Evaluation of identification methods for cryptic Bactrocera dorsalis (Diptera: Tephritidae) specimens: combining morphological and molecular techniques
Emeline Charbonnel, Marie-Pierre Chapuis, Andrea Taddei, Mark K Schutze, Melissa L Starkie, Laure Benoit, Raphaëlle Mouttet, David Ouvrard

To cite this version:

HAL Id: hal-04271012
https://hal.inrae.fr/hal-04271012v2
Submitted on 2 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives| 4.0 International License
Horticultural Entomology

Evaluation of identification methods for cryptic Bactrocera dorsalis (Diptera: Tephritidae) specimens: combining morphological and molecular techniques

Emeline Charbonnel\textsuperscript{1,2,3}, Marie-Pierre Chapuis\textsuperscript{1,2,}, Andrea Taddei\textsuperscript{3,4}, Mark K. Schutze\textsuperscript{4,}, Melissa L. Starkie\textsuperscript{4,}, Laure Benoit\textsuperscript{1,2,}, Raphaëlle Mouttet\textsuperscript{3,4}, David Ouvrard\textsuperscript{3,4}

\textsuperscript{1}CBGP, CIRAD, INRAE, Institut Agro, IRD, Univ Montpellier, Montpellier, France, \textsuperscript{2}CIRAD, CBGP, F 34398, Montpellier, France, \textsuperscript{3}ANSES, Plant Health Laboratory, Entomology and Botany Unit, Montferrier-sur-Lez, France, \textsuperscript{4}Queensland Department of Agriculture and Fisheries, Brisbane, QLD, Australia

Corresponding author, mail: david.ouvrard@anses.fr

Subject Editor: Kent Shelby

Received on 5 July 2023; revised on 5 July 2023; accepted on 8 September 2023

The potential for population genomics to elucidate invasion pathways of a species is limited by taxonomic identification issues. The Oriental fruit fly pest, Bactrocera dorsalis (Hendel) belongs to a complex in which several sympatric species are attracted to the same lure used in trapping and are morphologically cryptic and/or reported to hybridize. In this study, we evaluated the taxonomic ambiguity between B. dorsalis and 2 major cryptic species, based on morphological expertise and 289 target specimens sampled across the whole distribution range. Specimens were then subjected to DNA sequence analyses of the COI mitochondrial barcode and the EIF3L nuclear marker to evaluate the potential for molecular identification, in particular for specimens for which morphological identification was inconclusive. To this aim, we produced reference datasets with DNA sequences from target specimens whose morphological identification was unambiguous, which we complemented with 56 new DNA sequences from closest relatives and 76 published and curated DNA sequences of different species in the complex. After the necessary morphological observation, about 3.5% of the target dataset and 47.6% of the specimens from Southeast Asian islands displayed ambiguous character states shared with B. carambolae and/or B. occipitalis. Critical interpretation of DNA sequence data solved morphological ambiguities only when combining both mitochondrial and nuclear markers. COI discriminated B. dorsalis from 5 species; EIF3L and ITS from another species. We recommend this procedure to ensure correct identification of B. dorsalis specimens in population genetics studies and surveillance programs.

