Original Research Article

DNA barcoding of economically important fruit flies (Diptera: Tephritidae) from the lower Gangetic plains of eastern India

.

ABSTRACT

The Lower Gangetic Plains are India's central horticultural crop production regions. A major insect pest of several commercially valuable fruit and vegetable crops is fruit flies (Diptera: Tephritidae), which are, hence, one of the significant causes of hindering effective crop production. Therefore, rapid and precise species identification techniques are essential to controlling, managing, and quarantining these pests. Furthermore, reliably identifying fruit flies is exceedingly challenging due to their striking physical similarities. In this work, we investigated the effectiveness of mitochondrial cytochrome c oxidase I (COI) gene sequences for identifying seven fruit fly species, involving only the genus Bactrocera from the lower Gangetic Plains in eastern India. The sequences of the studied species showed 99%–100% similarity with sequences from within and outside the country. COI barcodes were able to distinguish between species, with K2P intraspecific genetic divergences ranging from 0% to 1% and K2P interspecific genetic divergences ranging from 3.65% to 28.46%, thus maintaining a proper barcode gap. Phylogenetic analyses were performed, and the results showed moderate to high supported monophyly among species. Therefore, COI barcodes have proven a highly successful alternative for quickly identifying fruit flies. This is also the first account of molecular identification of the tephritid fruit flies infesting fruits and vegetables in eastern India's agriculturally important lower Gangetic Plains.

Keywords: Bactrocera, COI gene, DNA barcode, Genetic divergence, Horticultural crops, Phylogenetic analysis

UNDER PEER REVIEW

1. INTRODUCTION

India's harvest of fruits and vegetables keeps growing because of its diversified climate and other advantageous conditions. From 2022–2023, India produced a record 212.91 MT (million tonnes) of vegetables and 108.34 MT (million tonnes) of fruits (https://pib.gov.in/PressReleseDetailm.aspx?PRID=1985479). West Bengal is the most productive potato-growing state and is well-known for its vegetable production. It also produces many other crops, including leafy greens, tomatoes, cucumbers, and pumpkins. Sundarbans are famous for brinjal production. West Bengal produced 30,555.993 tonnes of

vegetables and 3,857.630 tonnes of fruits in 2023. However, the effective production of these crops in the region is severely hindered by pests and crop illnesses. In this group of pests, tephritids are identified globally as a severe risk to the agriculture sector, including India [1, 2, 3]. The Tephritidae family has about five thousand species [4]. Out of all the fruit fly species, 325 species are known to exist in the Indian subcontinent, and 243 of them (across 79 genera) have been documented from India alone [5, 6]. Numerous of these are considered serious pests of commercial crops, causing billion-dollar losses every year [7]. Tephritid fruit flies are responsible for significant losses in fruits and vegetables, and to increase commercial output, they are frequently the focus of extensive pesticide treatments [8].

The genera *Anastrepha* Schiner, *Bactrocera* Macquart, *Ceratitis* Macleay, *Dacus* Fabricius, and *Rhagoletis* Loew include most Tephritidae pest species [9, 10]. Comprising 30 subgenera and over 500 known species, the genus Bactrocera is one of the largest in the Tephritidae family [11, 12, 13]. Within this genus, over 40 species are recognized as significant global pests affecting fruit and vegetable crops from an economic standpoint [14]. Due to its extensive host range and capacity to infiltrate and establish itself in new environments, the genus *Bactrocera* poses a significant risk to horticulture crops [15]. In India, the fruit and vegetable pests *Bactrocera dorsalis*, *Bactrocera correcta*, *Bactrocera zonata*, *Bactrocera cucurbitae*, and *Bactrocera tau* constitute significant economic losses [16]. Since fruits and vegetables are both export-oriented agricultural products, there is growing concern that fruit fly pest species may be migrating due to human intervention due to the increased commerce between nations brought about globally. As a result, both domestic and international quarantine organizations take this seriously [8]. Using an ovipositor, adult females pierce the delicate fruits to deposit their eggs beneath the skin. When fruit is afflicted, brown and resinous fluids seep out of the holes the female flies make for oviposition. The eggs develop into maggots, which feed within the fruit on pulp [3]. Fruits containing fruit fly larvae deteriorate fast. While it is occasionally feasible to remove the damaged portion of the fruit for domestic use, diseased fruits are usually not marketable and cannot be exported [3]. Crop losses might range from a few percent to 90% [17]. Since fruit fly maggots eat within the fruits, it is very hard to identify fruit flies accurately using only their larvae; instead, one must wait until the maggots are fully grown before making an accurate diagnosis. Thus, it is crucial for government and quarantine organizations everywhere in the globe to accurately and promptly identify the pest species at the port of entry.

Even for specialists, accurately identifying a fruit fly species can be challenging due to the notable similarity in physical characteristics across and within different species [9]. For insect pests like fruit flies (Diptera: Tephritidae), timely and precise identification is crucial [18]. It was suggested that species identification might be achieved by DNA barcoding based on the mitochondrial cytochrome C oxidase I (COI) [19]. Research on the effectiveness of DNA barcodes for fruit fly species identification has been conducted both in India and worldwide. While members of fruit fly species complexes had less success [20, 21, 22], most barcoding of fruit flies had high percentages of successfully classifying unknown specimens to recognized species [23, 24, 25, 26, 27, 28, 18, 8, 29].

The goal of this study is to generate mitochondrial COI sequencing data for rapid, precise, and appropriate DNA-based identification of fruit flies from Eastern India. Properly identifying economically important tephritids will help to know the diversity of these flies across the lower Gangetic plains of West Bengal, which is very important for both vegetable and fruit crop production. This is the first study from Eastern India and from the lower Gangetic Plains in West Bengal.

UNDER PEER REVIEW

2. MATERIAL AND METHODS

2.1 *Sample collection and identification*

Fruit flies were collected utilizing Cue lure-baited traps, Methyl Eugenol, and also from infested vegetables and fruits. To facilitate subsequent laboratory research, fly samples were collected in 100% pure high-grade ethanol and kept at -20 °C. From those samples, some samples were pinned, dried, and stored at the National Zoological Collection (NZC), Zoological Surveys of India (HQ), Kolkata. Before molecular research, sorted fruit fly specimens were morphologically identified using a stereomicroscope and utilizing taxonomic keys [5, 6, 30, 31]. Leica stereo-iso microscope M205A, coupled with Leica DFC 500 camera and Leica Application Suite LAS v3.6 software, was implemented for snapping images of the fly samples. According to Systema Dipterorum v5.0, valid species names were allocated [32]. The geographical coordinates of the study areas were provided using a Garmin GPS device, and the terrain map for this research was produced using ESRI's Arc GIS ® Desktop program (version 10.8). Fruit flies belonging to one genus (*Bactrocera* Macquart, 1835) and seven species were collected from various vegetable crop fields and fruit orchards of the Gangetic plains and coastal regions in West Bengal between October 2022 and August 2023 (Fig. 1). The maggots of fruit flies were nurtured to adults in the laboratory environment in the case of fruit and vegetable infestations. 13 sequences representing the seven species were generated from the collected specimens (Table 1), and the 30 other sequences (excluding outgroup) that corresponded to the aforementioned genera and subgenera were obtained from the Genbank database (Table 2).

Fig. 1 Map showing the sampling localities of tephritid species from different regions of Gangetic Plain and coastal areas of West Bengal, India.

