nad6* gene (Ka/Ks = 0.870 ± 0.858; range 0.103–2.763) and *nad4L* gene (Ka/Ks = 0.174 ± 0.329; range 0.041–1.360), the Ka/Ks ratio (an indicator of selective pressure on a PCG) was less than 1 for the other 11 PCGs in the six *Dacus* mitogenomes, indicating purifying selection (Table S6; Fig. S2). The sequence of the Ka/Ks ratio was *cob* < *cox1* < *nad4L* < *atp8* = *cox1* < *atp6* < *cox2* < *nad1* < *nad4* < *nad3* < *nad2* < *nad5* < *nad6*.

Ribosomal RNA genes and transfer RNA genes

Of the two rRNA genes in *D. vijaysegarani*, *rrnS* (792 bp) was much shorter than *rrnL* (1374 bp). The same condition was found in the other *Dacus* mitogenomes, with little variation [9–12]. Both the rRNA genes of all the *Dacus* mitogenomes were AT-rich, with negative AT skewness and positive GC skewness (Table S3).

Excepting the mitogenome of *D. conopsoides* with 23 tRNAs (with duplicated *trnF* gene and a pseudogene of partially duplicated *trnE* gene) [11], the other *Dacus* mitogenomes had 22 tRNAs (Fig. S3). The mitogenome of *D. vijaysegarani*, and the other *Dacus* species, had the three

main tRNA clusters common to other insects: (1) I-Q-M; (2) W-C-Y; and (3) A-R-N-S1-E-F (Fig. S1).

Most of the tRNAs in the *Dacus* mitogenomes had canonical clover-leaf secondary structure (Fig. S3). Histidine and phenylalanine did not possess a TΨC loop in *D. vijaysegarani*, while serine S1 had an aberrant DHU loop structure with loss of the stem (Fig. S3).

Control region

The length of the non-coding control region in the *Dacus* mitogenomes was variable, ranging from 812 bp in *D. conopsoides* [11] to 1343 bp in *D. longicornis* [9]; the length in *D. vijaysegarani* was 947 bp. All the *Dacus* species had positive AT skewness value and negative GC skewness value (Table S2).

The control region in the *D. vijaysegarani* mitogenome possessed relatively short polynucleotide stretches. There was significantly more poly-A than poly-T stretches: 13 poly-A and three poly-T stretches.

The simple tandem repeats present in the control region of *D. vijaysegarani* included $(ATT)_2$, $(AAT)_2$, $(AATT)_2$, $(ATAAA)_2$, $(TAA)_2$, $(TAAA)_2$, $(AAAT)_3$, $(CTA)_3$, $(TA)_3$, and $(TTA)_3$. The palindromes included AATTAA (n = 2), TAAAAT (n = 4), and TTAAAATT.

There were 13 nucleotide motifs in the control region of *Dacus* mitogenomes (Table S7). *D. vijaysegarani* had the highest number for the motifs TAAAAT palindrome (n = 13) and AAATT (n = 23).

Phylogenetic analysis

The phylogenetic trees based on 15 mt-genes (13 PCGs and 2 rRNA genes) revealed similar topology with good to very good nodal support based on BI (Fig. 2a) and ML (Fig. 2b) methods. The family Tephritidae comprising 32 species formed a monophyletic group, which was clearly separated from two other tephritoid families (Platystomatidae and Lonchaeidae).

