FINAL REPORT

Validation of the morphological and molecular diagnostic protocols for *Bactrocera dorsalis*

IPPC ISPM 27 Diagnostic protocols for regulated pests DP 29: *Bactrocera dorsalis*
and
EPPO PM 7/129 (2) DNA barcoding as an identification tool for a number of regulated pests

Version N° 01 – 18 October 2021

Summary

1. Introduction

The European Reference Laboratory for Insects and Mites has to select, adapt or develop reliable diagnostic protocols for the phytosanitary insect and mite species that are relevant for the European Union (included in the Commission Delegated Regulation (EU) 2019/1702 and in the EURL for Insects and Mites working programmes). One of the tasks of the EURL is to validate available diagnostic protocols before recommending their use to the National Reference Laboratories of the European Union.

According to the ISO/IEC 17025 standard, the validation of a test is defined as the "confirmation by examination and the provision of objective evidence that the particular requirements for a given intended use are met". In fact, this confirmation consists of comparing the values of the performance criteria determined during the test characterization study with those expected or assigned beforehand (limits of acceptability, objectives to be achieved), then declaring the analytical test valid or invalid. In the field of entomology, identification is qualitative, meaning that diagnostic protocols allow the identification at a given taxonomic level providing a response in terms of presence/absence.

The EURL for Insects and Mites focuses on the validation of tests published in international or regional standards, such as those issued by the International Plant Protection Convention (IPPC) or the European and Mediterranean Plant Protection Organization (EPPO).

Bactrocera dorsalis (Hendel, 1912) (Diptera: Tephritidae) is endemic and widespread in tropical areas of Southeast Asia and is part of the *Bactrocera dorsalis* complex, a group of closely related species that comprises 88 described species (Doorenweerd *et al.*, 2018). The complex is named after *Bactrocera dorsalis* because of the importance of this polyphagous commercial fruit pest worldwide. However, the complex as a whole does not represent a monophyletic group (Leblanc *et al.*, 2015).
An ongoing debate is taking place regarding the taxonomy of *B. papayae* and *B. invadens*: some experts consider them indistinguishable from *B. dorsalis* and thus, the same biological species; others consider them as valid taxonomic entities distinct, although extremely similar, from *B. dorsalis* (Clarke *et al.*, 2005; Chen and Hui, 2007; Schutze *et al.*, 2015a, b; Drew & Romig, 2016; Schutze *et al.*, 2017). In this paper, these three species are considered as the same species and treated under the name of *B. dorsalis sensu lato* (*s.l.*). For a list of other *B. dorsalis* synonyms, see Pest information in IPPC, 2019.
Due to its high reproductive and biotic potential, a rapid dispersal ability and a broad host range, *B. dorsalis* is considered a species with a high invasive capacity. Since the first report in Kenya in 2003, the species has rapidly colonized almost the entire African continent. It is locally present in the United States and the recent, repeated interceptions in Italy (2018, 2019), Austria (2014-2019) and France (2019) keep the European Plant Protection Organisations on alert (Egartner *et al.* 2019; CABI, 2021; EPPO, 2021a).

Bactrocera dorsalis (Diptera: Tephritidae) is a European Union regulated species, listed among the EU quarantine pests (Annex II of the Commission Implementing Regulation (EU) 2019/2072) and among the EU priority pests (Commission Delegated Regulation (EU) 2019/1702).

2. Scope of validation and diagnostic protocols

2.1 Scope

The scope of this validation study is to provide objective evidence that the selected diagnostic protocols are suitable to perform routine identification of *Bactrocera dorsalis* by the staff of the EU National Reference Laboratories.
Note that in this document, when reference is made to "*Bactrocera dorsalis*" simply, it means "*Bactrocera dorsalis s.l.* (*sensu lato*)". For a brief explanation of the meaning of *B. dorsalis s.l.*, see the introduction.

2.2 Description of the diagnostic protocols under validation

This validation study is focused on two diagnostic protocols for the morphological and molecular identification of *Bactrocera dorsalis,* i.e.:

- IPPC ISPM 27 Diagnostic protocols for regulated pests DP 29: *Bactrocera dorsalis* (IPPC, 2019), which includes:
  - tables of characters and keys for the morphological identification of adults of the *B. dorsalis* complex
  - a molecular test to distinguish *B. carambolae* from other species of *B. dorsalis* complex
- EPPO PM 7/129 (2) DNA barcoding as an identification tool for a number of regulated pests (EPPO, 2021b), which includes tests for the DNA barcoding of arthropods.

Validation was conducted according to the EPPO PM7/ 98(4) Specific requirements for laboratories preparing accreditation for a plant pest diagnostic activity (EPPO, 2019).

2.2.1 Morphological identification of adults

Protocol: IPPC ISPM 27 Diagnostic protocols for regulated pests DP 29: *Bactrocera dorsalis*

The identification at the level of the species for the *Bactrocera dorsalis* complex requires morphological examination of adult flies. The identification is possible both on male and female specimens. The use of a stereomicroscope is needed (≥ 20 magnification).
The protocol provides guidance for the identification of the species for the *Bactrocera dorsalis* complex starting from the subgenus level:

- paragraph 4.2.1, characters for the identification of the **subgenus *Bactrocera* (*Bactrocera*)**;
- paragraph 4.2.2, list of characters (Table 2) that are useful for the identification of **the *B. dorsalis* complex**. A specimen must have all the characters that match the description provided to be identified as belonging to the *B. dorsalis* complex;
- paragraph 4.2.4, diagnostic key to **six economically important species** belonging to the *Bactrocera dorsalis* complex. The species included in the key are *B. caryeae*, *B. kandiensis*, *B. occipitalis*, *B. pyrifoliae*, *B. carambolae* and *B. dorsalis s.l.*. Results obtained by means of this key have to be confirmed by checking the **list of morphological characters** included in Table 3 (paragraph 4.2.3).

The validation planned in this document took into account the list of characters for the identification of the *B. dorsalis* complex (4.2.2), the diagnostic key to six economically important species (4.2.4) and Table 3 – 'Diagnostic morphological characters of adult fruit flies of six economically important species of the *Bactrocera dorsalis* complex' (4.2.3). **The list of characters for the identification of the subgenus *Bactrocera* (*Bactrocera*), as well as the observation of male and female genitalia, was not subject of this study**, due to the following practical reasons:

- the dissection of genitalia must be performed in advance by supervisor (for definition of staff roles, see 5 Time schedule and staff) and, if the whole abdomen has to be removed, that means that the characters of the abdomen are not available anymore for the operators to be checked;
- handling of male genitalia by supervisor and operators risks damaging/breaking the aedeagus, with a considerable impact of repeatability and reproducibility of the analysis;
- as it is stated in the protocol by authors themselves, the aedeagus length "does not always provide a clear diagnosis because of overlap in the range of aedeagus (M) and aculeus (F) size between *B. dorsalis s.l.* and *B. carambolae*" (see 4.2.3).

This choice was supported by the opinion of an internationally renowned expert, Marc de Meyer (Royal Museum for Central Africa, Tervuren, Belgium), directly questioned on the matter. In particular, his opinion was that "No, I don't think that the reliability of an identification to species level is reduced if you cannot confirm the subgenus first. As long as the character state data set used for identification of the species is large enough and allow excluding species, a species identification can be done based on this without knowing the subgenus".

2.2.2 Molecular identification of adults, larvae and pupae

Molecular tests can support morphological identifications of adults. Furthermore, these tests can especially be used when dealing with other developmental stages than adults (e.g. larvae, pupae). Two protocols were validated.

- Protocol: EPPO PM 7/129 (2) DNA barcoding as an identification tool for a number of regulated pests (EPPO, 2021b), Appendix 1 – DNA barcoding of arthropods

DNA barcoding is used to identify arthropods at a certain taxonomic level. The chosen marker region is the mitochondrial cytochrome c oxidase I (*COI*) gene. Two different primer sets (LCO1490/HCO2198 and LepF/LepR), targeting this gene, were validated.