\textbf{Key words:} mtDNA, nDNA, haplotype diversity, ambiguous identification

Introduction

Over the last 2 decades, the Oriental fruit fly, Bactrocera dorsalis (Hendel) (Diptera: Tephritidae), has emerged as one of the most invasive and destructive insect pests of tropical and subtropical fruits and vegetables, particularly in sub-Saharan Africa and the Indian Ocean (Drew and White 2005, Schutze et al. 2015). Bactrocera dorsalis is highly polyphagous, with more than 300 host plants (Allwood et al. 1999, Clarke et al. 2005, Ekesi and Billah 2007), although its major host is mango (Mangifera indica) (Sileshi et al. 2019, Ndela et al. 2022). Population genetics has been, and increasingly is, a powerful tool for providing new insights into its invasion biology, in particular introduction sources, invasion success, and colonization dynamics (Aketerawong et al. 2007, Khamis et al. 2009, Wan et al. 2011, Li et al. 2012, Shi et al. 2012, Choudhary et al. 2016, Kim et al. 2021). Most previous studies omitted to detail the method used for specimen identification, despite a clearly articulated need to do so as outlined by Schutze et al. (2017) advocating an integrative taxonomic approach to tephritid fruit fly species delimitation and diagnoses (but see Choudhary et al. 2016, Qin et al. 2018, Zhang et al. 2022). There is a risk of confusion with several sympatric and morphologically cryptic species, since B. dorsalis belongs to a complex of about 90 species including roughly 30 species attracted to methyl-eugenol (ME), a male-targeted chemical lure used for sampling in population genetics studies and surveillance programs. The undetected presence of a species closely related to B. dorsalis in population genetics analyses skews estimations of genetic structure and
diversity, and thereby compromises the understanding of underlying processes or induces discrepancies between studies.

*Bactrocera dorsalis* identification based on morphological characters alone is sometimes difficult because its intraspecific variability overlaps with other members of the complex, especially *B. carambolae* Drew & Hancock and *B. occipitalis* (Bezzi) (Leblanc et al. 2015, Pieterse et al. 2017, Taddei et al. 2023). Moreover, cases of hybridization have been reported between *B. dorsalis* and *B. carambolae*, *B. kandiensis* Drew & Hancock, *B. occipitalis*, and *B. raiensis* Drew & Hancock (Wee and Tan 2005, Schutze et al. 2013, Schutze, Mahmood et al. 2015, Nuges et al. 2018, San Jose et al. 2018, Zhang et al. 2022, Doorenweerd et al. 2020). Since accurate identification is crucial for surveillance of *B. dorsalis*, standard diagnostic protocols (e.g., EPPO 2021) incorporated a DNA barcode approach, often based on the single mitochondrial Cytochrome c Oxidase subunit I (COI), which has become a widely used identification tool (Hebert et al. 2003). However, the ability of the mtDNA barcode to distinguish *B. dorsalis* from other species within the complex remains controversial and its issues relating to tephritid diagnoses are well documented (Krosch et al. 2019, 2020), some studies indicating a substantial overlap in intra- and interspecific sequence variation for many species in the complex (e.g., Jiang et al. 2014 and references within; San Jose et al. 2018, Doorenweerd et al. 2020). The contradictory and unclear results may be explained by differences in taxonomic coverage and intraspecific sampling effort, in particular of the target species, and by identification errors, in particular when DNA entries from public databases are used (e.g., Asadi et al. 2019).

As a consequence, *B. dorsalis* identifications could possibly be validated by combining the mtDNA barcode with independent nDNA markers. Finding appropriate nuclear barcodes is challenging, however, since nuclear genes often lack sufficient species-level diagnostic variability. In this context, IPPC (2019) recommends using the nuclear Internal Transcribed Spacer 1 (ITS1) (Boykin et al. 2014), which discriminates *B. dorsalis* from *B. carambolae* (only) via length polymorphism. Furthermore, Plant Health Australia (2020) provides reference sequences for 4 nuclear markers, including the Eukaryotic translation Initiation Factor 3 subunit L gene (EIF3L), developed from anchored hybrid enrichment research and which seems the most promising for discriminating the species closest to

![Fig. 1. Distribution maps and sampling points for *B. dorsalis* (A), *B. carambolae* and *B. raiensis* (B), and *B. kandiensis* and *B. occipitalis* (C). The maps are adapted from the April 2023 EPPO database, completed with the addition of India for *B. kandiensis*. The size of dots is proportional to the number of samples collected in each country. (B) Sampling dot in Thailand is for *B. raiensis*. Note that the Plant Health Australia collection samples are not represented here because no locality information was provided with the reference sequences.](https://academic.oup.com/jee/article/116/6/2193/7308789)
Table 1: Final identification for the ten specimens with ambiguous morphological characters, with pieces of information from morphology and each molecular marker. Six lesser-known species not attracted by methyl-eugenol are marked with an asterisk.