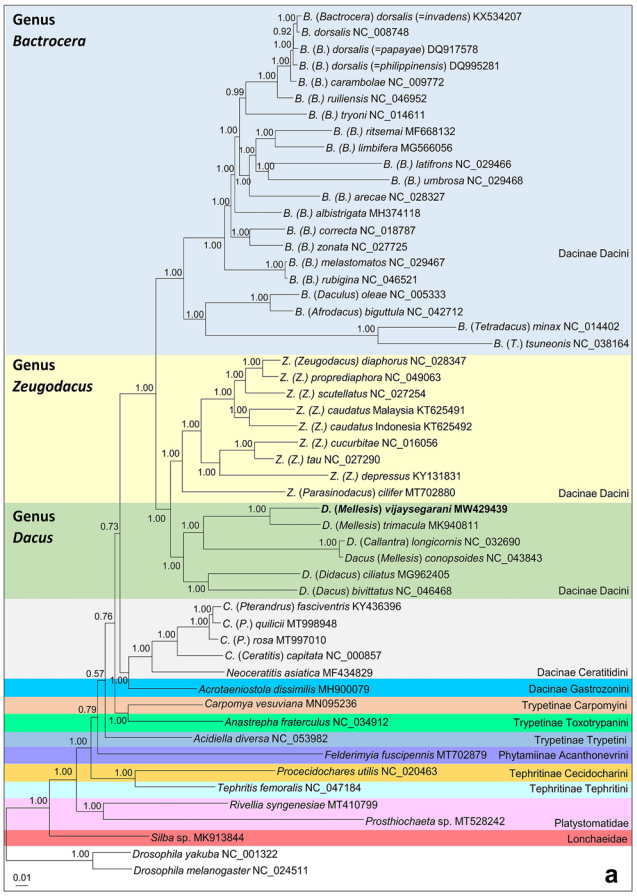

Fig. 2 **a** Bayesian inference (BI) phylogenetic tree based on 15 mitochondrial genes (13 PCGs and 2 rRNAs) of the whole mitogenomes of *Dacus* and other tephritoid taxa with *Drosophila* taxa as outgroup. Numeric values at the nodes are Bayesian posterior probabilities. **b** Maximum likelihood (ML) tree based on 15 mitochondrial genes (13 PCGs and 2 rRNAs) of the whole mitogenomes of *Dacus* and other tephritoid taxa with *Drosophila* taxa as outgroup. Numeric values at the nodes are bootstrap values

The genera *Dacus*, *Zeugodacus* and *Bactrocera* formed a distinct clade from the other tephritid taxa (Fig. 2). The genus *Dacus* formed a monophyletic group in the subclade containing also the genus *Zeugodacus* (excepting *Z. cilifer* which was an outlier in the ML tree); this *Dacus-Zeugodacus* subclade was distinct from the *Bactrocera* subclade. *D.* (*Mellesis*) *vijaysegarani* formed a lineage with *D.* (*Mellesis*) *trimacula* in the subcluster containing also the lineage of *D.* (*Mellesis*) *conopsoides* and *D.* (*Callantra*) *longicornis*. *D.* (*Dacus*) *bivittatus* and *D.* (*Didacus*) *ciliatus* formed a distinct subcluster.

The subgenus *Bactrocera* in the *Bactrocera* subclade was monophyletic, forming a distinct cluster from that containing the lineage of the subgenera *Daculus* and *Afrodacus* as well as the subgenus *Tetradacus* (Fig. 2).

Based on the near complete *cox1* sequences of *Dacus* taxa, the Malaysia and Vietnam taxa of *D. vijaysegarani* were nested in different lineages (Fig. 3). They were genetically distinct, with a large genetic distance of 9.15 %.

Discussion

Mitochondrial genomes of insects have been very extensively studied and applied particularly to studies regarding phylogeny and evolution [34]. To date, the complete mitogenomes of 36 species of tephritid fruit flies include: 17 species of genus *Bactrocera* (excluding 3 conspecific species of *B. dorsalis*); 5 of *Dacus*, 9 of *Zeugodacus* (including the cryptic species of *Z. caudatus* [35]); 4 of *Ceratitis*; and 1 each of *Acidiella*, *Anastrepha*, *Carpomya*, *Neoceratitis*, *Procecidochares* and *Tephritis* (Fig. 2). The present study has added an additional complete mitogenome for the genus *Dacus*.

In the present study, the family Tephritidae comprising 32 species formed a monophyletic group (Fig. 2). In an earlier study based on mitochondrial 12 S, 16 S and COX2 genes, only the BI tree reveals Tephritidae (n = 79 species) as a monophyltic group; the ME (minimum evolution) tree does not support this result [36].