- Protocol: IPPC ISPM 27 Diagnostic protocols for regulated pests DP 29: *Bactrocera dorsalis*

In this protocol, molecular tests are not recommended as standalone test in order to discriminate the six economically relevant species mentioned in the standard. When identifying *B. carambolae* and *B. dorsalis s.l.* specimens using this protocol, this

molecular test is necessary for accurate identification whenever adult morphology alone cannot distinguish between the two species.
DNA sequencing of either the internal transcribed spacer 1 (*ITS*1) or 2 (*ITS*2) nuclear DNA regions has been proposed as a reliable way to distinguish between the species *B. carambolae* and *B. dorsalis s.l.* (Boykin et al., 2014; Schutze et al., 2015a) (Paragraph 4.3.2) . The *ITS*1 test as described by Boykin et al. (2014) for distinguishing the two species is included in the current protocol. This test is designed to diagnose a fly as *B. carambolae* based on the presence of a unique DNA insert that is not present in *B. dorsalis s.l.*. The *ITS*1 test in the IPPC protocol has not been tested to distinguish *B. carambolae* from all other *Bactrocera dorsalis* complex species.

2.3 Composition of the sample set

A set of 40 samples was used. It consisted of 40 adult specimens belonging to the target and to the non-target species (11 taxa). Table 1 provides a summary of the sample set. For the detailed composition of the sample set, see Appendix 1 of this document. Target specimens came from 6 different countries (Laos, Mali, Senegal, Taiwan, Thailand and Vietnam). Non-target specimens belonged all to the family Tephritidae and were selected primarily based on the close similarity to the target species and the availability in the partner laboratories reference collections. The origin of the non-target specimens was variable, including Asian, African and European countries. Each sample was re-labelled with a number from 1 to 40 by supervisors, after randomization (for definition of staff roles, see 5 - Time schedule and staff). Original codification of samples was available only to supervisors. All samples were preserved in single tubes, filled with 95% ethanol.
The composition of the set was chosen to allow the evaluation of sensitivity, specificity, repeatability, reproducibility and accuracy of the tests.
For the validation of the molecular protocols additional, smaller sample sets were prepared (see 3.3.2 – Molecular tests).

Table 1: Summary of the composition of the sample set

Species	Number	Provider	Number	Provider	Total Number
B. dorsalis	12	ANSES	2	AGES	**14**
B. carambolae	5	ANSES	1	AGES	**6**
B. caryeae	0	ANSES	2	AGES	**2**
B. kandiensis	5	ANSES	0	AGES	**5**
B. occipitalis	0	ANSES	2	AGES	**2**
B. pyrifoliae	0	ANSES	1	AGES	**1**
Anastrepha suspensa	0	ANSES	2	AGES	**2**
Anastrepha obliqua	0	ANSES	2	AGES	**2**
Zeugodacus cucurbitae	2	ANSES	0	AGES	**2**
Bactrocera oleae	2	ANSES	0	AGES	**2**
Dacus ciliatus	2	ANSES	0	AGES	**2**
					40

3. Validation of the diagnostic protocols

3.1 Performance characteristics assessed

According to the guidance given in PM 7/98 (4) (EPPO, 2019) and the definitions given in PM 7/76 (5) (EPPO, 2018a), PM 7/122 (1) (EPPO, 2014) and EPPO PM 7/129 (2) (EPPO, 2021b), validation of diagnostic tests relies on the evaluation of the following performance characteristics: sensitivity, specificity, reproducibility, repeatability and accuracy.

Table 2 shows the criteria that were used to calculate the performance characteristics of the tests.

Table 2: Definition and calculation of performance characteristics

Performance criteria	Definition	Calculation
Diagnostic specificity	The proportion of non-target samples (true negatives) testing negative compared with results from an alternative test (or combination of tests) *Comments: as far as possible, the evaluation of specificity must include samples from non-target organisms that can be confused with the target species*	Diagnostic specificity = true negatives/(true negatives + false positives)
Analytical specificity	**Inclusivity:** The performance of a test with a range of target organisms covering genetic diversity, different geographical origin and hosts	-
	Exclusivity: The performance of a test with regards to cross-reaction with a range of non-targets (e.g. closely related organisms)	-
Diagnostic sensitivity	The proportion of target samples (true positives) testing positive compared with results from an alternative test (or combination of tests)	Diagnostic sensitivity = true positives/(true positives + false negatives)
Analytical sensitivity	The smallest amount of target that can be detected reliably. In the case of molecular test, it is referred to as "limit of detection", i.e. the lowest DNA concentration of the target organism that can be reliably detected). For DNA barcoding the limit of detection is the DNA concentration that is sufficient to generate an amplicon which can be sequenced and leading to a HQ consensus sequence (Phred score > 40) of at least 99%.	-
Repeatability	The level of agreement between replicates of a sample tested under the same conditions	% level of agreement
Reproducibility	The ability of a test to provide consistent results when applied to aliquots of the same sample tested under different conditions (e.g. time, persons, equipment, location)	% level of agreement
Accuracy	The proportion of target samples (true positives) testing positive and non-target samples (true negatives) testing negative compared with the total number of samples. It is worth noting that the accuracy is a global criterion which can be subdivided, to refine the analysis, into three other criteria: sensitivity, specificity and repeatability.	Accuracy = (true positives + true negatives)/(true positives + false negatives + true negatives + false positives)

3.2 Performance characteristics already available

Performance characteristics obtained within this study were compared with performance characteristics already available for the respective tests. For the EPPO PM 7/129 (2) DNA barcoding as an identification tool for a number of regulated pests (EPPO, 2021b) performance characteristics were already available in Appendix 1, paragraph 4 of the standard itself. For the other molecular tests and for the morphological test, performance characteristics were not available. In this latter case, the expected performance characteristics were considered equal to 100%, with the exclusion of the molecular analytical sensitivity, which consists in a measure of concentration expressed in ng/µl.

EPPO PM 7/129 (2) DNA barcoding as an identification tool for a number of regulated pests (EPPO, 2021b), performance characteristics:

- **Analytical sensitivity:** DNA concentration (PCR amplicon) of 4 ng/µl sufficient for high quality amplicon sequencing – HQ quality (Phred score >40) consensus sequence of at least 99%.
- **Analytical specificity:** The interspecific variation of the gene locus was determined to be sufficient for identification at species level.
    - Inclusivity: Summary list of identified arthropods in Appendix 1 (Table 1) of the standard and EPPO validation sheet (http://dc.eppo.int/tps.php).
    - Exclusivity: n.a.
- **Diagnostic sensitivity:** 98%-100%

Additional performance characteristics in literature: no additional information available.

3.3 Validation protocol

3.3.1 Morphological test

The set of 40 specimens was analyzed by three operators (for definition of staff roles, see 5 Time schedule and staff), belonging to the two different institutes (AGES and ANSES). The set composition was defined by the supervisors and known to the supervisors only.
Supervisors provided operators with the Check Lists and Summary Results sheet in Appendix 2, but did not provide operators with origin and host plants data. During the analysis, to be carried out at a stereomicroscope, operators have filled the Check List for each sample and record the identification results on the Summary Results sheet. For a better understanding of some morphological characters, especially concerning their colour, operators observed each specimens both in ethanol and dry. The results of the identification were expressed as:
- POSITIVE, if **all** the characters of the specimens matched with those of *B. dorsalis s.l.*;
- NEGATIVE, if **not all** the characters of the specimens matched with those of *B. dorsalis s.l.*;
- NOT DETERMINED (n.d.), if the matching of characters was ambiguous. In this case, operators were required to highlight which characters lead to the ambiguous results, i.e. the impossibility of identification (Notes column in the Summary Results sheet).
After the analysis, the Summary Results sheet were retrieved by the supervisors. In case of deviations of the results from the expected ones, the Check List allowed the supervisors to precisely identify any critical issues within the protocol.

Performance characteristics were assessed according to the following *a priori* established plan:

- Diagnostic sensitivity and specificity were assessed on the basis of the analysis of the whole sample set carried out by operator 3 (ANSES);
- Repeatability was assessed on the basis of the analysis of the whole sample set carried out by operator 3 (ANSES) (three repetitions of analysis).
- Reproducibility was assessed on the basis of the analysis of the whole sample set carried out by operator 1, 2 (AGES) and 3 (ANSES) (first of the three repetitions of analysis).

While performing the morphological analysis for the third and last repetition, operator 3 removed one leg from each specimen and placed it in an Eppendorf vial, in 70% ethanol, keeping the respective code. The leg samples were shipped to AGES for the DNA extraction and the molecular analysis. For some of the specimens, DNA extraction was repeated on the whole specimen due to the fact that DNA of unsufficient quality and quantity was purified from leg.
Figure 1 provides a scheme of the activity.

3.3.2 Molecular tests

DNA extraction
For DNA extraction of whole specimens (e.g. analytical sensitivity) the DNeasy Blood & Tissue Kit (Qiagen) was used. For the DNA extraction from single legs the QIAamp DNA Micro Kit (Qiagen) was used.