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Ambiguous morphological characters</th>
<th>Possible species confusion</th>
<th>COI identification</th>
<th>EIF3L identification</th>
<th>ITS1 identification</th>
<th>Final identification</th>
</tr>
</thead>
</table>
| IDN_322   | - Costal band confluent with R2 + 3 vein, very broad around apex of wing.  
- Thin “T” pattern on abdomen.  
- Rectangular anterolateral black markings on tergite 4. | B. dorsalis, B. carambolae, B. propinquus, B. irvingiae* | B. carambolae, B. dorsalis, B. kandiensis, B. raensis | NA | Not B. carambolae (~500bp) | B. dorsalis |
| IDN_323   | - Costal band confluent with R2 + 3 vein, very broad around apex of wing. | B. carambolae | B. carambolae | B. carambolae | B. carambolae | B. carambolae |
| IDN_331   | - Costal band slightly overlapping R2 + 3 vein. | B. dorsalis, B. carambolae, B. propinquus* | B. carambolae, B. dorsalis, B. kandiensis, B. raensis | B. carambolae | Not B. carambolae (~500bp) | B. dorsalis |
| IDN_373   | - Intermediate costal band between confluent and overlapping with R2 + 3 vein.  
- Rectangular anterolateral black markings on tergite 4.  
- Fulvous femora and tibia, with dark spot on fore femora and dark half of the tibia of the mid leg. | B. dorsalis, B. carambolae, B. propinquus* | B. carambolae, B. dorsalis, B. kandiensis, B. raensis | B. carambolae, B. endiandrae, B. kandiensis, B. dorsalis, B. latineola, B. occipitalis, B. opiliae, B. raensis | Not B. carambolae (~500bp) | B. dorsalis, B. propinquus* |
| IDN_376   | - Costal band slightly overlapping R2 + 3 vein.  
- Dark spot on the fore femora.  
- Large “T” pattern on abdomen. | B. dorsalis, B. carambolae, B. propinquus* | B. carambolae, B. dorsalis, B. kandiensis, B. raensis | B. carambolae, B. endiandrae, B. kandiensis, B. dorsalis, B. latineola, B. occipitalis, B. opiliae, B. raensis | Not B. carambolae (~500bp) | B. dorsalis, B. propinquus* |
| MYS_352   | - Costal band confluent with R2 + 3 vein, broad around apex of wing.  
- Rectangular anterolateral black markings on tergite 4. | B. dorsalis, B. carambolae | B. carambolae, B. dorsalis, B. kandiensis, B. raensis | B. carambolae | B. carambolae | B. carambolae |
| MYS_354   | - Costal band confluent with R2 + 3 vein, broad around apex of wing.  
- Rectangular anterolateral black markings on tergite 4. | B. carambolae | B. carambolae | B. carambolae | B. carambolae | B. carambolae |
| MYS_356   | - Costal band confluent with R2 + 3 vein, broad around apex of wing.  
- Rectangular anterolateral black markings on tergite 4. | B. carambolae | B. carambolae | B. carambolae | B. carambolae | B. carambolae |
Table 1. Continued

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Final identification</th>
<th>Ambiguous morphological characters</th>
<th>Possible species confusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHL_268</td>
<td>Not B. carambolae</td>
<td>- Not easy to tell if there are rectangular anterolateral black markings on tergite 4.</td>
<td>B. philippinensis, B. dorsalis</td>
</tr>
<tr>
<td></td>
<td>B. dorsalis</td>
<td>- Fulvous femora.</td>
<td>B. invadens, B. papayae, B. philippinensis</td>
</tr>
<tr>
<td>PHL_269</td>
<td>Not B. carambolae</td>
<td>- Shining spots on abdomen neither light nor dark.</td>
<td>B. invadens, B. papayae, B. philippinensis</td>
</tr>
</tbody>
</table>