The present findings on *Dacus* phylogeny, although based on very limited number of *Dacus* taxa, agrees with the findings of Leblanc et al. [2] based on seven genes, which group members of different subgenera in the same lineage: *D.* (*Mellesis*) *discophorus* forming a lineage with *D.* (*Callantra*) *longicornis* and *D.* (*Callandra*) *axanus*; and *D.* (*Didacus*) *ciliatus* forming a lineage with *D.* (*Dacus*) *bivittatus* and other species of the subgenus *Dacus*. However, *D.* (*Mellesis*) *discophorus*, before the subgeneric revision, was included as a member of the subgenus *Callantra*, viz. *D.* (*Callantra*) *discophorus* [6].

It is noteworthy that *D.* (*Callantra*) *longicornis* and *D.* (*Mellesis*) *conopsoides* in the present study show a very close genetic affinity, with an exceptionally low genetic distance of 0.86 % based on 15 mt-genes; the closely related *D. vijaysegarani* and *D. trimacula* have a genetic distance of 8.23 % based on 15 mt-genes (Table 2). An example of such a low genetic distance is between the sibling species *Bactrocera carambolae* and *B. dorsalis*: 15 mt-genes, $p = 1.2\,\%$ [13].

D. longicornis is morphologically similar to *D. conopsoides*, *D. insulosus* and *D. trimacula*, and has been regularly misidentified [6]. Before the revised classification [3], which places *D. conopsoides* under the subgenus *Mellesis*, it was earlier treated as a member of the subgenus *Callantra*, viz. *D.* (*Callantra*) *conopsoides* [6]. Our present dataset cannot resolve the question of possible misidentification of *D. longicornis* and the subgenus status of *D. conopsoides*. An extensive taxon sampling and phylogeography study is

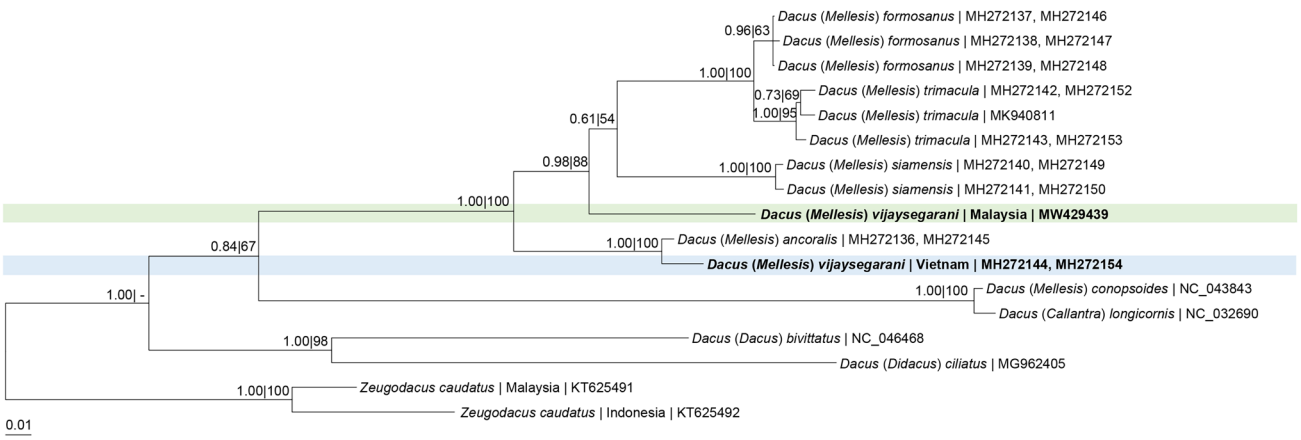

Fig. 3 Phylogenetic tree based on near complete *cox1* sequences (1465 bp) of *Dacus* taxa with *Zeugodacus caudatus* as outgroup taxa. Numeric values at the nodes are Bayesian posterior probabilities and ML bootstrap values

Table 2 Pairwise genetic distance (%) of *Dacus* taxa based on 13 protein-coding genes (PCGs, below diagonal) and 15 mitochondrial genes (13 PCGs and 2 rRNA genes, above diagonal)

Taxon	*D. vijaysegarani*	*D. trimacula*	*D. conopsoides*	*D. longicornis*	*D. bivittatus*	*D. ciliatus*
D. vijaysegarani	–	8.23	17.13	17.38	15.73	15.73
D. trimacula	8.95	–	17.19	17.39	15.90	15.90
D. conopsoides	18.72	18.77	–	0.86	17.02	17.37
D. longicornis	19.01	19.01	0.94	–	17.21	17.21
D. bivittatus	17.28	17.34	18.60	18.84	–	13.67
D. ciliatus	17.35	18.00	18.90	18.98	15.02	–

needed to elucidate the genetic affinity of *D. conopsoides* and *D. longicornis* as well as other *Dacus* taxa.