UNDER PEER REVIEW

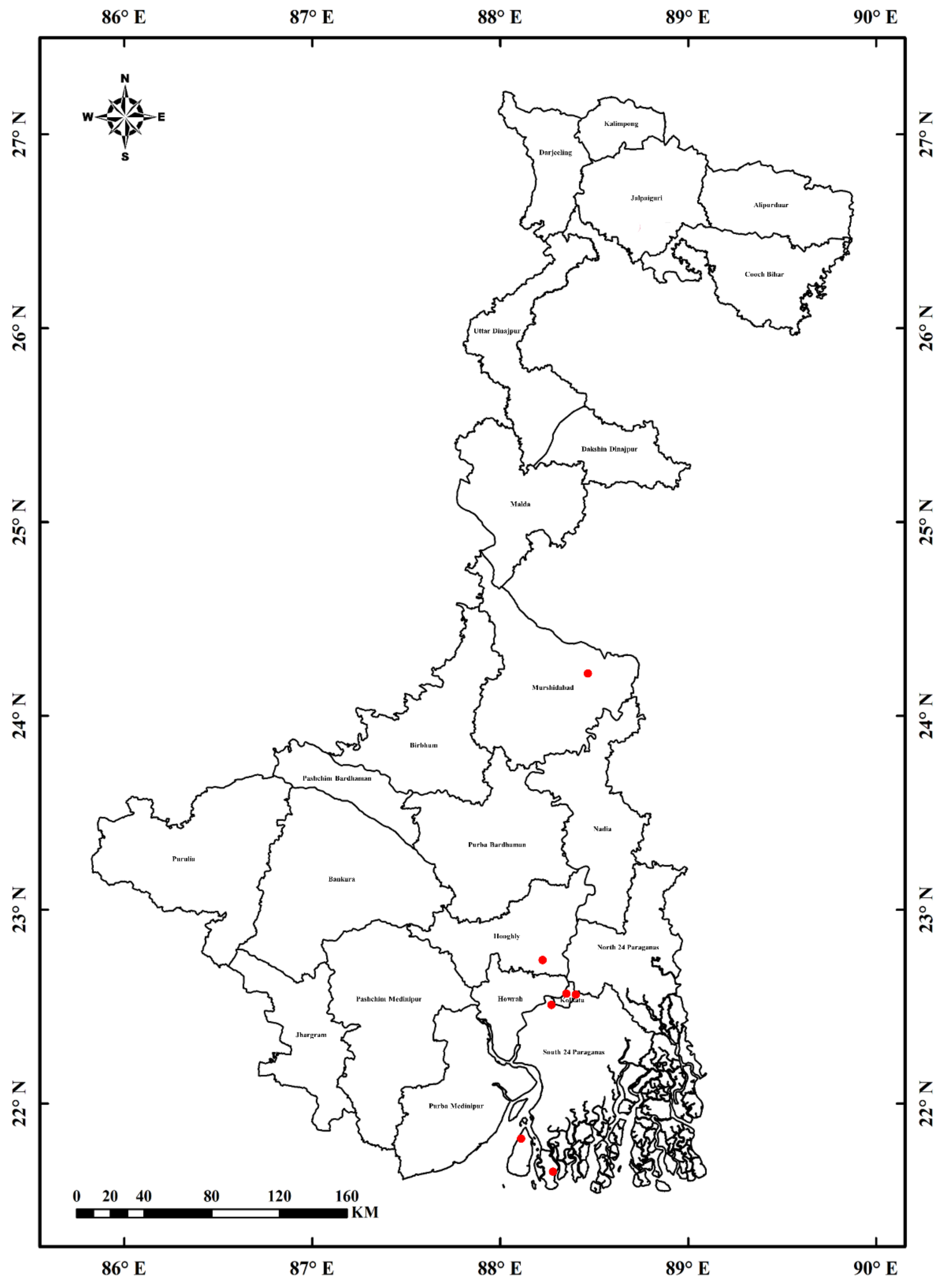


Table 1. List of the collected fruit fly specimens from the lower Gangetic plains of West Bengal:

Sl No.	Location	Accession Number	Name of species	Voucher Number	Source of collection
1.	Bakkhali, South 24 Parganas	OL442108	*Bactrocera* (*Zeugodacus*) *cucurbitae* (Coquillett, 1899)	T_2	Infested cucurbit (Spine Gourd)
2.	Sapkhali, Sagar Island	PP499254	*Bactrocera* (*Zeugodacus*) *cucurbitae* (Coquillett, 1899)	T_1	Cue Lure
3.	East Calcutta Wetlands/ Dhapa fields, Kolkata	PP499257	*Bactrocera* (*Zeugodacus*) *cucurbitae* (Coquillett, 1899)	T_4	Infested cucurbit (Bottle Gourd)
4.	Chinsurah, Hooghly	OL440711	*Bactrocera* (*Bactrocera*) *latifrons* (Hendel, 1915)	T_3	Infested Capsicum (Solanaceae)
5.	Ruhia, Murshidabad	PP499261	*Bactrocera* (*Bactrocera*) *latifrons* (Hendel, 1915)	T_5	Methyl Eugenol
6.	Sapkhali,	PP499276	*Bactrocera*	T_6	Infested cucurbit

UNDER PEER REVIEW

	Sagar Island		(*Zeugodacus*) *tau* (Walker, 1849)		(Squash)
7.	East Calcutta Wetlands/ Dhapa fields, Kolkata	PP499284	*Bactrocera* (*Bactrocera*) *dorsalis* (Hendel, 1912)	T_7	Infested Indian jujube
8.	Chinsurah, Hooghly	PP499288	*Bactrocera* (*Bactrocera*) *dorsalis* (Hendel, 1912)	T_8	Infested Guava
9.	Ruhia, Murshidabad	PP500529	*Bactrocera* (*Bactrocera*) *zonata* (Saunders, 1842)	T_9	Methyl Eugenol
10.	Sapkhali, Sagar Island	PP500532	*Bactrocera* (*Bactrocera*) *zonata* (Saunders, 1842)	T_10	Methyl Eugenol
11.	Raja Bazar, Kolkata	PP500621	*Bactrocera* (*Bactrocera*) *correcta* (Bezzi, 1916)	T_11	Infested Guava
12.	Maheshtala, South 24 Paraganas	PP500726	*Bactrocera* (*Zeugodacus*) *caudata* (Fabricius,	T_12	Infested cucurbit (Pumpkin)

UNDER PEER REVIEW

			1805)		
13.	Ruhia, Murshidabad	PP500968	*Bactrocera* (*Zeugodacus*) *caudata* (Fabricius, 1805)	T_13	Cue Lure

Table 2. GenBank sequences available in other parts of India and worldwide for the species under present study:

UNDER PEER REVIEW

Serial Number	Location	Species name	Accession Number
1	Thailand	*Bactrocera* (*Zeugodacus*) *cucurbitae* (Coquillett, 1899)	MN256074
2	Meghalaya, India	*Bactrocera* (*Zeugodacus*) *cucurbitae* (Coquillett, 1899)	MH198034
3	Australia	*Bactrocera* (*Zeugodacus*) *cucurbitae* (Coquillett, 1899)	MF970804
4	Tamil Nadu, India	*Bactrocera* (*Zeugodacus*) *cucurbitae* (Coquillett, 1899)	MN016983
5	China	*Bactrocera* (*Bactrocera*) *latifrons* (Hendel, 1915)	ON586734
6	Africa	*Bactrocera* (*Bactrocera*) *latifrons* (Hendel, 1915)	GQ154146
7	China	*Bactrocera* (*Bactrocera*) *latifrons* (Hendel, 1915)	KJ753915
8	Tamil Nadu, India	*Bactrocera* (*Zeugodacus*) *tau* (Walker, 1849)	KT151119
9	Bangladesh	*Bactrocera* (*Zeugodacus*) *tau* (Walker, 1849)	MH973720
10	China	*Bactrocera* (*Zeugodacus*) *tau* (Walker, 1849)	KF660192
11	Meghalaya, India	*Bactrocera* (*Bactrocera*) *dorsalis* (Hendel, 1912)	KT151116