Analytical specificity
The same set of specimens used for the morphological analysis was used for the validation of the molecular tests - see 2.3 for further specifications.
Inclusivity: 14 targets
Exclusivity: 26 non-targets, 16 of which belong to the *Bactrocera dorsalis* complex

The primer sets and PCR parameters are described in Appendix 3.
SANGER sequencing was outsourced to a certified sequencing service provider (EUROFINS Genomics).
Data-analysis: The software Geneious prime® 10.1.3 was used for the consensus sequence preparation. For sequence alignment following genetic databases were consulted: NCBI-GenBank, Bold and EPPO Q-Bank.

In silico testing of analytical specificity for molecular tests:

LCO1490/HCO2198 and LepF/LepR: Search for Tephritidae *in silico* by a database alignment (NCBI GenBank) (see App. 7).

*ITS*6/7 primer: Search for *Bactrocera* sp. *in silico* by a database alignment (NCBI GenBank) (see App. 7).

Analytical sensitivity
4 samples consisting of one adult, a leg, a larva and a pupa from *B. dorsalis* were prepared in different dilutions. Three experimental repetitions were performed with this sample set.

Sample set:
1 adult specimen of *B. dorsalis* (333/20)
1 leg of *B. dorsalis* (334/20)
1 larva of *B. dorsalis* (335/20)
1 pupa of *B. dorsalis* (336/20)

Dilutions (1:10, 1:100, 1:1.000; 1:10.000; 1: 100.000, 1:1.000.000).
To define the limit of detection for DNA barcoding, the two highest dilutions which resulted in an amplicon were sequenced and analysed.

Repeatability
Four biological replicates of *B. dorsalis* (adult, leg, larva, pupa) in three different dilutions (last dilution near by the detection limit) were analysed with 3 technical replicates to determine the repeatability.

Reproducibility

Testing reproducibility of the PCR tests:

Three targets (adult, larva and pupa) and three non targets (adult, larva) were used to test the reproducibility of the PCR tests (Table 3). These tests were performed with three replicates and under different conditions (two operators at different days and using different thermocycler machines).

Table 3: Sample set used to test reproducibility

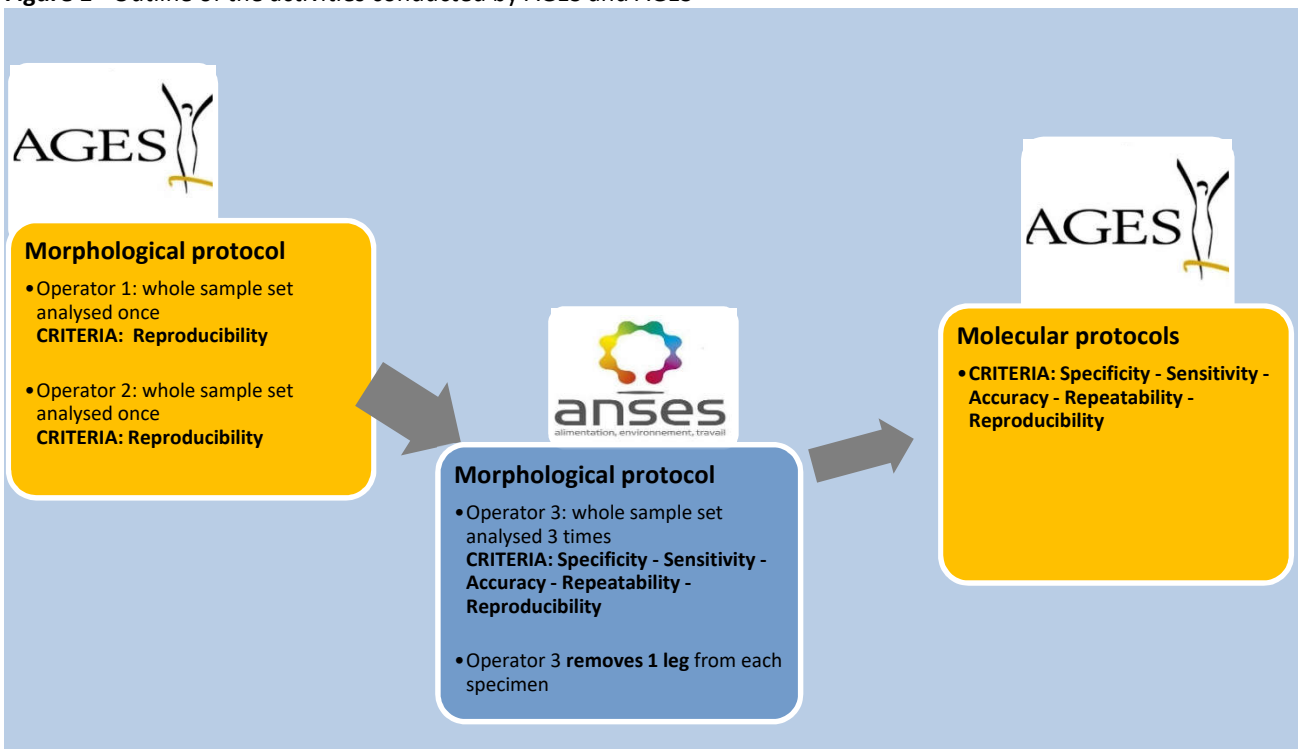
Target	Non target	Origin
B. dorsalis adult	/	Thailand /Saraburi
B. dorsalis larva	/	Thailand /Saraburi
B. dorsalis pupa	/	Thailand /Saraburi
	Bactrocera correcta larva	India
	Bactrocera carambolae adult	French Guyana
	Bactrocera latifrons larva	Thailand

Testing reproducibility of the SANGER sequence analysis:

The reproducibility of the SANGER sequence analysis was tested with the same sample set. The sequence analysis was performed by two operators on different days. The alignment of the consensus sequence was performed in three different databases (NCBI GenBank, Bold, Q-Bank).

Specifications and parameters for the molecular tests are provided in Appendix 3.
Figure 1 provides a scheme of the activity.

Figure 1 - Outline of the activities conducted by AGES and AGES



4. Performance adequacy and validation

The performance values obtained by the diagnostic protocol/ test were compared with the predetermined, expected performance characteristics.
The adequate expected performance characteristics are shown in Table 4. They are also referred to as "limits of acceptability" of the test. If the obtained performance characteristics did not reach the expected values, a cause analysis was carried out to identify the critical steps in the test(s) that led to the unexpected results (i.e., false negatives, false positives, not determined).
Some positive deviations were expected, as DNA barcoding is according to the IPPC standard insufficient to discriminate *B. carambolae* from *B. dorsalis.* Due to this, the lowest calculated value of expected performance characteristics with the current sampel panel is 77% (diagnostic specificity). This also influences the accuracy.

Table 4: Expected performance characteristics (limits of acceptability).

Performance criteria	Expected performance characteristics		
	IPPC 27 -DP 29 *Bactrocera dorsalis* – morphological identification	EPPO PM 7/129 DNA barcoding	IPPC 27 -DP 29 *Bactrocera dorsalis* – *ITS1* primer
Diagnostic specificity	100%	77%	100%
Analytical specificity (Inclusivity)	-	-	100%
Diagnostic sensitivity	100%	100%	100%
Analytical sensitivity	1 adult specimen	4 ng/µl	4 ng/µl
Repeatability	100%	100%	100%
Reproducibility	100%	100%	100%
Accuracy	100%	85%	100%

[a] as from Appendix 1, paragraph 4 of EPPO PM 7/129 (2) DNA barcoding as an identification tool for a number of regulated pests (EPPO, 2021b)

5. Time schedule and staff

The trial period was from May to August 2020 and involved staff from the EURL for Insects and Mites.