Monophyly of the subfamily Dacinae is not supported by our study (Fig. 2). The tribes Ceratitidini and Gastrozonini do not form a monophyletic group with the tribe Dacini. They form a subclade in the clade which contains also the subclade comprising the subfamilies Tephritinae and Trypetinae. Two recent studies based on 15 mt-genes also show the Ceratitidini tribe to be closer to *Anastrepha* (Trypetinae) than to the Dacini tribe [10, 11]. In some earlier taxonomic treatments, the tribes Ceratitidini and Gastrozonini have been placed under the subfamily Ceratitidinae [37, 38]. As our present study included only a single species of Gastrozonini, a more extensive taxon sampling is needed to address this taxonomic issue.

Based on *cox1* gene [2] and concatenation of *cox1* and six nuclear genes [8], *D. vijaysegarani* from Vietnam is closely related to *D. ancoralis* from Sri Lanka. The Vietnam taxon of *D. vijaysegarani* also forms a monophyletic COXI cluster with specimens from Bangladesh resembling *D. jacobi* from India [39]. Our present study based on near complete *cox1* gene shows a large genetic distance of 9.15 % between the Vietnam and Malaysia taxa of *D. vijaysegarani*. The Vietnam taxon can therefore be reasonably considered to be not conspecific with *D. vijaysegarani* Malaysia. As Peninsular Malaysia is the type locality of *D. vijaysegarani* [6], the Vietnam taxon warrants to be accorded different specific status as a component taxon of the *D. vijaysegarani* species complex.

In addition to the Vietnam taxon, a male specimen from Sabah (Borneo Island) named as *D. vijaysegarani* [40] may be another member of the *D. vijaysegarani* complex. The scutum of the Sabah taxon is all black [40] whilst that of *D. vijaysegarani* type taxon is black with a narrow dark red-brown area along the posterior margin [6]. Molecular markers will help to differentiate this and other morphologically very similar taxa.

It is evident that studies on the mitogenomes of an extensive taxa sampling of various taxonomic orders of tephritid fruit flies are needed to provide a potentially more robust phylogeny and systematics. Compared to partial individual genes, mitogenome provides more gene contents for phylogenetic and systematics analyses.

Conclusions

In summary, we have successfully sequenced the whole mitochondrial genome of *D. vijaysegarani* from Peninsular Malaysia by next generation sequencing. The genome features are similar to other *Dacus* fruit flies. Phylogenetic analysis based on 15 mitochondrial genes (13 PCGs and two rRNA genes), reveals *Dacus*, *Zeugodacus* and *Bactrocera* forming a distinct clade; these three genera are monophyletic. *D.* (*Mellesis*) *vijaysegarani* forms a lineage with *D.* (*Mellesis*) *trimacula* in the subcluster containing also the lineage of *D.* (*Mellesis*) *conopsoides* and *D.* (*Callantra*) *longicornis*. *D.* (*Dacus*) *bivittatus* and *D.* (*Didacus*) *ciliatus* form a distinct subcluster. *D.* (*Callantra*) *longicornis* and *D.* (*Mellesis*) *conopsoides* show a very close genetic affinity. The subfamily Dacinae, as presently constituted, is not monophyletic. Based on the near complete *cox1* sequences, the Malaysia and Vietnam taxa of *D. vijaysegarani* are genetically distinct and therefore may not be conspecific. The tribes Ceratitidini and Gastrozonini do not form a monophyletic group with the tribe Dacini. In sum, this study characterized the complete mitochondrial genome of *D. vijaysegarani* and contributes to our understanding of the mitochondrial gene evolution within tephritid fruit flies. More importantly, the data provided valuable information for phylogenetic analysis and species differentiation among other *Dacus* species in the future.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11033-021-06608-2.