UNDER PEER REVIEW

12	China	*Bactrocera* (*Bactrocera*) *dorsalis* (Hendel, 1912)	MW410926
13	Karnataka, India	*Bactrocera* (*Bactrocera*) *dorsalis* (Hendel, 1912)	KX259497
14	China	*Bactrocera* (*Zeugodacus*) *diversa* (Coquillett, 1904)	KJ833987
15	China	*Bactrocera* (*Zeugodacus*) *diversa* (Coquillett, 1904)	KJ833985
16	Meghalaya, India	*Bactrocera* (*Bactrocera*) *zonata* (Saunders, 1842)	KT151121
17	Punjab, India	*Bactrocera* (*Bactrocera*) *zonata* (Saunders, 1842)	MK564013
18	Iran	*Bactrocera* (*Bactrocera*) *zonata* (Saunders, 1842)	MG881697
19	Iran	*Bactrocera* (*Bactrocera*) *zonata* (Saunders, 1842)	MG881760
20	Srilanka	*Bactrocera* (*Bactrocera*) *correcta* (Bezzi, 1916)	MT257533
21	Cambodia	*Bactrocera* (*Bactrocera*) *correcta* (Bezzi, 1916)	MT257525
22	Vietnam	*Bactrocera* (*Bactrocera*) *correcta* (Bezzi, 1916)	MT257321
23	Assam, India	*Bactrocera* (*Zeugodacus*) *caudata* (Fabricius,	KM505013

UNDER PEER REVIEW

		1805)	
24	Thailand	*Bactrocera* (*Zeugodacus*) *caudata* (Fabricius, 1805)	JX559676
25	Thailand	*Bactrocera* (*Zeugodacus*) *caudata* (Fabricius, 1805)	JX297540
26	Meghalaya, India	*Bactrocera* (*Zeugodacus*) *cilifera* (Hendel, 1912)	MH395849
27	China	*Bactrocera* (*Zeugodacus*) *cilifera* (Hendel, 1912)	KF660025
28	Meghalaya, India	*Bactrocera* (*Zeugodacus*) *scutellaris* Bezzi, 1913	KT151118
29	Nepal	*Bactrocera* (*Zeugodacus*) *scutellaris* Bezzi, 1913	OP804513
30	China	*Bactrocera* (*Zeugodacus*) *scutellaris* Bezzi, 1913	KF660073

UNDER PEER REVIEW

2.2 *DNA extraction*

DNA was extracted utilizing the QIAmp DNA extraction kit (QIAGEN, Germany) from the crushed leg tissue of the fly sample. The whole procedure was done according to the manufacturer's protocol [39]. Every gDNA extraction batch had a blank extraction (one without a sample) to safeguard against cross-contamination during the DNA extraction process. Voucher specimens were submitted to the National Zoological Collection (NZC) of the Diptera Section in ZSI, Kolkata, India. The quantity of DNA was estimated using a Qubit Fluorometer (Life Technologies, USA), and the extracted DNA was stored at -20°C for further examination [39].

2.3 *Polymerase Chain Reaction (PCR) and sequencing of COI gene*

Amplification of about 700 base pairs from the 5' end of the mitochondrial COI gene was performed using roughly 20 ng of genomic DNA employing primers forward LCO-1490 (F) (GGT CAA CAA ATC ATA AAG ATA TTG G) and reverse HCO-2198 (R) (TAA ACT TCA GGG TGA CCA AAA AAT CA) [34]. PCR was carried out in a 50µl total reaction volume comprising 20 Pico moles of each primer, 100 mM KCl, 20 mM Tris–HCl (pH 8.0), 1 mM DTT, 0.1 mM EDTA, 2.0 mM $MgCl_2$, 0.25 mM of each dNTP, primer cocktail, and 1U of Taq polymerase (Takara BIO Inc., Japan) with the following cycling parameters: 5 min at 94◦C; followed by 40 cycles of 30 s at 94◦C, 40 s at 53◦C, 1 min at 72◦C and final extension for 5 min at 72◦C [35]. To verify the size of the amplicon, the products that were amplified were observed on a 1% agarose gel, dyed with SYBR@safe DNA gel dye, and captured on an Invitrogen-safe gel imager. The PCR-amplified products were purified with the QIAquick Gel Extraction Kit (Qiagen, Germany) following the manufacturer's instructions. About 15 ng of the purified PCR products were employed for cycle sequencing. Using both forward and reverse PCR primers, the BigDye®Terminator ver. 3.1Cycle Sequencing Kit (Applied Biosystems, Inc.) was utilized to perform cycle sequencing on an ABI thermal cycler with the following parameters: 96°C for 1 min, then 25 cycles of 96°C for 10 s, 50°C for 5 s, and final extension at 60°C for 1 min 15 s. After cycle sequencing, the products were cleaned with the BigDye X terminator kit (Applied Biosystems Inc.) and placed into an ABI 3730 capillary Genetic analyzer at the Zoological Survey of India sequencing laboratory [36, 35, 37].

2.4 *Sequence analyses and dataset formation*

The forward and reverse chromatogram analysis was performed using MEGAX software. The sequence editing data was obtained by trimming both ends to eliminate any ambiguous bases and noisy portions. This was followed by comprehensive annotation based on the forward and reverse sequences of every specimen [39]. MEGA X was used to manually edit the sequences from each specimen [40]. All sequences were matched to identical reported sequences in the NCBI database utilizing the BLAST (Basic Local Alignment Search Tool) (https://blast.ncbi.nlm.nih.gov) algorithm [38] and the ORF finder of NCBI (https://www.ncbi.nlm.nih.gov/orffinder/) [8] was used to study the accurate amino acid codes devoid of any stop codon or indels (insertions and deletions). Using the similarity search result performed through the NCBI BLASTn [39] and the BOLD (Barcode of Life Data system) search IDs, a dataset is formed using 43 sequences (excluding out-group), some of which were extracted from the GenBank database. Each of our sequences was uploaded to the GenBank library, and unique accession numbers were issued to each one (Table 1). For comparability, other COI sequences obtainable for the current species worldwide were acquired from GenBank (Table 2). The out-group was used as *Drosophila melanogaster* (KC750827, USA) a Drosophilidae, which was taken from the GenBank database also. The ClustalW algorithm program of MEGAX software aligned the dataset at the beginning [40]. To avoid any form of incongruent outcomes, the dataset was made to be 574 base pairs long [41].