Participating staff:

- for morphological tests:
  - Experts/ Supervisors: Sylvia Blümel, Valérie Balmès, Raphaëlle Mouttet
    Role: definition, randomization and blind-codification of sample set, preparation of check-lists, collection and analyses of results, drafting of final report
  - Technical staff/ Operators: Christa Lethmayer, Alois Egartner, Andrea Taddei
    Role: performance of analyses, help to supervisor in the interpretation and analysis of results, drafting of final report

- for molecular tests:
  - Experts/ Supervisors: Richard Gottsberger, Helga Reisenzein
    Role: definition, randomization and blind-codification of sample set, collection and analyses of results, drafting of final report
  - Technical staff/ Operators: Claudia Heiss, Christina Lippitz, Chiara Pohn
    Role: performance of analyses, help to supervisor in the interpretation and analysis of results, drafting of final report

6. Results of the validation analysis

6.1 Morphological test

Protocol: IPPC ISPM 27 Diagnostic protocols for regulated pests DP 29: *Bactrocera dorsalis* (IPPC, 2019)

The values obtained for diagnostic specificity, diagnostic sensitivity, accuracy and repeatability met the expected value of 100% (Table 5). The test was found to be inclusive for target specimens from Laos, Mali, Senegal, Taiwan, Thailand and Vietnam and exclusive for a range of non-target specimens belonging to the *B. dorsalis* complex (*B. carambolae, B. caryeae, B. kandiensis, B. occipitalis, B. pyrifoliae*), the *Bactrocera* genus (*B. oleae*) and non-*Bactrocera* Tephritidae (*Anastrepha obliqua, Anastrepha suspensa, Dacus ciliatus, Zeugodacus cucurbitae*).
The value obtained for reproducibility did not meet the expected value of 100%, but reached a value of 87.5%. The cause was found in the divergent results obtained for 9 specimens either by 1 (5 specimens) or 2 operators (4 specimens), as summarized in Table 6.The check lists compiled by operators during the performance of the analyses allowed to track back the critical steps in the protocol that led to the deviation from the expected results.
Appendix 4 of this document shows the results obtained by the three operators.
Appendix 5 shows the calculations for the performance characteristics.

Table 5: Summary of the results obtained for the morphological protocol

Performance criteria	Definition	Calculation	Expected performance characteristics	Obtained performance characteristics
Diagnostic specificity	The proportion of non-target samples (true negatives) testing negative compared with results from an alternative test (or combination of tests)	Diagnostic specificity = true negatives/(true negatives + false positives)	**100%**	**100%**
Analytical specificity	**Inclusivity:** The performance of a test with a range of target organisms covering genetic diversity, different geographical origin and hosts	-	**-**	**Laos Mali Senegal Taiwan Thailand Vietnam**
	Exclusivity: The performance of a test with regards to cross-reaction with a range of non-targets (e.g. closely related organisms)	-	**-**	***B. carambolae B. caryeae B. kandiensis B. occipitalis B. pyrifoliae B. oleae Dacus ciliatus Zeugodacus cucurbitae Anastrepha obliqua Anastrepha suspensa***
Diagnostic sensitivity	The proportion of target samples (true positives) testing positive compared with results from an alternative test (or combination of tests)	Diagnostic sensitivity = true positives/(true positives + false negatives)	**100%**	**100%**
Analytical sensitivity	The smallest amount of target that can be detected reliably	-	**1 adult specimen**	**1 adult specimen**
Repeatability	The level of agreement between replicates of a sample tested under the same conditions	% level of agreement	**100%**	**100%**
Reproducibility	The ability of a test to provide consistent results when applied to aliquots of the same sample tested under different conditions (e.g. time, persons, equipment, location)	% level of agreement	**100%**	**87,5%**
Accuracy	The proportion of target samples (true positives) testing positive and non-target samples (true negatives) testing negative compared with the total number of samples	Accuracy = (true positives + true negatives)/(true positives + false negatives + true negatives + false positives)	**100%**	**100%**

6.1.1 *Analysis of critical steps in the protocol*

The morphological characters, as described in the protocol, which were recognized at the basis of divergent results are listed in Table 6. Among them, the descriptions of the following characters were particularly relevant:

I. Character "Costal band", relevant for the divergent results for 7 specimens (5 *B. carambolae* and 2 *B. occipitalis*);
II. Character “Transversal black band on tergite 3”, relevant for the divergent results for 3 specimens (3 *B. carambolae*).

With reference to these diagnostic characters, it has been possible to detect that their description in the protocol (either in the diagnostic key or in the character table or both) is sometimes prone to uncertain interpretation, so that it can sometimes be misleading for the protocol user. An in-depth analysis is provided below.

Table 6: Samples for which divergent results were obtained with respect to *a priori* assigned value

Sample code	Assigned value	Result by Operator 1	Result by Operator 2	Relevant character	Description in the protocol
4	Negative (*B. carambolae*)	Negative	**Positive**	Costal band confluent/ overlapping	Table 3 (4.2.3), page 11
8	Negative (B. carambolae)	**Positive**	**Positive**	Costal band confluent/ overlapping Transverse band on abdominal tergite 3	Table 3 (4.2.3), page 11 Table 3 (4.2.3), page 9; Diagnostic key (4.2.4), page 12
18	Positive	Positive	**Negative**	Postpronotal lobe	Table 3 (4.2.3), page 10
22	Negative (B. carambolae)	**Not determined**	**Positive**	Transverse band on abdominal tergite 3 Costal band confluent/ overlapping	Table 3 (4.2.3), page 9; Diagnostic key (4.2.4), page 12 Table 3 (4.2.3), page 11
23	Negative (B. occipitalis)	Negative	**Not determined**	Costal band distinctly overlapping R2+3 and expanding broadly around apex of wing reaching mid-point between R2+3 and R4+5	Table 3 (4.2.3), page 11 Diagnostic key (4.2.4), page 12
27	Negative (B. carambolae)	**Not determined**	**Positive**	Scutum color Costal band confluent/ overlapping	Table 3 (4.2.3), page 10 Table 3 (4.2.3), page 11
36	Positive	**Not determined**	Positive	-	-
39	Negative (B. carambolae)	**Not determined**	**Positive**	Transverse band on abdominal tergite 3 Costal band confluent/ overlapping	Table 3 (4.2.3), page 9; Diagnostic key (4.2.4), page 12 Table 3 (4.2.3), page 11
40	Negative (B. occipitalis)	Negative	**Positive**	Costal band distinctly overlapping R2+3 and expanding broadly around apex of wing reaching mid-point between R2+3 and R4+5	Table 3 (4.2.3), page 11 Diagnostic key (4.2.4), page 12

Character "Costal band" in the identification of *B. occipitalis* and discrimination with *B. carambolae*

In the Diagnostic key to six economically important species belonging to the *Bactrocera dorsalis* complex (paragraph 4.2.4, page 12), at point 3 the description states:

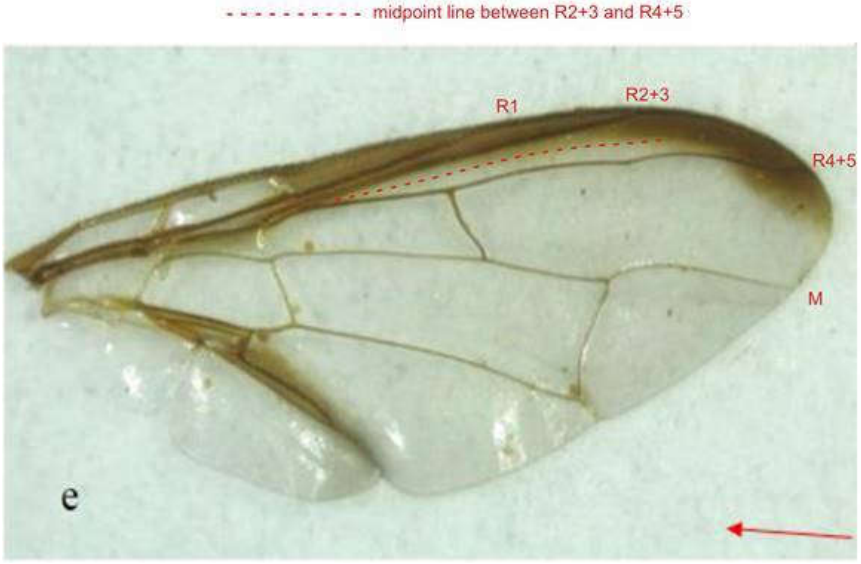
3. Costal band distinctly overlapping R_{2+3} and expanding broadly around apex of wing reaching mid-point between R_{2+3} and R_{4+5} (Figure 16(e))……B. occipitalis,

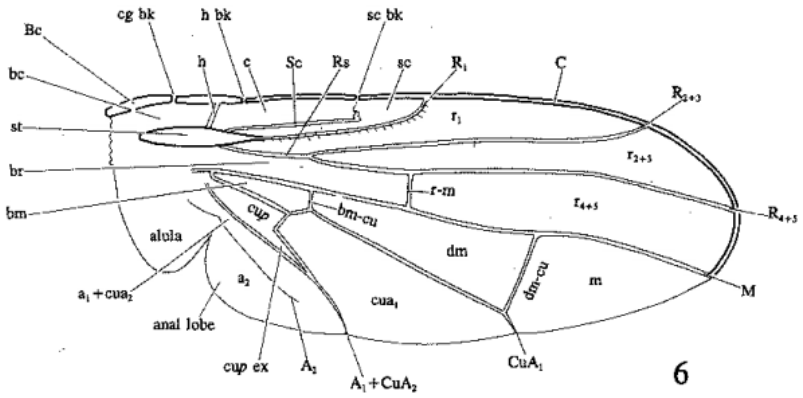
This description can lead to misunderstanding, as one might be led to think that "mid-point" refers to a point along the edge of the wing and consider the only option of a point between R_{4+5} and M as possible. This misunderstanding would consequently lead to the belief that the wing veins are misnamed in the description, which is not the case. As international Tephritidae experts Norman Barr and Camiel Doorenweerd explained, in fact "mid-point” refers to the imaginary line that divides the cell in half between R_{2+3} and R_{4+5} (Fig. 2) and it should be better considered as a “mid-line” rather than a “mid-point”.