Acknowledgements We thank our institutions for providing various research facilities and other support. This study was funded in part by the Universiti Malaya (H-5620009). H-S Yong thanks David Hancock for taxonomic identification of the specimen, and Universiti Malaya Deputy Vice Chancellor Kamila Ghazali for permission to carry out research during the Covid-19 pandemic.

Author contributions HSY, SLS and KOC conceived and designed the study. SLS carried out experimental work. HSY, SLS and KOC carried out data analysis. HSY, SLS and KOC wrote the manuscript. All authors contributed to interpretation of results, read and approved the final draft.

Declarations

Conflict of interest The authors declare that they have no conflicts of interest.

Ethical approval *Dacus vijaysegarani* is not endangered or protected by law. No permits are required to study this fruit fly.

Informed consent The authors declare their consent to participate in this study.

Consent for publication The authors declare their consent for publication of this study.

References

1. Doorenweerd C, Leblanc L, Norrbom AL, San Jose M, Rubinoff D (2018) A global checklist of the 932 fruit fly species in the tribe Dacini (Diptera, Tephritidae). ZooKeys 730:19–56
2. Leblanc L, Doorenweerd C, San Jose M, Sirisena UGAI, Hemachandra KS, Rubinoff D (2018) Description of a new species of *Dacus* from Sri Lanka, and new country distribution records (Diptera, Tephritidae, Dacinae). ZooKeys 795:105–114
3. Hancock DL, Drew RAI (2006) A revised classification of subgenera and species groups in *Dacus* Fabricius (Diptera, Tephritidae). Instrumenta Biodivers VII:167–205
4. Hancock DL (2009) Additions and amendments to a recent classification of *Dacus* Fabricius (Diptera: Tephritidae: Dacinae). Aust Entomol 36(2):67–70
5. Hancock DL (2010) Supplementary additions to a recent classification of *Dacus* Fabricius (Diptera: Tephritidae: Dacinae), with notes on the *D* .(*Neodacus*) *newmani* group. Aust Entomol 37(2):45–46
6. Drew RAI, Hancock DL, White IM (1998) Revision of the tropical fruit flies (Diptera: Tephritidae: Dacinae) of South-east Asia. II. *Dacus* Fabricius. Invertebr Taxon 12:567–654
7. Virgilio M, de Meyer M, White IM, Bacleljau T (2009) African *Dacus* (Diptera: Tephritidae): molecular data and host plant associations do not corroborate morphology-based classifications. Mol Phylogenet Evol 51:531–539
8. San Jose M, Doorenweerd C, Leblanc L, Barr N, Geib SM, Rubinoff D (2018) Incongruence between molecules and morphology: A seven-gene phylogeny of Dacini fruit flies paves the way for reclassification (Diptera: Tephritidae). Mol Phylogenet Evol 121:139–149
9. Jiang F, Pan X, Li X, Yu Y, Zhang J, Jiang H, Dou L, Zhu S (2016) The first complete mitochondrial genome of *Dacus longicornis* (Diptera: Tephritidae) using next-generation sequencing and mitochondrial genome phylogeny of Dacini tribe. Sci Rep 6:36426
10. Zhang Y, Feng S, Fekrat L, Jiang F, Khathutshelo M, Li Z (2019) The first two complete mitochondrial genome of *Dacus bivittatus* and *Dacus ciliatus* (Diptera: Tephritidae) by next-generation sequencing and implications for the higher phylogeny of Tephritidae. Int J Biol Macromol 140:469–476
11. Song S-L, Yong H-S, Suana IW, Lim P-E, Eamsobhana P (2019) Complete mitochondrial genome of *Dacus conopsoides* (Insecta: Tephritidae) with tRNA gene duplication and molecular phylogeny of Dacini tribe. J Asia-Pacif Entomol 22:997–1003
12. Ren Y-L, Wang T, Dai R, Wu Z-Y, Wang J (2019) Complete mitochondrial genome of *Dacus trimacula* (Diptera: Tephritidae) using next-generation sequencing from China. Mitochondrial DNA Part B 4(2):2355–2356
13. Yong H-S, Song S-L, Lim P-E, Eamsobhana P, Suana IW (2016) Complete mitochondrial genome of three *Bactrocera* fruit flies of subgenus *Bactrocera* (Diptera: Tephritidae) and their phylogenetic implications. PLoS ONE 11(2):e0148201
14. Song S-L, Yong H-S, Suana IW, Lim P-E (2018) Complete mitochondrial genome of *Bactrocera ritsemai* (Insecta: Tephritidae) and phylogenetic relationship with its congeners and related tephritid taxa. J Asia-Pacif Entomol 21:251–257
15. Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge
16. Dierckxsens N, Mardulyn P, Smits G (2017) NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res 45:e18
17. Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritsch G, Pütz J, Middendorf M, Stadler PF (2013) MITOS: inproved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol 69:313–319
18. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410
19. Hung J-H, Weng Z (2016) Sequence alignment and homology search with BLAST and ClastalW. Cold Spring Harb Protoc. https://doi.org/10.1101/pdb.prot093088
20. Kumar K, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549
21. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A (2017) DnaSP 6: DNA Sequence Polymorphism Analysis of Large Datasets. Mol Biol Evol 34:3299–3302
22. Perna NT, Kocher TD (1995) Patterns of nucleotide composition in fourfold degenerate sites of animal mitochondrial genomes. J Mol Evol 41:353–358
23. Benson G (1999) Tandem repeats finder: a program to analyse DNA sequences. Nucleic Acids Res 27:573–580
24. Alikhan N-F, Petty NK, Zakour NLB, Beatson SA (2011) BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genom 12:402
25. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol 30:772–780
26. Zhang D, Fangluan G, Jaković I, Zou H, Zhang J, Li W-X, Wang GT (2020) PhyloSuite: and integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour 20:348–355
27. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. Mol Biol Evol 32:268–274
28. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14:587–589
29. Schwarz G (1978) Estimation of the dimension of a model. Ann Stat 6:461–464
30. Huelsenbeck JP, Ronquist F F (2001) MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755
31. Tanabe AS (2007) Kakusan: a computer program to automate the selection of a nucleotide substitution model and the configuration of a mixed model on multilocus data. Mol Ecol Notes 7:962–964