2.5 Sequence divergence and phylogenetic analysis

Applying the Kimura-2-Parameter (K2P) in MEGAX, the genetic divergences between and among species were calculated. JModelTest v2.1.10 was used to identify the nucleotide substitution model that suited the data the best [42] through the CIPRES server [43] and concerning the lowest AIC (Akaike Information Criterion) score of -3185.2058 [36]. The General Time Reversible Model across Lineages, in conjunction with Gamma (GTR+G) (NST=6), was determined to be the most appropriate model for nucleotide substitution. MEGA X was used to investigate nucleotide substitution rate and composition [40]. The transition and transversion rate ratios were calculated using MEGA X software and the Maximum Composite Likelihood (MCL) technique. To represent the divergence between the

UNDER PEER REVIEW

fruit fly species, phylogenetic trees were constructed via Neighbour-Joining (NJ), Maximum Likelihood (ML), and Bayesian Analysis (BA). The NJ tree was made in MEGAX, where the bootstrap consensus tree was inferred from 1000 replicates. For ML-tree, the dataset was designed and analyzed in IQ-TREE on XSEDE (2.1.2v) [44, 45] via the CIPRES website [43]. The FigTree v1.4.4 program (http://tree.bio.ed.ac.uk/software/fgtree/) was used to modify the produced files. As a result, the sequence divergence between the specimens was visually represented. The Bayesian Inference (BI)-tree was constructed in Mr. Bayes v3.2.7a with nst=6 for the GTR+G model and metropolis-coupled Markov Chain Monte Carlo (MCMC), which was run for 1,000,000 generations with 25% burn-in and trees saved every 100 generations, to test the reciprocal monophyletic criteria for species delimitation [46]. Posterior probability (PP) was utilized to evaluate branch support. The iTOL v6 online tool (https://itol.embl.de/) was utilized to form a tree from the produced files, facilitating the visual display [47]. DnaSP v5.10 [48] was the tool used for estimating the haplotype diversity, nucleotide diversity, and the number of haplotypes.

3. RESULTS AND DISCUSSION

Fruit fly species *B. B. zonata* was not reported from vegetables or crops (infested) and was caught with the help of Methyl Eugenol (ME). *Bactrocera* (*Bactrocera*) *dorsalis* was captured from infested guava and Indian jujube fruit. This report had similarities with the works of Manger et al. (2018). *Bactrocera* (*Zeugodacus*) *tau* and *B. Z. cucurbitae* were exclusively reported from infested cucurbits. *Bactrocera* (*Zeugodacus*) *cucurbitae* was also captured through Cue-Lure. *Bactrocera* (*Bactrocera*) *latifrons* were captured from crops of the Solanaceae family and Methyl Eugenol. *Bactrocera* (*Bactrocera*) *correcta* was captured from infested fruit and *B. B. caudata* was captured in Cue- Lure and infested vegetable. Our reports had similarities with the works of Manger et al. (2018) and Arya et al. (2022). The seven species of fruit flies, namely *B. Z. cucurbitae*, *B. B. correcta*, *B. B. latifrons*, *B. Z. tau*, *B. B. dorsalis*, *B. B. zonata*, and *B. Z. caudata*, which were captured from various fruit and vegetable crop fields were identified properly through DNA barcoding. The identification of various fruit fly species has proven to be challenging, even with the support of existing literature [52]. This is because certain species are cryptic, and the color morphology of different species varies [12, 52].

The aligned dataset included 574 base pairs from the COI gene of 7 species of fruit flies belonging to the genera *Bactrocera* Macquart, 1835, which are significant destructive insect pests. The present study showed that the sequences for the cytochrome oxidase I gene have 51 conserved sites, 125 variable or polymorphic sites, 2 variable singleton sites, and 123 parsimony informative sites, excluding insertions, deletion, and other gap areas [8]. Seven species of fruit flies were identified by the nucleotide basic local alignment search (BLASTn) analysis of the COI sequences generated in the study. These species include *B. B. correcta, B. Z. cucurbitae, B. B. latifrons, B. Z. tau, B. B. dorsalis, B. B. zonata,* and *B. Z. caudata*. The published sequences in the GenBank database of the National Centre for Biotechnology Information (NCBI) revealed 99–100% identity for every identified species.

The analysis of the COI gene fragment during the study demonstrated that the average percentage of each nucleotide was T = 34.9%, C = 19.1%, A = 28.9%, and G = 17.1%. This indicates that the content of A+T is significantly larger than that of G+C. The higher A + T content seen in insect mtDNA is the feature that is usually mentioned as unique [49]. The transition/transversion rate ratios were k1 = 7.94 (purines) and k2 = 7.513 (pyrimidines). The overall transition/transversion bias was R = 3.56, {where $R = [A*G*k1 + T*C*k2]/[(A+G) * (T+C)]$} as calculated by Maximum Composite Likelihood method in MEGA X [40, 50]. This analysis involved 43 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated

UNDER PEER REVIEW

(complete deletion option). This result shows that the rate of transitions was more than the rate of transversions. Nucleotide transitional substitution patterns ranged between 13.65-27.54, and transversional substitution patterns ranged between 1.72-3.67 (Table 3). Maximum composite likelihood (MCL) with the Kimura-2 parameter [51, 40] was used to estimate the genetic distance between and within species. All positions with gaps and missing data were removed from the dataset using the complete deletion option. Intraspecific divergences of values < 3.00% were observed in all species, with values ranging from 0.00% to 1.00% (Table 4). Intraspecific divergences of value 1% were shown by *B. Z. tau, B. B. correcta,* and *B. Z. cilifera*. All the other species showed 0% intraspecific distance. The range of interspecific divergences was 3.65% to 28.46% (Table 4). The maximum genetic divergence was seen between *B. B. correcta* and *B. B. dorsalis* (28.46%). The least genetic divergences were seen between *B. Z. cucurbitae* and *B. Z. tau* (3.65%). All other species showed proper interspecific genetic divergence values.

Table 3. Maximum Composite Likelihood Estimate of the Pattern of Nucleotide Substitution:

	A	T/U	C	G
A	-	*3.67*	*1.92*	**13.65**
T/U	*2.99*	-	**14.45**	*1.72*
C	*2.99*	**27.54**	-	*1.72*
G	**23.76**	*3.67*	*1.92*	-

Each entry shows the probability of substitution ® from one base (row) to another base (column). For simplicity, the sum of r values is made equal to 100. Rates of different transitional substitutions are shown in bold and those of transversional substitutions are shown in italics.

Table 4 The average interspecific genetic distances (K2P model) of the COI sequences of 84 tephritid fly specimens collected throughout West Bengal and the GenBank database.

Species	Intra Specific Distance	1	2	3	4	5	6	7	8	9	10
1. *B. Z. cucurbitae*	0.00										
2. *B. B.*	0.00	21.48									

UNDER PEER REVIEW

latifrons											
3. *B. Z. tau*	1	3.65	22.51								
4. *B. B. dorsalis*	0.00	15.92	13.05	15.40							
5. *B. Z. diversa*	0.00	13.82	20.16	14.65	19.90						
6. *B. B. zonata*	0.00	16.98	14.02	17.38	7.81	20.30					
7. *B. B. correcta*	1	28.20	27.61	25.14	28.46	28.31	26.34				
8. *B. Z. caudata*	0.00	12.02	19.23	13.25	18.23	7.03	17.95	24.80			
9. *B. Z. cilifera*	1	13.31	22.21	11.98	18.44	14.72	20.11	21.60	13.25		
10. *B. Z. scutellaris*	0.00	12.98	18.55	11.35	19.35	9.82	20.16	24.82	10.06	13.56	

UNDER PEER REVIEW

One critical aspect of DNA-based species identification methods, according to Hebert et al. (2003), is their ability to discriminate between intraspecific and interspecific changes. To ensure the reliability of the gene being used for the process of species-level identification, the interspecific genetic divergence and intraspecific genetic divergence for any particular insect sample should be ≥3% and ≤1%, respectively [54]. Because of its ability to distinguish between species with greater clarity and greater interspecific divergence [55], COI has emerged as a common barcode for insect identification [56, 57]. In our studied dataset, although >3% interspecific genetic diversity was seen between *B. Z. cucurbitae* and *B. Z. tau*, the value was very low (3.65%) with an intraspecific value of ≤1 [26]. Strikingly *B. B.*

correcta showed very high intraspecific values with *B. Z. cucurbitae* (28.20%), *B. B. latifrons* (27.61%), *B. Z. tau* (25.14%), *B. B. dorsalis* (28.46), *B. Z. diversa* (28.31%), *B. B. zonata* (26.34%), *B. Z. caudata* (24.80%), *B. Z. cilifera* (21.60%), and *B. Z. scutellaris* (24.82%). There were substantial barcode gaps between the different subgenera and remaining species that showed that they differed from one another because of enough COI genetic divergences. When compared to other arthropod species, insects generally have less genetic variation, and intraspecific divergence in all animal taxa hardly exceeds 2% [53]. The large barcoding gap in the present study implies that all species examined here, except *B. Z. cucurbitae* and *B. Z. tau*, exhibit sufficient COI sequence divergencies for their accurate species identification.