This misinterpretation of the term "mid-point" contributed to the divergent results for the two *B. occipitalis* specimens in the fact that the operators, when looking for a mid-point, failed to locate it between R_{2+3} and R_{4+5}, refusing to choose *B. occipitalis.* Possible options for the costal band in the *Bactrocera* genus are as follows (Camiel Doorenweerd's communication):

1) Costal band confluent with R_{2+3}
2) Costal band faintly (slightly) crosses R_{2+3}
3) Costal band reaches midway (mid-point) between R_{2+3} and R_{4+5} [*B. occipitalis*]
4) Costal band reaches up to R_{4+5}
5) Costal band confluent with R_{4+5}

Figure 2 – *B. occipitalis* wing pattern (e), as from DP 29 (IPPC, 2019), modified by and courtesy of Camiel Doorenweerd: the dotted line indicates the mid-point between R_{2+3} and R_{4+5} ; Tephritidae wing venation (6), as from White & Elson-Harris (1992)





Figs 5-6. Adult morphology; 5, head; 6, wing. Abbreviations are listed on p. 31.

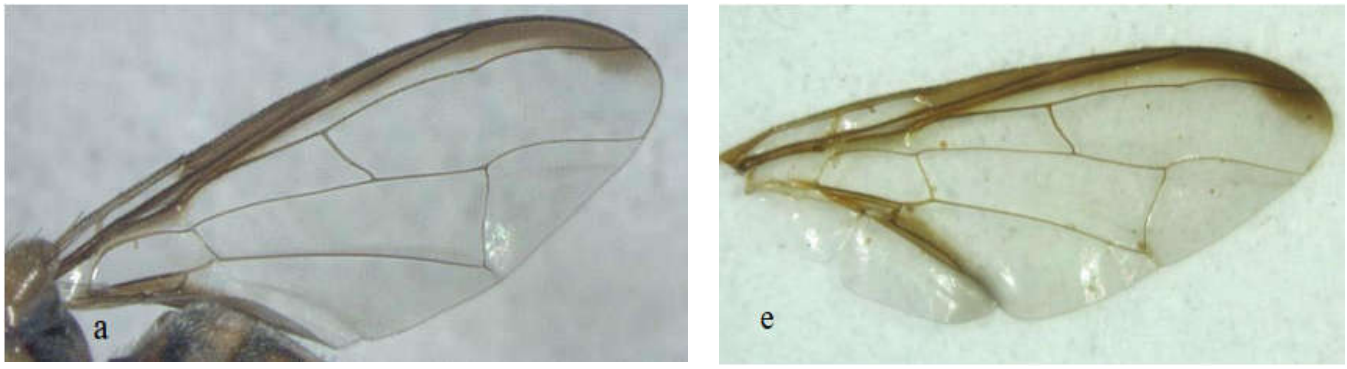
In addition, in the Diagnostic key (paragraph 4.2.4, page 12), the costal band of *B. occipitalis* and *B. carambolae* is described as follows:

3. Costal band distinctly overlapping R_{2+3} and expanding broadly around apex of wing reaching mid-point between R_{2+3} and R_{4+5} (Figure 16(e))……B. occipitalis

5. Costal band slightly overlapping R_{2+3}, moderately broad around apex of wing (Figure 16(a)); […]……B. carambolae

The difference between the shape of costal band of *B. occipitalis* ("distinctly overlapping R_{2+3}") and *B. carambolae* ("slightly overlapping R_{2+3}") is not sufficiently clear from Figures 16 (a) and 16 (e) in the protocol (Figure 3 in this document). This unclear difference contributed to the divergent results obtained for the two *B. occipitalis* specimens included on the set.
These descriptions are confirmed in Table 3 (paragraph 4.2.3, page 11).

Figure 3 – Costal band in *B. carambolae* (a) and *B. occipitalis* (e), as from DP 29 (IPPC, 2019)



Character "Costal band" in the discrimination between *B. dorsalis s.l.* and *B. carambolae*

In Table 3 (paragraph 4.2.3, page 11), the description of the costal band of *B. dorsalis s.l.* states:
Narrow, generally confluent with R_{2+3} (inter- or intra-regionally variable), narrow to moderately broad around apex of wing (Figure 16(c)),

whereas in the Diagnostic key to six economically important species belonging to the *Bactrocera dorsalis* complex (paragraph 4.2.4, page 12), at point 5, costal band is *confluent with R_{2+3}, narrow to moderately broad around apex of wing*, the adverb "generally" is missing. In this way, the description in the Table 3 includes an element of uncertainty that is not present in the key, i.e. that generally the costal band is confluent with R_{2+3}, but in some cases may not be confluent (*Bactrocera carambolae* – like overlapping costal band?)
The adverb "generally" was relevant for the divergent results obtained for 5 *B. carambolae* specimens.
After consultation with Tephritidae experts Norman Barr and Luc Leblanc, the term "generally" would mean "typically" in the table. A diagnosis of *B. dorsalis* requires confluence. *B. dorsalis* populations which do not display a costal band confluent with R_{2+3} are not known. Although rare, some populations might have 'aberrant' specimens with costal band that crosses vein R_{2+3}, but more information regarding those specimens are needed and studies are currently ongoing (Norman Barr's communication).
Description of *B. dorsalis* costal band from Schutze et al. (2015a) is recalled here: *"Wing costal band width from vein subcostal to slightly below vein R_{4+5} at wing apex; confluent with vein R_{2+3} in depth."and "narrow fuscous costal band confluent with R_{2+3} and remaining very narrow or widening slightly if it overlaps this vein, to end just beyond apex of R_{4+5} (in some specimens there is an expansion around extremity of R_{4+5}, which may be slight or expanding into a hook-like pattern)*".

Character "Transversal black band on tergite 3"

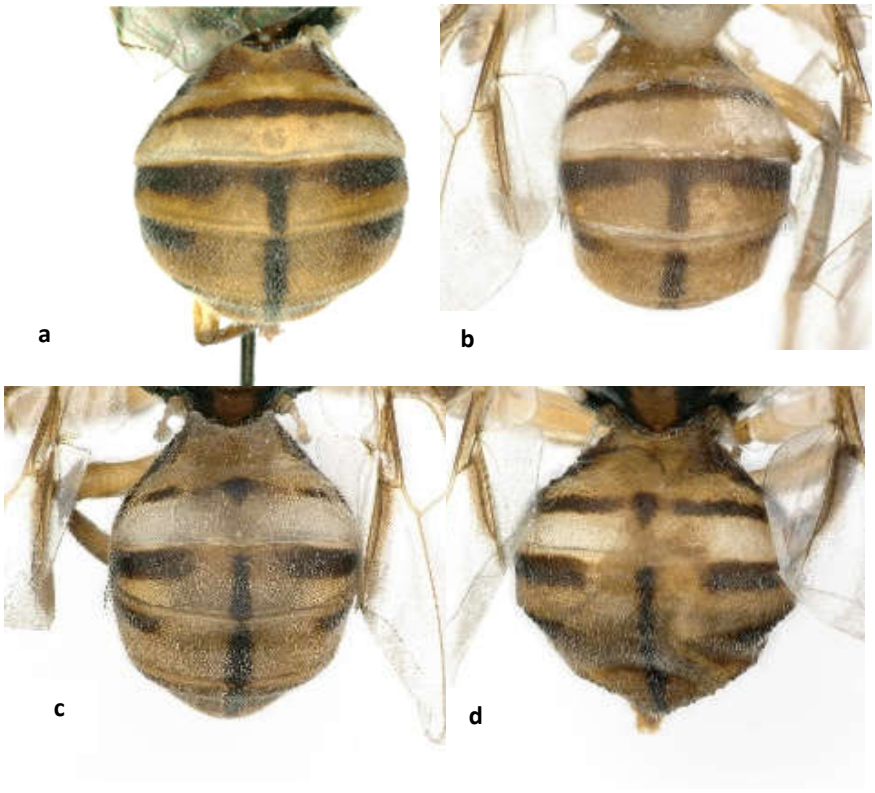
In the Diagnostic key to six economically important species belonging to the *Bactrocera dorsalis* complex (paragraph 4.2.4, page 12, point 5) and in Table 3 (paragraph 4.2.3, page 9), the character tergite 3 of *B. carambolae* is described as follows:

with a narrow transverse black band across anterior margin (constituting a "T" pattern) widening to cover lateral margins.