32. Rambaut A. Tree figure drawing tool, version 1.4.0 [computer program]. http://tree.bio.ed.ac.uk/software/figtree/. Accessed 11 May 2020
33. Ojala D, Montoya J, Attardi G (1981) tRNA punctuation model of RNA processing in human mitochondria. Nature 290:470–474
34. Cameron SL (2013) Insect mitochondrial genomics: implications for evolution and phylogeny. Annu Rev Entomol 59:95–117
35. Yong H-S, Song S-L, Lim P-E, Eamsobhana P, Suana IW (2016) Differentiating sibling species of *Zeugodacus caudatus* (Insecta: Tephritidae) by complete mitochondrial genome. Genetica 144:513–521
36. Han H-Y, Ro K-E (2009) Molecular phylogeny of the family Tephritidae (Insecta: Diptera): new insight from combined analysis of the mitochondrial 12S, 16S, and COII genes. Mol Cells 27:55–66
37. Permkam S, Hancock DL (1994) Australian Ceratitinae (Diptera: Tephritidae). Invertebr Taxon 8:1325–1341
38. Hancock DL, McGuire DJ (2002) New species and records of non-dacine fruit flies (Insecta: Diptera: Tephritidae) from south and southeast Asia. Steenstrupia 27(1):1–17
39. Doorenweerd C, San Jose M, Leblanc L, Barr N, Geib S, Chung AYC, Dupuis J, Ekayanti A, Fiegalan ER, Hemachandra KS, Hossain AA, Huang C-L, Hsu Y-F, Morris KY, Mustapeng AM, Niogret J, Pham TH, Sirisena UGAI, Rubinoff D (2020) DNA barcodes and reliable molecular identifications in a diverse group of invasive pests: lessons from *Bactrocera* fruit flies on variation across the COI gene, introgression, and standardization. bioRxiv. https://doi.org/10.1101/2020.11.23.394510
40. Chua TH (2010) Fruit flies (Diptera: Tephritidae) from Malaysia and Brunei Darussalam: new species and records. Fla Entomol 93:483–488