The reciprocal monophyly of the COI sequences derived from the previously morphologically examined species was confirmed by both Neighbour-Joining (NJ) (Fig. 2) and Maximum Likelihood (ML) (Fig. 3) analyses, with notable branch support for both techniques. Based on monophyletic separation, the Neighbour Joining (NJ) tree and the Maximum-Likelihood (ML) tree demonstrated congruency and successfully separated the seven tephritid species. At the species level, the high bootstrap values of nearly 100% for both NJ and ML trees in deep branches suggested support for monophyly in both methods. *Bactrocera* (*Zeugodacus*) *cucrbitae* showed high bootstrap support values of 63%- 83% (NJ) and 63%-100% (ML) with samples from Thailand, Australia, northeastern India (Meghalaya), and southern India (Tamil Nadu). *Bactrocera* (*Zeugodacus*) *tau* had low to moderate bootstrap support values of 40%-100% (ML) and 51% (NJ), with samples from Bangladesh and China. *Bactrocera* (*Zeugodacus*) *cilifera*, *B. Z. scutellaris*, and *B. Z. diversa* showed strong bootstrap values of 100% (ML) and 52%-92% (NJ) between samples of China, Nepal, Thailand, and northeastern India (Meghalaya) but did not consist any of our species. *Bactrocera* (*Bactrocera*) *correcta* also showed similar strong support values of 93%-100% (ML) and 64%-100% (NJ) with samples from Vietnam, Sri Lanka, and Cambodia. *Bactrocera* (*Bactrocera*) *latifrons* and *B. B. dorsalis* had similar strong support of 94%-100% (ML) and 65%-100% (ML) and 75%-81% (NJ) and 98% (NJ) respectively with inside and outside Indian samples. *Bactrocera* (*Bactrocera*) *zonata* and *B. B. caudata* both expressed moderate to high bootstrap supports of 42%-100% (ML) and 96%-99% (NJ), and 47%-100% (ML) and 60%-92% (NJ) with outside-India and within-India samples respectively. Two subgenera (Bactrocera and Zeugodacus) and seven of our species were differentiated properly except for *B. Z. caudata* in both the NJ and ML trees and *B. B. correcta* in the ML tree. In the Bayesian Inference (BI) tree (Fig. 4) analysis, all the branches showed well-supported monophyly. The posterior probability for most of the deep branches ranged from 0.8-1 in their respective clades, which indicated proper reciprocal monophyly. Only *B. B. correcta* and *B. B. caudata* showed 0.5-1 posterior probability values, which indicated weak to high monophyly. Further haplotype analysis demonstrated a significant level of haplotype diversity, with 21 haplotypes identified in 43 tephritid samples, including those retrieved from GenBank and our specimens (Table 5). COI gene revealed that *B. Z. caudata* had the highest number of haplotypes (4). The highest haplotype diversity value of 1.00 was shown by *B. Z. cilifera* and *B. Z. diversa,* and the highest nucleotide diversity value of 0.01163 was shown by *B. Z. cilifera*. *Bactrocera* (*Zeugodacus*) *caudata* and *B. Z. cilifera* showed the highest segregating sites of value 4.

Table 5 Haplotype diversity and number of mitochondrial haplotypes in the tephritid specimens.

Name of	Number of	Average number of	Number of	Haplotype Diversity	Nucleotide Diversity	Number of segregating

UNDER PEER REVIEW

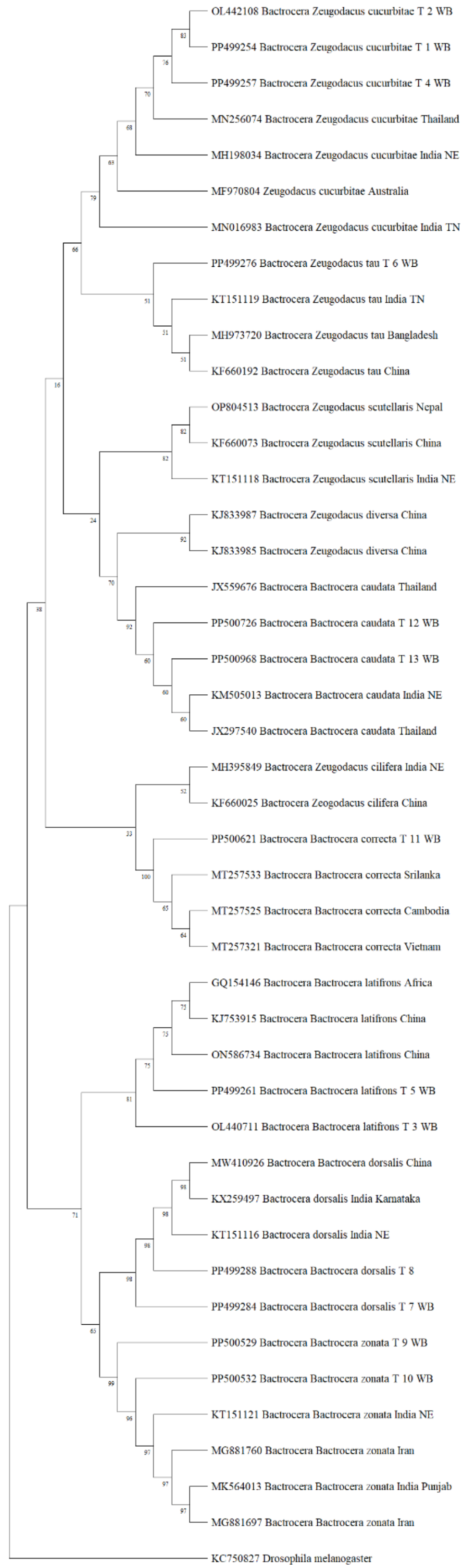
specimens	sequences	differences (K)	haplotypes	(H_d)	(P_i)	sites (S)
B. Z. cucurbitae	7	0.00	1	0.00	0.00	0
B. B. latifrons	5	0.40	2	0.40	0.00116	1
B. Z. tau	4	0.50	2	0.50	0.00145	1
B. B. dorsalis	5	0.00	1	0.00	0.00	0
B. Z. diversa	2	1.00	2	1.00	0.00291	1
B. B. zonata	6	0.3333	2	0.3333	0.00097	1
B. B. correcta	4	1.66667	3	0.8333	0.00484	3
B. Z. caudata	5	1.60	4	0.90	0.00465	4
B. Z. cilifera	2	4.00	2	1.00	0.01163	4
B. Z. scutellaris	3	1.33	2	0.66667	0.00388	2

UNDER PEER REVIEW

The three most efficient phylogenetic tree-building techniques (NJ, ML, and BI) were utilized to analyze our data and establish monophyly among the species. With strong branch support for every approach, NJ, ML, and BI analyses reinstate the reciprocal monophyly of the COI sequences derived from the previously morphologically investigated taxa. Our phylogenetic analyses are in agreement with various works done both in India [26, 28, 8] and worldwide [24, 25, 58, 59, 22, 18]. The evolutionary relationships of the seven tephritid species were established through the three analyses. In each analysis, the tree inferred proper distinct clades. The *Zeugodacus* subgenus was properly differentiated from the *Bactrocera* subgenus [24, 25, 26, 8]. However, in our study, *B. B. caudata* entered into the *Zeugodacus* clade in all three analyses (NJ, ML, and BI). Overall, COI highlights the significance of swift identification without the lengthy procedure of taxonomic identification of both adults and larvae. The high haplotype diversity from our study of the fruit fly species indicates that much work on characterizing these species with other genes is needed in future works.