However, this description does not take into account the variation of the T pattern that is described in FruitFly ID Australia (Plant Health Australia, 2021). This variation consists in a non-continuous transverse band on tergite 3 (Fig.4 (b)). The fact that this variation is not mentioned in the standard led the operator to refuse to identify some specimens displaying this variation as *B. carambolae* (Fig. 5 (c) and (d)).

Figure 4 – (a) *B. carambolae* abdomen as from DP 29 (IPPC, 2019); (b) *B. carambolae* abdomen variation as from FruitFly ID Australia (Plant Health Australia, 2021)

Figure 5 – Detail of abdomen in some *B. carambolae* specimens included in the sample set for this study; (a) Sample 6 and (b) sample 27 display a continuous band on tergite 3; (c) sample 8 and (d) sample 39 display the abdominal variation for this character.



6.2 Molecular tests

For the goal of species identification in animals and some protists the *cytochrome c oxidase* subunit 1 (*COI* or cox1) gene of the mitochondrial DNA has been introduced as standard marker. DNA sequencing of the *COI* DNA barcode can be applied to distinguish several *Bactrocera* species like *B. oleae* and non-*Bactrocera* Tephritidae (e.g. *Anastrepha obliqua, Anastrepha suspensa, Dacus ciliatus, Zeugodacus cucurbitae*). Nevertheless, it was described as not providing adequate resolution to identify many species in the *B. dorsalis* complex (*B. carambolae, B. caryeae, B. kandiensis, B. occipitalis, B. pyrifoliae*) (IPPC, 2019).

According to the recommendations in the IPPC protocol, the sequencing of the *ITS*1 to distinguish *B. dorsalis* from *B. carambolae* was applied. *B. carambolae* has a unique 44bp insert that is lacking in other *Bactrocera dorsalis* complex species. Furthermore in the standard it is stated, that if there is no insert in a sample, *B. carambolae* can be excluded, but it can not be attributed to another species in the *Bactrocera dorsalis* complex.

In contrast to the validation of specificity (sample set used from morphological validation) the samples for sensitivity, repeatability and reproducibility consisted of fresh specimens of different developmental stages (e.g. adults, larvae and pupae). Furthermore, samples consisting of DNA extracted from only one leg were also included to demonstrate the usual suitability of such kind of material. *Bactrocera pyrifoliae* was excluded from the analysis, as no amplicon could be obtained with sample 21 (the only sample of *B. pyrifoliae* in this study).

Protocol: EPPO PM 7/129 (2) DNA barcoding as an identification tool for a number of regulated pests (EPPO, 2021b)

Appendix 1 – DNA barcoding of arthropods (sequencing of *COI* locus, LCO1490/HCO2198 primer set and additionally LepF/LepR primer set.

In silico testing of analytical specificity by a database alignment (NCBI GenBank) was performed with the DNA barcoding primer sets (LCO1490/HCO2198 and LepF/LepR). The search set was limited to Tephritidae. The results showed suitablity of both primer sets (see Appendix 6) for identification *of B. dorsalis*, although we have to state that barcoding is a generic test including targets and non-targets.

The values obtained for diagnostic specificity, diagnostic sensitivity, and accuracy met the expected values (Table 7).
The values for the diagnostic specificity were higher (96% and 95%) than the expected one (77%). It has to be mentioned that the expected value for the diagnostic specificity was calculated based on the current sample set and the assumed number of possible misidentifications between *B. dorsalis* and *B. carambolae* using DNA barcoding standard only.
The values of the performance characteristics showed the sequencing of the *COI* locus cannot fully discriminate all listed species. The test was found to be 100% inclusive for *B. dorsalis* from Laos, Mali, Senegal, Taiwan, Thailand and Vietnam.
For the exclusivity several non-targets were tested (including *B. dorsalis* complex species: *B. carambolae, B. caryeae, B. kandiensis, B. occipitalis, B. pyrifoliae*, and other non-targets: *B. oleae, Dacus ciliatus, Zeugodacus cucurbitae, Anastrepha obliqua, Anastrepha suspensa*). Only one *B. kandiensis* sample (16) was misidentified (false positive).
Contrary to the IPPC standard, in our study it was possible to discriminate *B. carambolae* from *B. dorsalis* and other species of *B. dorsalis* complex using sequence data on *COI* only.

The analytical sensitivity with both primer sets also easily met the expected value of 4 ng/µl. It is noteable, that the analytical sensitivity was higher (up to 100-fold) with LepF/LepR primers for certain matrices (legs and pupa).
For the reproducibility tests specimens of *B. correcta* and *B. latifrons* were included. The reproducibility of the PCR tests using two different primer sets and reproducibility of the SANGER sequence analysis were 100% in all cases.
The performance characteristics of the repeatability and the analytical sensitivity were different. Whereas the repeatability for the amplicon production of LepF/LepR primer set was 100%, the repeatability for the LCO1490/HCO2198 primer set was only 91.66%.

Appendix 7 of this document shows the results for diagnostic specificity.
Appendix 8 shows the results for analytical sensitivity, repeatability and reproducibility.
Appendix 9 shows the calculations for the performance characteristics.

Table 7: Summary of the results obtained for the molecular protocol – EPPO PM7/129 (2), Appendix 1, *COI* gene locus.

Performance criteria	Definition	Calculation	Expected performance characteristics (EPPO, 2021b)	Obtained performance characteristics for sequencing of *COI* (primer set LCO1490/HCO2198)	Obtained performance characteristics for sequencing of *COI* (primer set LepF/LepR)
Diagnostic specificity	The proportion of non-target samples (true negatives) testing negative compared with results from an alternative test (or combination of tests)	Diagnostic specificity = true negatives/(true negatives + false positives)	**77%**	**96%**	**95%**
Analytical specificity	**Inclusivity:** The performance of a test with a range of target organisms covering genetic diversity, different geographical origin and hosts	-	-	**Laos Mali Senegal Taiwan Thailand Vietnam**	**Laos Mali Senegal Taiwan Thailand Vietnam**
	Exclusivity: The performance of a test with regards to cross-reaction with a range of non-targets (e.g. closely related organisms)	-	-	***B. carambolae B. caryeae B. kandiensis B. occipitalis B. pyrifoliae B. oleae Dacus ciliatus Zeugodacus cucurbitae Anastrepha obliqua Anastrepha suspensa B. latifrons B. correcta***	***B. carambolae B. caryeae B. kandiensis B. occipitalis B. pyrifoliae B. oleae Dacus ciliatus Zeugodacus cucurbitae Anastrepha obliqua Anastrepha suspensa B. latifrons B. correcta***
Diagnostic sensitivity	The proportion of target samples (true positives) testing positive compared with results from an alternative test (or combination of tests)	Diagnostic sensitivity = true positives/(true positives + false negatives)	**100%**	**100%**	**100%**
Analytical sensitivity	The smallest amount of target that can be detected reliably	-	**4ng/µl**	**0.325ng/µl**	**0.325/µl**
Repeatability	The level of agreement between replicates of a sample tested under the same conditions	% level of agreement	**100%**	**91.66%**	**100%**
Reproducibility	The ability of a test to provide consistent results when applied to aliquots of the same sample tested under different conditions (e.g. time, persons, equipment, location)	% level of agreement	**100%**	**100%**	**100%**
Accuracy	The proportion of target samples (true positives) testing positive and non-target samples (true negatives) testing negative compared with the total number of samples	Accuracy = (true positives + true negatives)/(true positives + false negatives + true negatives + false positives)	**85%**	**97%**	**97%**

In silico testing of analytical specificity by a database alignment (NCBI GenBank) was performed with the *ITS*1 sequencing primer set (*ITS*6/7). The search set was limited to Tephritidae. The results showed suitablitiy of the primer set (see App. 7) for identification of *B. dorsalis*.