Materials and Methods

We sampled 289 adult fruit fly specimens across the whole distribution range of B. dorsalis (Fig. 1, Supplementary Table S1): 213 specimens from the collection of the French Agency for Food, Environmental and Occupational Health & Safety, Plant Health Laboratory (ANSES – LSV), reared from larvae intercepted at European borders on infested imported fruits, and 76 specimens collected worldwide in orchards using ME traps, including 3 specimens from the Queensland Primary Industries Insect Collection (QDPC). Specimens were identified morphologically using IPPC (2019), completed with Drew and Romig (2013, 2016) and White and Elson-Harris (1992), by at least 1 expert in taxonomic entomology and/or person trained in morphological techniques. Following Schütze, Aketarawong et al. (2015), we considered B. invadens, B. papayae, and B. philippinensis as synonyms of B. dorsalis.

We used nondestructive DNA extraction protocols on whole specimens to conserve morphological characters, as detailed in Supplementary Table S1. The 658 bp 5' standard “barcoding region” of the COI gene (Hebert et al. 2003) and the 581 bp of the EIF3L gene (Plant Health Australia 2020) were amplified using polymerase chain reactions (PCRs) performed as detailed in Supplementary Table S2 and sequenced by Eurofins Genomics. All nucleotide sequences were checked for ambiguous bases, the presence of frame shift mutations and stop codons, edited using CodonCode Aligner v.10.0.2 and aligned using Seaview v.5.0.5 (Gouy et al. 2021). PCR products of the ITS1 marker (Boykin et al. 2014), obtained as detailed in Supplementary Table S2, were run in 2.5% electrophoresis agarose gel for 90 min to double blind read the fragment size, expected to be 544 bp for B. carambolae and 300 bp for other B. dorsalis complex species.

For specimens for which morphological identification was inconclusive, molecular identification was achieved by visualizing the position of their COI and EIF3L sequences in tree graphics relatively to curated sequences of species of the B. dorsalis complex. The reference datasets included 406 (COI) and 309 (EIF3L) sequences from target specimens with unambiguous morphological identifications, specimens of the B. dorsalis species complex published by Plant Health Australia (2020) (Supplementary Table S1), and additional specimens from species morphologically confusable and/or with reported hybridizations with B. dorsalis (see Fig. 1, Supplementary Tables S1 and S3 for sample size details). Maximum likelihood (ML) trees were constructed on unique haplotypes, recovered with the sidier package (Muñoz-Pajares 2013), using MEGA v.X (Kumar et al. 2018) and the T92+G mutational model (Tamura 1992). We performed 500 bootstrap replications and condensed the ML trees to show clades supported by bootstrap values over 50%. For the
haploid COI barcode, we also constructed a minimum spanning network among haplotypes using Popart v.1.7 (Leigh and Bryant 2015).

**Results and Discussion**

Based on morphological keys, we confidently identified 279 specimens as *B. dorsalis* out of the 289 studied. The 10 remaining specimens, ~3.5% of the target dataset, possessed character states intermediate between *B. dorsalis* and other species of the complex, including *B. carambolae*, *B. occipitalis* and several little-known species rarely cited in literature and indifferent to ME (but attracted to Cue-lure, except *B. irvingiae* Drew & Hancock, for which no lure is known), preventing conclusive morphological identification (Table 1). All these specimens originated from either Indonesia, Malaysia, or the Philippines, where the rate of ambiguous individuals reaches 47.6%. The characters that diverged from the conventional morphological description of *B. dorsalis* were mainly the extent of the wing costal band and apex spot (Fig. 2G and H) and the abdominal color pattern (Fig. 2I and J).

COI barcoding confidently discriminated *B. dorsalis* from *B. cacuminata* (Hering), *B. endiandrae* (Perkins & May), *B. latilineola* Drew, *B. occipitalis*, and *B. opiliae* (Drew & Hardy) (Fig. 3A) whereas *B. carambolae* and *B. raiensis* specimens branched individually among *B. dorsalis* individuals in the ML tree without clustering according to species. The minimum spanning network of the mitochondrial haplotypes from *B. dorsalis* and its closest relatives was largely congruent with the ML tree: while the most frequently observed haplotypes clearly segregated according to species, the star-like shapes of the *B. dorsalis*, *B. carambolae*, and *B. raiensis* haplogroups overlapped to the point of preventing species resolution (Supplementary Fig. S1).