Fig. 2 A Neighbour-joining (NJ) tree of the studied seven tephritid species with bootstrap supports. *Drosophila melanogaster* (KC750827) is used as an outgroup for the analysis.

UNDER PEER REVIEW



ER REVIEW

Fig. 3 The figure depicts the ML IQ-Tree based on COI sequences of 43 fruit flies and an outgroup of Drosophilidae, using 1000 bootstrap support. The branch lengths show the bootstrap values for the tree formation. The WB annotation denotes the 7 species of tephritid flies with one genus (*Bactrocera*), two subgenera, and 1 outgroup (Drosophilidae).

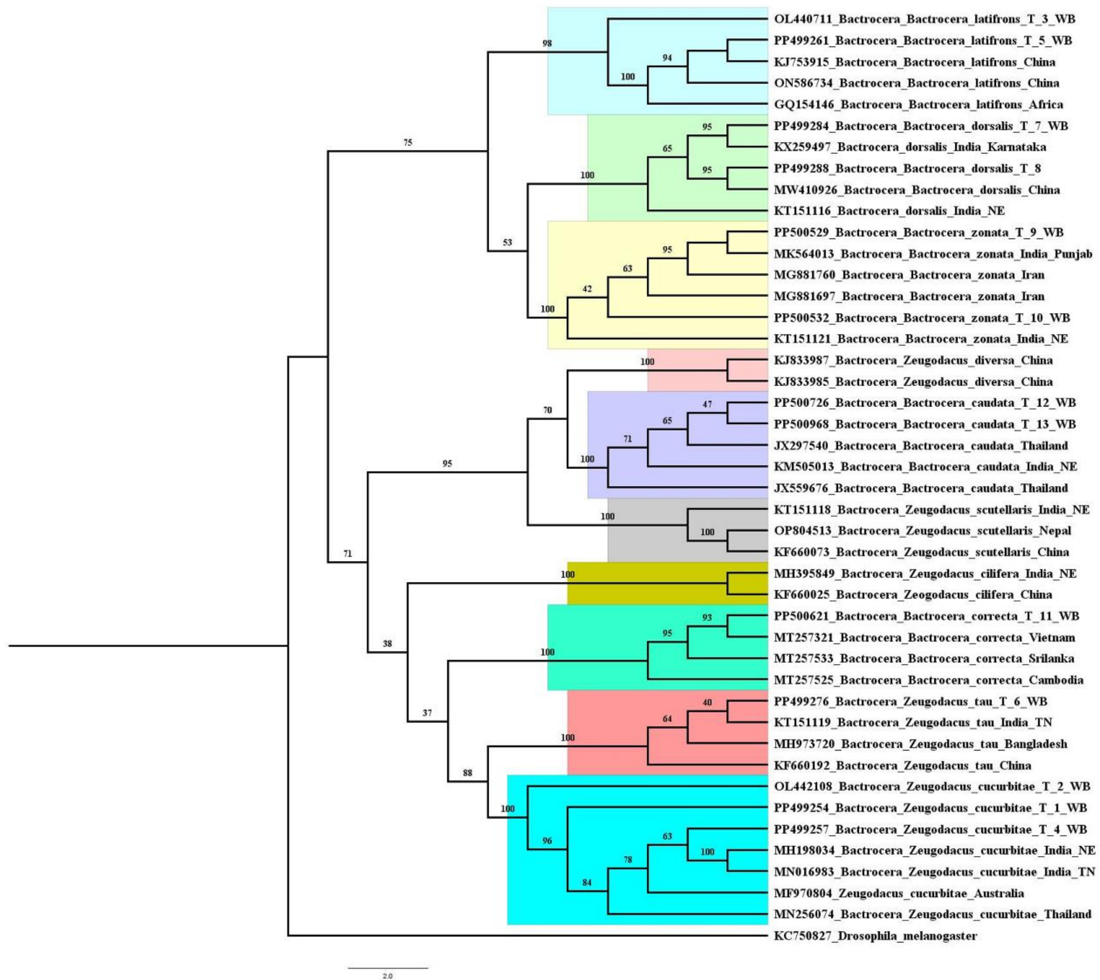
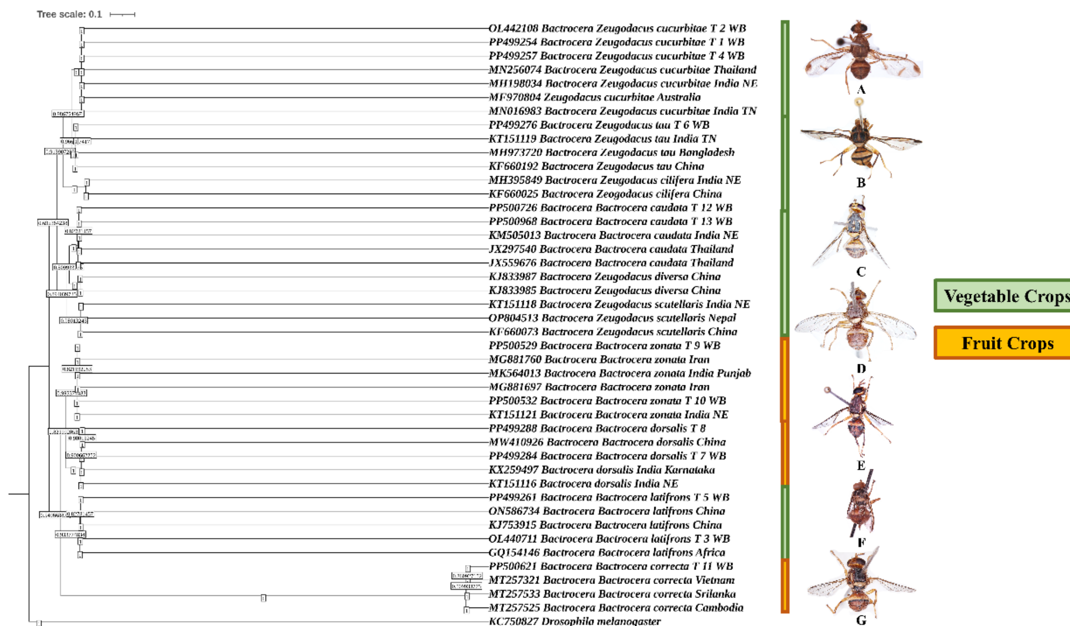


Fig. 4 The Bayesian (BA) Tree of 43 tephritid sequences with posterior probabilities from Bayesian analysis. Drosophilidae (*Drosophila melanogaster*) has been used as outgroup. The different species are shown in this figure namely- (A) *Bactrocera (Zeugodacus) cucurbitae* (Coquillett, 1899), (B) *Bactrocera (Zeugodacus) tau* (Walker, 1849), (C) *Bactrocera (Zeugodacus) caudata* (Fabricius, 1805), (D) *Bactrocera (Bactrocera) zonata* (Saunders, 1842), (E) *Bactrocera (Bactrocera) dorsalis* (Hendel, 1912), (F) *Bactrocera (Bactrocera) latifrons* (Hendel, 1915), and (G) *Bactrocera (Bactrocera) correcta* (Bezzi, 1916).