Sequencing of *ITS*1 internal transcribed spacer (*ITS*6/7 primer set)

This test was not applied for samples which beforehand were unambiguously identified as *Anastrepha spp., B. oleae, Dacus ciliatus* or *Zeugodacus cucurbitae* by barcoding. *Bactrocera pyrifoliae* was excluded from the analysis, because no amplicon could be obtained with sample 21 (only sample of *B. pyrifoliae* in this study). We analysed all samples belonging to the *B. dorsalis* complex according to the *COI* barcoding results. This is not in line with the described procedure of the IPPC standard, as the *ITS*6/7 sequencing is only recommended as a follow up step after morphological identification of adult specimens. It is described as the tool for molecular discrimination of *B.dorsalis* and B. *carambolae*.

The values obtained for diagnostic sensitivity, analytical sensitivity, reproducibility and repeatability met the expected performance characteristics of 100% (Table 8). The analytical sensitivity was very high (amplification could be achieved with all samples at a 1: 100.000 dilution (3.25pg/µl). The test was found to discriminate *B. dorsalis* from *B. carambolae.* However, in the case of sample 38 (assigned as *B. caryeae*) the sample could be discriminated from *B. carambolae* (no 44bp insert). Nevertheless, there was a match with *B. dorsalis* (NCBI GenBank query), resulting in a false positive result. This false positive result obtained for sample 38 (summarized in Table 8) also influenced the performance characteristics of the diagnostic specificity and accuracy. Therefore, the value obtained for diagnostic specificity and accuracy did not meet the expected value of 100%, but reached a value of 93% and 96% respectively.
In our study all samples of *B. kandiensis* could unambiguously identified with the *ITS*1 locus.

Appendix 7 of this document shows the results for diagnostic specificity.
Appendix 8 shows the results for analytical sensitivity, repeatability and reproducibility.
Appendix 9 shows the calculations for the performance characteristics.

Table 8: Summary of the results obtained for the molecular protocol – IPPC ISPM 27 DP29, *ITS*1

Performance criteria	Definition	Calculation	Expected performance characteristics	Obtained performance characteristics
Diagnostic specificity	The proportion of non-target samples (true negatives) testing negative compared with results from an alternative test (or combination of tests)	Diagnostic specificity = true negatives/(true negatives + false positives)	**100%**	**93%**
Analytical specificity	**Inclusivity:** The performance of a test with a range of target organisms covering genetic diversity, different geographical origin and hosts	-	**100%**	**100%**
	Exclusivity: The performance of a test with regards to cross-reaction with a range of non-targets (e.g. closely related organisms)	-	-	***B. carambolae*** ***B. kandiensis*** ***B. occipitalis***
Diagnostic sensitivity	The proportion of target samples (true positives) testing positive compared with results from an alternative test (or combination of tests)	Diagnostic sensitivity = true positives/(true positives + false negatives)	**100%**	**100%**
Analytical sensitivity	The smallest amount of target that can be detected reliably	-	**4ng/µl**	**3.25pg/µl**
Repeatability	The level of agreement between replicates of a sample tested under the same conditions	% level of agreement	**100%**	**100%**
Reproducibility	The ability of a test to provide consistent results when applied to aliquots of the same sample tested under different conditions (e.g. time, persons, equipment, location)	% level of agreement	**100%**	**100%**
Accuracy	The proportion of target samples (true positives) testing positive and non-target samples (true negatives) testing negative compared with the total number of samples	Accuracy = (true positives + true negatives)/(true positives + false negatives + true negatives + false positives)	**100%**	**96%**

6.2.1 *Analysis of critical issues in the molecular identification of B. dorsalis*

DNA barcoding and *ITS* sequencing

a. Quality of DNA and consensus sequence

Accurate comparison of sequences usually requires reliable consensus sequences. Due to the low phylogenetic resolution of some *Bactrocera* species, it was indispensable to work with high quality of DNA and consensus sequences (use sequence data from forward and reverse reaction for assembly, trimmed consensus sequences with correct orientation, length of the consensus sequences close to the expected value).

In this study we had to work on one leg per specimen (usually from a collection of long stored specimens), which resulted in low quality and/or fragmented DNA for some samples. In the case of sample 21, the DNA extraction failed in spite of several extraction repetitions (even using non destructive DNA extraction on the entire specimen), for other samples the quality of the DNA was poor (e.g. sample 33 and 38). Sample 40 was contaminated by human DNA.

This sometimes resulted in no amplification products (sample 21) or in bad or short consensus sequences (e.g. sample 16). For sample 16 the database alignment led to a false positive result (see chapter b). The database alignment of the *ITS*1 consensus sequence was correct (*B. kandiensis*). This is an example of poor DNA quality masking the *COI* result. Only *ITS*1 could resolve this sample originally assigned to *B. kandiensis*.

Two samples assigned as *B. caryeae* (sample 33 and 38) could not be resolved correctly although DNA extractions and PCR amplifications were repeated several times. DNA barcoding identified both samples as *B. carambolae*. This identification was wrong, but had no impact on the performance characteristics (true negative). Additionally, identification using *ITS* sequencing led to no consensus sequence (sample 33) and a false positive result (sample 38).

The contaminated sample 40 led to wrong results with barcoding. The more specific *ITS*1 sequencing allowed correct molecular identification of the sample to the assigned value.

In the light of these results it is important for the routine diagnosis to use adequate DNA extraction procedures especially for sequencing techniques (see EURL verification report: Verification of DNA Extraction Procedures for Insects 2021 (https://sitesv2.anses.fr/en/minisite/insects-and-mites/approved-reagents).

b. Availability and reliability of sequence data in NCBI GenBank, Bold and Q-Bank

During our validation study the number of correct hits and ability to exclude incorrect hits were a critical issue for identification of *Bactrocera* species.

***COI* gene sequences**

At the time of our query NCBI GenBank and Bold have the highest number of deposited sequence data for the selected *Bactrocera* species). Contrary to the IPPC standard mentioning zero entries for *B. pyrifoliae*, one entry of the *COI* gene has meanwhile been made available for this species.

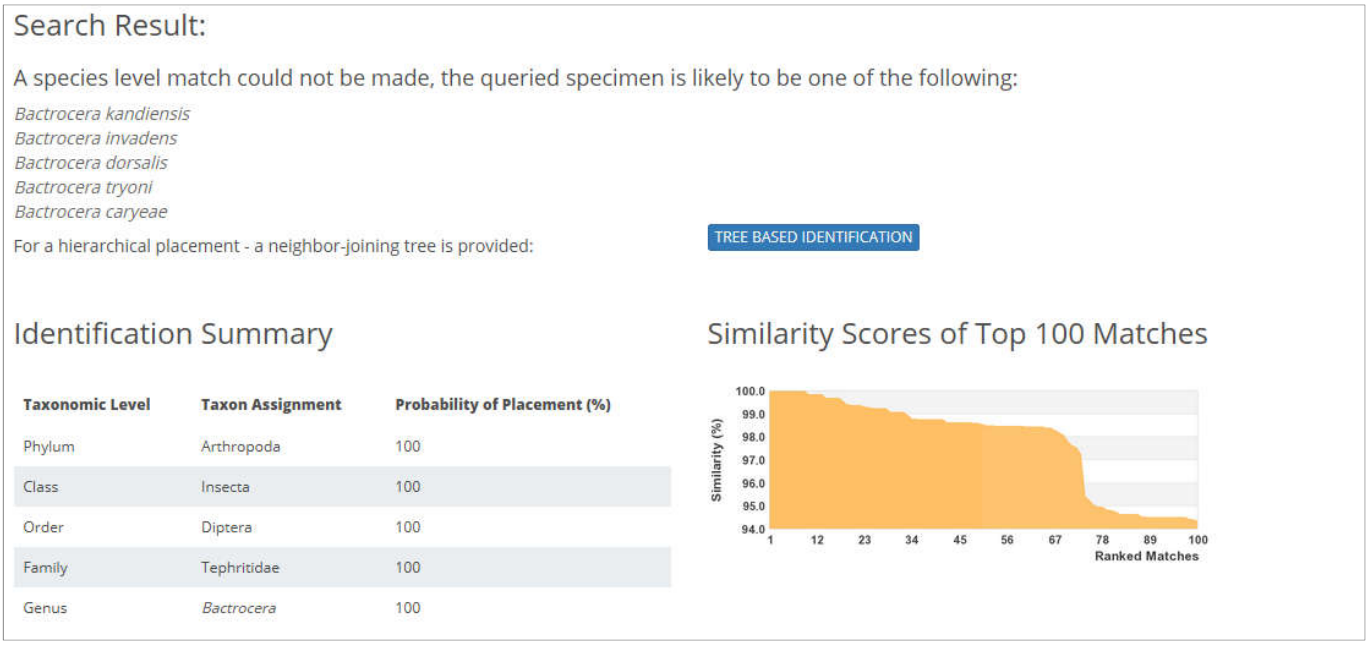
The EPPO Q-Bank currently lacks *COI* gene sequences of relevant species for *B. dorsalis* identification (e.g. no *B. kandiensis* sequences are available). Nevertheless, the sequence data included are obtained from properly documented and identified specimen (see Table 9). The Bold database has a comprehensive number of *COI* sequences and questionable results are indicated (see Fig. 6). The query in the NCBI GenBank also revealed a high number of sequences, but there is no information on the search result.