In addition, 7 specimens from India, Sri Lanka, Cameroon and the Congo that were confidently identified morphologically and with the EIF3L marker (see Fig. 3B and text below) as *B. dorsalis* clustered in the ML tree with *B. kandiensis* reference mitochondrial sequences (Fig. 3A). These represented 2.4% of the target specimens, 23.5% of which originated from the *B. kandiensis* range, and 2.3% of those from the invaded African continent. The minimum spanning network showed not only these *B. dorsalis* haplotypes in the *B. kandiensis* haplogroup, but also 1 distant haplotype in the *B. carambolae* haplogroup, and *B. raiensis* haplotypes in the *B. dorsalis* haplogroup (Supplementary Fig. S1). This supports previous reports of mitochondrial introgression events between *B. dorsalis* and *B. carambolae*, *B. kandiensis*, and *B. raiensis*.

Regarding the EIF3L marker, 34 specimens out of 272 (12.5%) failed the PCR amplification step due to poor DNA quality (Supplementary Table S1). Among those that worked, EIF3L confidently discriminated *B. carambolae* and *B. endiandrae* from *B. dorsalis* (Fig. 3B). However, *B. cacuminata*, *B. kandiensis* (1 specimen, *Bactrocera kandiensis_LKA_154, with heterozygous sites on diagnostic bases; see below), *B. latilineola*, *B. occipitalis*, *B. opiliae*, and *B. raiensis* could not be discriminated from *B. dorsalis*. Concerning ITS1, length polymorphism differentiated *B. carambolae* from other species in the complex and PCR amplification was more successful than with EIF3L (~99%) (Supplementary Table S1).

![Fig. 2. Photographs of the wing (above) and abdomen (below) from typical *B. carambolae* (A, D), *B. dorsalis* (B, E), and *B. occipitalis* (C, F) specimens and from specimens with morphological traits intermediate between *B. carambolae* and *B. dorsalis* (G, I) and between *B. carambolae*, *B. dorsalis* and *B. occipitalis* (H, J).](https://academic.oup.com/jee/article-lookup/16621617/308769)
Two molecular markers (COI + EIF3L or ITS1) thus successfully discriminated B. dorsalis from 6 species of the complex, including the morphologically similar pest species B. carambolae and B. occipitalis (Table 2). Bactrocera dorsalis had bases different from B. carambolae at positions 99 and 348 for EIF3L, and from B. occipitalis at positions 178 and 530 for COI. In both species pairs, the different bases were rare and often not borne by all 278 B. dorsalis specimens. Identification of the 10 ambiguous specimens was thus achieved using the ML trees constructed from each marker (see list in Table 1): the COI barcode confirmed IDN_322, IDN_331, IDN_373, PHL_268, and PHL_269 as not B. occipitalis (Fig. 3A), and the EIF3L marker and ITS1 length polymorphism confirmed IDN_323, MYS_352, MYS_354, and MYS_356 as B. carambolae and IDN_322, IDN_331 and PHL_268 as B. dorsalis (Table 1; Fig. 3B). Specimens IDN_373, IDN_376, and PHL_269 were neither B. occipitalis nor B. carambolae according to COI and EIF3L (Fig. 3), but they could be either B. dorsalis or 1 of the 6 lesser-known species unattracted to ME (Table 1).