4. CONCLUSION

It is essential to identify and distinguish the economically significant tephritid flies to comprehend and regulate their impacts in the fields of agriculture and horticulture. The COI gene shows potential as a DNA-based identification approach for Indian fruit flies, especially from immature larval or adult specimens. Since several *Bactrocera* species are very invasive and polyphagous, they should be kept under quarantine in eastern India. This approach bypasses the limitations of conventional taxonomy for tephritid species and aids in identifying species complexes, including cryptic and visually similar species. More significantly, our study emphasizes how crucial it is to create a comprehensive barcode database for fruit fly species in GenBank from this area, as they act as pests for important horticulture crops.

This study is a breakthrough initiative in eastern India, offering thorough coverage of the Gangetic Plains in West Bengal, which is important for horticultural crops. Also, COI gene sequences from other parts of India and worldwide have been included in the study for better understanding. Further, our results emphasize the necessity of adding more genetic (nuclear and mitochondrial) markers to improve the robust phylogeny of fruit fly species and other genera.

REFERENCES

1. Barnes BN. Proceedings of the 6th International Symposium on fruit flies of economic importance, Stellenbosch, South Africa, 6-10 May 2002. 2004.

2. Ekesi S, Billah MK. A field guide to the management of economically important tephritid fruit flies in Africa. ICIPE Science Press, Nairobi ISBN. 2007;92:209 http://hdl.handle.net/123456789/661

3. Laskar N, Sinha DK, Hath TK, Chatterjee H. Diversity of Tephritid Flies in Sub-Himalayan Region of West Bengal: Baseline Data for Developing Rational Management Practices. In:

UNDER PEER REVIEW

Chakravarthy AK, Sridhara S (eds) Arthropod Diversity and Conservation in the Tropics and Sub-tropics. Springer Singapore. 2016;271–284 https://doi.org/10.1007/978-981-10-1518-2_16

4. Uchôa MA. Fruit flies (Diptera: Tephritoidea): biology, host plants, natural enemies, and the implications to their natural control. Integrated Pest Management and Pest Control: Current and Future Tactics InTech, Rijeka, Croatia. 2012;271–300.

5. David KJ, Ramani S. An illustrated key to fruit flies (Diptera: Tephritidae) from Peninsular India and the Andaman and Nicobar Islands. Zootaxa. 2011; 3021(1):1–31.

6. David KJ, Ramani S, Singh SK. Taxonomy and Diversity of Indian Fruit Flies (Diptera: Tephritidae). In: Indian Insects. CRC Press. 2019;305–324.

7. Papadopoulos NT. Fruit Fly Invasion: Historical, Biological, Economic Aspects and Management. In: Shelly T, Epsky N, Jang EB, et al (eds) Trapping and the Detection, Control, and Regulation of Tephritid Fruit Flies. Springer Netherlands, Dordrecht. 2014;219–252

8. Manger A, Behere GT, Firake DM, et al. Genetic characterization of Bactrocera fruit flies (Diptera: Tephritidae) from Northeastern India based on DNA barcodes. Mitochondrial DNA Part A. 2018;29(5):792–799. https://doi.org/10.1080/24701394.2017.1357713

9. Van Houdt JKJ, Breman FC, Virgilio M, De Meyer M. Recovering full DNA barcodes from natural history collections of Tephritid fruitflies (Tephritidae, Diptera) using mini barcodes. Molecular Ecology Resources. 2010.10(5):459–465. https://doi.org/10.1111/j.1755-0998.2009.02800.x

10. White IM, Elson-Harris MM. Fruit flies of economic significance: their identification and bionomics. 1992.

11. Drew RA. The tropical fruit flies (Diptera: Tephritidae: Dacinae) of the Australasian and Oceanian regions. Memoirs of the Queensland Museum. 1989;26.

12. Drew RA, Hancock DL. The *Bactrocera dorsalis* complex of fruit flies (Diptera: Tephritidae: Dacinae) in Asia. Bulletin of entomological research supplement series. 1994;2:1–68.

13. Jose MS, Leblanc L, Geib SM, Rubinoff D. An evaluation of the species status of Bactrocera invadens and the systematics of the Bactrocera dorsalis (Diptera: Tephritidae) complex. Annals of the Entomological Society of America. 2013;106(6):684–694. https://doi.org/10.1603/AN13017

14. Drew RA, Hancock DL. Phylogeny of the tribe Dacini (Dacinae) based on morphological, distributional, and biological data. In: Fruit Flies (Tephritidae). Crc Press. 1999;509–522

15. Clarke AR, Armstrong KF, Carmichael AE, et al. INVASIVE PHYTOPHAGOUS PESTS ARISING THROUGH A RECENT TROPICAL EVOLUTIONARY RADIATION: The *Bactrocera dorsalis* Complex of Fruit Flies. ANNUAL REVIEW OF ENTOMOLOGY. 2005;50:293–319. https://doi.org/10.1146/annurev.ento.50.071803.130428

UNDER PEER REVIEW

16. Verghese A, Madhura HS, Kamala Jayanthi PD, Stonehouse JM. Fruit flies of economic significance in India, with special reference to *Bactrocera dorsalis* (Hendel). In: Proceedings of 6th International Fruit fly Symposium. 2002;6–10.

17. Dhillon MK, Singh R, Naresh JS, Sharma HC. The melon fruit fly, *Bactrocera cucurbitae*: A review of its biology and management. Journal of Insect Science. 2005;5(1):40. https://doi.org/10.1093/jis/5.1.40

18. Kunprom C, Pramual P. DNA barcoding of fruit flies (Diptera: Tephritidae) in Thailand: ambiguity, misidentification and cryptic diversity. Mitochondrial DNA Part A. 2019;30:861–873. https://doi.org/10.1080/24701394.2019.1693550

19. Hebert PD, Cywinska A, Ball SL, DeWaard JR. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London Series B: Biological Sciences. 2003;270:313–321. https://doi.org/10.1098/rspb.2002.2218

20. Barr NB, Islam MS, De Meyer M, McPheron BA. Molecular identification of *Ceratitis capitata* (Diptera: Tephritidae) using DNA sequences of the COI barcode region. Annals of the Entomological Society of America. 2012;105(2):339–350. https://doi.org/10.1603/AN11100

21. Jiang F, Jin Q, Liang L, et al. Existence of species complex largely reduced barcoding success for invasive species of T ephritidae: a case study in B actrocera spp. Molecular Ecology Resources. 2014;14(6):1114–1128. https://doi.org/10.1111/1755-0998.12259

22. Kunprom C, Pramual P. DNA barcode variability and host plant usage of fruit flies (Diptera: Tephritidae) in Thailand. Genome. 2016;59(10):792–804. https://doi.org/10.1139/

23. Asokan R, Rebijith KB, Singh SK, et al. Molecular identification and phylogeny of Bactrocera species (Diptera: Tephritidae). Florida Entomologist. 2011;1026–1035.