B. kandiensis and B. raiensis could not be discriminated from B. dorsalis with COI nor EIF3L (Fig. 3), but their morphology is sufficient for identification (Table 2). The limited individual diagnostic capacity of these molecular markers was confirmed by the considerable overlap between the distributions of intraspecific and interspecific distances (Supplementary Fig. S2) and the low sampling coverage of haplotype diversity (Supplementary Fig. S3). Using next-generation sequencing will produce large genomic resources with which to target diagnostic nuclear loci. However, we demonstrate that combining a mtDNA (i.e., COI) with a nDNA (i.e., EIF3L) marker is effective for identifying B. dorsalis among morphologically similar specimens of sibling and other closely related taxa. To assign taxonomic status to specimens, we therefore recommend combining expert morphological description with both mitochondrial and nuclear DNA barcodes, particularly when studying B. dorsalis in Southeast Asian islands.

Table 2: Success (√) and fail (X) of each morphological and molecular tool to discriminate B. dorsalis from other species in the complex, to highlight the importance of a multidisciplinary approach when studying the B. dorsalis complex.

<table>
<thead>
<tr>
<th>Morphology</th>
<th>COI</th>
<th>EIF3L/ITS1</th>
</tr>
</thead>
<tbody>
<tr>
<td>B. cacuminata</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>B. carambolae</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>B. endiandrae</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>B. kandiensis</td>
<td>√</td>
<td>X</td>
</tr>
<tr>
<td>B. latilineola</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>B. occipitalis</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>B. opiliae</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>B. raiensis</td>
<td>√</td>
<td>X</td>
</tr>
</tbody>
</table>

Acknowledgments

This work was funded by the French Agricultural Research Centre for International Development (CIRAD) and the French Agency For Food, Environmental and Occupational Health & Safety (ANSES) (project BACTRACK). We are thankful to the many collaborators named in Supplementary Table S1 for providing us with fruit fly specimens, the phytosanitary border inspectors at Roissy airport, and Anais Chailleux in particular for coordinating sampling in Africa. We thank the CBGP (Centre de Biologie pour la Gestion des Populations) for giving us access to Molecular Biology and Collection platforms. We express our deeply felt gratitude to Emma Artige (INRAE) for having taken on the colossal administrative workload to obtain the legal agreements between CIRAD and 21 different government agencies required by the Nagoya Protocol for the Convention of Biological Diversity, and Servane Baufumé (CIRAD) for her assistance and insightful tips.
Author Contributions
Emeline Charbonnel (Conceptualization [Equal], Data curation [Lead], Formal analysis [Equal], Investigation [Equal], Methodology [Equal], Validation [Equal], Visualization [Equal], Writing – original draft [Equal], Marie Pierre Chapuis (Conceptualization [Lead], Funding acquisition [Lead], Methodology [Lead], Project administration [Lead], Resources [Equal], Supervision [Lead], Validation [Equal], Writing – original draft [Equal]), Andrea Taddei (Data curation [Supporting], Resources [Supporting], Writing – original draft [Supporting]), Laure Benoit (Data curation [Supporting], Resources [Equal], Writing – original draft [Supporting]), Raphaëlle Mouttet (Data curation [Supporting], Writing – original draft [Supporting]), and David Ouvrard (Conceptualization [Equal], Funding acquisition [Lead], Methodology [Supporting], Project administration [Equal], Resources [Equal], Supervision [Lead], Validation [Equal], Writing – original draft [Equal])

Data Availability
All COI and EIF3L sequences produced in this study were deposited in GenBank under accession numbers OR208262–OR208544, OQ547452–OG547739, OR554265, and OR197650–OR197700.

Supplementary Material
Supplementary material is available at Journal of Economic Entomology online.

References


Drew RAI, Romig MC. Tropical Fruit Flies (Tephritidae: Dacinae) of South-East Asia: Indomala/ to North-West Australasia. Oxfordshire, UK: CABI; 2013.

Drew RAI, Romig MC. Keys to the tropical fruit flies (Tephritidae: Dacinae) of South-East Asia: Indomala/ to North-West Australasia. Oxfordshire, UK: CABI; 2016.


EPPO. PM 7/129 (2) DNA barcoding as an identification tool for a number of regulated pests. EPPO Bull. 2021;51:100–143.