24. Afroz S, Nomana S, Zhang Y, et al. Species identification of economic important adult fruit flies based on DNA Barcoding (MT DNA COI) and larvae based on species specific primers from Central and South part of Bangladesh. Malaysian Journal of Sustainable Agriculture. 2022;6:103–109. http://doi.org/10.26480/mjsa.02.2022.103.109

25. Ahmad JN, Sharif T, Ahmad SJ, et al. Molecular identification and characterization of fruit flies of genus Bactrocera (Diptera: Tephritidae) in Pakistan. Pakistan Journal of Zoology. 2019;51(9):2275. http://dx.doi.org/10.17582/journal.pjz/2019.51.6.2275.2280

26. Arya V, Narayana S, Tyagi S, et al. DNA barcoding of fruit flies associated with cucurbit ecosystem and combination of Cue-Lure and Methyl Eugenol in trap is not effective for mass trapping of responsive fruit flies. Phytoparasitica. 2022;50:683–695. https://doi.org/10.1007/s12600-022-01003-4

27. Blacket MJ, Semeraro L, Malipatil MB. Barcoding Queensland Fruit Flies (*Bactrocera tryoni*): impediments and improvements. Molecular Ecology Resources. 2012;12(3):428–436. https://doi.org/10.1111/j.1755-0998.2012.03124.x

28. Choudhary JS, Naaz N, Lemtur M, et al. Genetic analysis of *Bactrocera zonata* (Diptera: Tephritidae) populations from India based on cox1 and nad1 gene sequences. Mitochondrial DNA Part A. 2018;29(5):727–736. https://doi.org/10.1080/24701394.2017.1350952

UNDER PEER REVIEW

29. Virgilio M, Jordaens K, Breman FC, et al. Identifying insects with incomplete DNA barcode libraries, African fruit flies (Diptera: Tephritidae) as a test case. PloS one. 2012;7(2):e31581. https://doi.org/10.1371/journal.pone.0031581

30. Nair N, Bhattacharjee T, Thangjam B, et al. Species diversity of Dacine fruit flies (Diptera: Tephritidae: Dacinae: Dacini) in Tripura, NE India. Journal of Entomology and Zoology Studies. 2018;6(1):297–302.

31. Prabhakar CS, Sood P, Mehta PK. Pictorial keys for predominant *Bactrocera* and *Dacus* fruit flies (Diptera: Tephritidae) of north western Himalaya. Arthropods. 2012;1(3):101.

32. Evenhuis NL, Pape T. Systema Dipterorum, Version (5.0). http://www.diptera.org/. Accessed 15 Feb 2024. 2024.

33. Ghosh D, Kar O, Pramanik D, et al. Molecular identification and characterization of Muscid flies (Diptera: Muscidae) of medico-veterinary importance from the Gangetic plains of Eastern India. International Journal of Tropical Insect Science. 2022;42(6):3759-3769. https://doi.org/10.1007/s42690-022-00900-9

34. Folmer O, Black M, Hoeh W, et al. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology. 1994;3(5):294–299.

35. Banerjee D, Kumar V, Maity A, et al. Identification through DNA barcoding of Tabanidae (Diptera) vectors of surra disease in India. Acta Tropica. 2015;150:52–58. https://doi.org/10.1016/j.actatropica.2015.06.023

36. Achint R, Singh D. Application of COI gene for identification of some economically and forensically important muscid flies of India (Diptera: Muscidae). International Journal of Tropical Insect Science. 2021;41:3023–3029. https://doi.org/10.1007/s42690-021-00494-8

37. Tyagi K, Kumar V, Singha D, et al. DNA Barcoding studies on Thrips in India: Cryptic species and Species complexes. Scientific Reports. 2017;7:1–14. https://doi.org/10.1038/s41598-017-05112-7

38. Chakraborty R, Singha D, Kumar V, et al. DNA barcoding of selected Scirtothrips species (Thysanoptera) from India. Mitochondrial DNA Part B. 2019;4:2710–2714. https://doi.org/10.1038/s41598-017-05112-7

39. Johnson M, Zaretskaya I, Raytselis Y, et al. NCBI BLAST: a better web interface. Nucleic acids research. 2008;36(Issue suppl_2, 1):W5–W9. https://doi.org/10.1093/nar/gkn201

40. Kumar S, Stecher G, Li M, et al. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Molecular biology and evolution. 2018;35(6):1547-1549. https://doi.org/10.1093/molbev/msy096

41. Kundu S, Chandra K, Tyagi K, et al. DNA barcoding of freshwater fishes from Brahmaputra River in Eastern Himalaya biodiversity hotspot. Mitochondrial DNA Part B. 2019;4(2):2411–2419. https://doi.org/10.1080/23802359.2019.1637290

42. Darriba D, Posada D. jModelTest 2.0 Manual. Nature Methods. 2012;9:772. https://www.phylo.org/pdf_docs/jmodeltest-2.1.6-manual.pdf

UNDER PEER REVIEW

43. Miller MA, Pfeiffer W, Schwartz T. The CIPRES science gateway: enabling high-impact science for phylogenetics researchers with limited resources. In: Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the extreme to the campus and beyond. 2012;1–8. https://doi.org/10.1145/2335755.2335836

44. Minh BQ, Schmidt HA, Chernomor O, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Molecular biology and evolution. 2020;37(5):1530–1534. https://doi.org/10.1093/molbev/msaa015

45. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution. 2015;32(1):268–274. https://doi.org/10.1093/molbev/msu300

46. Ronquist F, Teslenko M, Van Der Mark P, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic biology. 2012;1(3):539–542. https://doi.org/10.1093/sysbio/sys029

47. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic acids research. 2021;49(W1):W293–W296. https://doi.org/10.1093/nar/gkab301

48. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25:1451–1452. https://doi.org/10.1093/bioinformatics/btp187

49. Lunt DH, Hyman BC. Animal mitochondrial DNA recombination. Nature. 1997;387:247–247. https://doi.org/10.1038/387247a0

50. Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighborjoining method. Proceedings of the National Academy of Sciences. 2004;101(30):11030–11035. https://doi.org/10.1073/pnas.040420610

51. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution. 1980;16:111–120. https://doi.org/10.1007/BF01731581

52. Schutze MK, Aketarawong N, Amornsak W, et al. 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. Systematic Entomology. 2015;40:456–471. https://doi.org/10.1111/syen.12113

53. Hebert PD, Ratnasingham S, De Waard JR. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proceedings of the Royal Society of London Series B: Biological Sciences. 2003;270(Issue Suppl_1):S96–S99. https://doi.org/10.1098/rsbl.2003.0025

54. Wells JD, Sperling FA. DNA-based identification of forensically important Chrysomyinae (Diptera: Calliphoridae). Forensic Science International. 2001;120(1-2):110–115. https://doi.org/10.1016/S0379-0738(01)00414-5

UNDER PEER REVIEW

55. Pentinsaari M, Hebert PD, Mutanen M. Barcoding beetles: a regional survey of 1872 species reveals high identification success and unusually deep interspecific divergences. PLoS One. 2014;9(9):e108651. https://doi.org/10.1371/journal.pone.0108651

56. Lin X, Stur E, Ekrem T. Exploring genetic divergence in a species-rich insect genus using 2790 DNA barcodes. PloS one. 2015;10(9):e0138993. https://doi.org/10.1371/journal.pone.0138993

57. Twinkle T, Shashank PR, Chattopadhyay PC. DNA barcoding and Taxonomic account on some selected species of subfamily Plusiinae (Lepidoptera: Noctuidae) from India. Zootaxa. 2020;4845(4):zootaxa-4845. https://doi.org/10.11646/zootaxa.4845.4.1

58. Doorenweerd C, San Jose M, Barr N, et al. Highly variable COI haplotype diversity between three species of invasive pest fruit fly reflects remarkably incongruent demographic histories. Scientific Reports. 2020;10(1):6887. https://doi.org/10.1038/s41598-020-63973-x

59. Karimi J, Darsouei R. Presence of the endosymbiont Wolbachia among some fruit flies (Diptera: Tephritidae) from Iran: A multilocus sequence typing approach. Journal of Asia-Pacific Entomology. 2014;17(1):105–112. https://doi.org/10.1016/j.aspen.2013.11.002

UNDER PEER REVIEW