Therefore and due to quality assurance reasons, the database alignment of this study was performed in three different databases (NCBI GenBank, Bold and Q-Bank). In addition, reference alignment was performed using Geneious prime® 10.1.3..

Table 9: Numbers of *COI* gene sequences of relevant *Bactrocera* species represented in different databases – (Query from 22 February 2021).

Species	NCBI GenBank	Bold	EPPO Q-Bank
B. dorsalis	5384	7314	412
B. kandiensis	34	132	0
B. carambolae	219	408	6
B. pyrifoliae	1	0	0
B. occipitalis	82	239	0
B. caryeae	14	15	0

Figure 6 - Inconclusive results indicated by Bold database



An example for possible misidentification due to lack of sequence can be shown with barcoding results of sample 16. The *COI* consensus sequence of sample 16 was aligned in all three databases. NCBI GenBank gave three equal hits of *B. dorsalis* and *B. kandiensis* (equal scores, query covers, E-values and percentage identities). Since the *B. dorsalis* accession comprises a sequence of 676bp instead of 658bp, it is ordered above the *B. kandiensis* accessions. The database alignment in Bold revealed *B. kandiensis* only, but it was indicated that a species level match could not be made (see Fig. 6). In EPPO Q-Bank the query resulted in *B. dorsalis*. Because, two out of three database alignments resulted in *B. dorsalis*, the final judgment for this sample was *B. dorsalis*. Our false positive assignment can mainly be ascribed to the result of the BLAST search in NCBI GenBank and to the false result in EPPO Q-Bank at the date of query. At this time there were no sequences for *B. kandiensis* deposited in EPPO Q-Bank, which might be the reason for the false assignment.

Figure 7 - Blast results of NCBI GenBank, Bold and EPPO Q-Bank for sample 16.

Database	Result	Documentation
NCBI GenBank	*Organism: Bactrocera dorsalis* Accession Nb.: MK314052.1 %identiy: 100% e-value: 0.0 Score:1029	
Bold	Organism *Bactrocera kandiensis* Accession Nb.: GBMIN62999-17 %identiy: 100%	
EPPO Q-Bank	Organism *Bactrocera dorsalis* Accession Nb.: CO1/VBAL_1001036 %identiy: 100%	

NCBI GenBank

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident
Bactrocera dorsalis isolate Ugb5 cytochrome c oxidase subunit 1 (COI) gene, partial cds; ...	Bactrocera d...	1029	1029	100%	0.0	100.00%
Bactrocera kandiensis mitochondrial COI gene for cytochrome oxidase subunit 1, partial c...	Bactrocera k...	1029	1029	100%	0.0	100.00%
Bactrocera kandiensis voucher Bk4 cytochrome oxidase subunit 1 (COI) gene, partial cds...	Bactrocera k...	1029	1029	100%	0.0	100.00%
Bactrocera kandiensis voucher Bd1590 cytochrome oxidase subunit I (COI) gene, partial ...	Bactrocera k...	1024	1024	100%	0.0	99.82%
Bactrocera invadens voucher BiSLT4.1 cytochrome oxidase subunit 1 (COI) gene, partial ...	Bactrocera d...	1024	1024	100%	0.0	99.82%
Bactrocera kandiensis voucher Bk7 cytochrome oxidase subunit 1 (COI) gene, partial cds...	Bactrocera k...	1024	1024	100%	0.0	99.82%

Bold

Phylum	Class	Order	Family	Genus	Species	Subspecies	Similarity (%)	Status
Arthropoda	Insecta	Diptera	Tephritidae	*Bactrocera*	*kandiensis*		100	Published
Arthropoda	Insecta	Diptera	Tephritidae	*Bactrocera*	*kandiensis*		100	Published
Arthropoda	Insecta	Diptera	Tephritidae	*Bactrocera*	*kandiensis*		100	Published
Arthropoda	Insecta	Diptera	Tephritidae	*Bactrocera*	*kandiensis*		100	Published
Arthropoda	Insecta	Diptera	Tephritidae	*Bactrocera*	*kandiensis*		100	Published
Arthropoda	Insecta	Diptera	Tephritidae	*Bactrocera*	*invadens*		100	Published

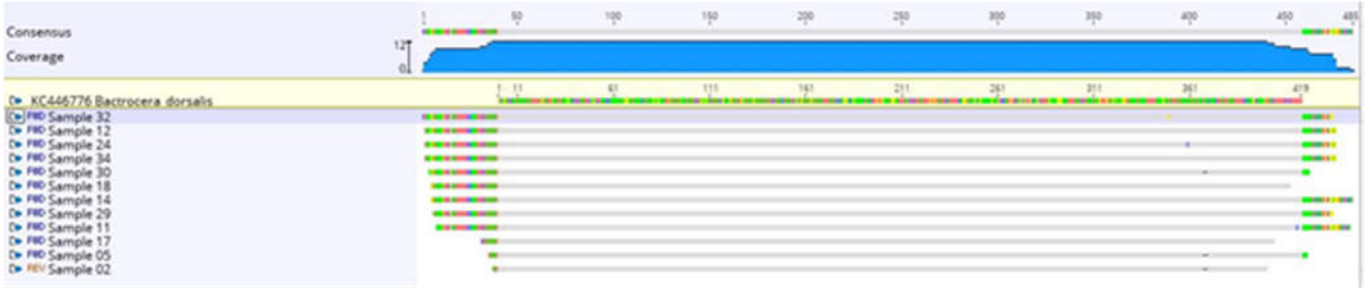
EPPO Q-Bank

#	Reference	Score	Similarity %	Overlap %	Direction
	Search...	Search	Search...	Search...	Search...
1	CO1/VBAL_1001036_COI - Bactrocera dorsalis	884.409	100	100	+/+
2	CO1/VBAL_1000481_COI - Bactrocera dorsalis	881.239	99.82	100	+/+
3	CO1/CCOC09707_0101_COI - Bactrocera dorsalis	881.239	99.82	100	+/+
4	CO1/VBAL_1000295_COI - Bactrocera dorsalis	881.239	99.82	100	+/+
5	CO1/VBAL_1101045_COI - Bactrocera dorsalis	878.069	99.641	100	+/+
6	CO1/VBAL_1101524_COI - Bactrocera dorsalis	878.069	99.641	100	+/+

***ITS* sequences**

The IPPC standard mentions that *B. carambolae* is distinguishable from *B. dorsalis* on the *ITS*1 due to the presence of a unique 44bp insert in *B. carambolae*. However, this was based on the *ITS*1 comparison of only four species in the *Bactrocera dorsalis* complex: *B. dorsalis s.l., B. occipitalis, B. opiliae* and *B. cacuminata.* Guidance is given in the IPPC standard for reference alignments including reference sequences (NCBI accession Nb. KC446737.1 for *B. carambolae* and KC446776.1 for *B. dorsalis*).

Figure 8 - *ITS1* reference alignment: *B. dorsalis* samples of this study aligned to the recommended reference sequence (KC446776.1) according to the IPPC standard.



To elucidate the quality and reliability of the *ITS*1 sequencing, a comprehensive search of available sequences in databases was performed. Sequences on the *ITS*1 region are only available in the NCBI GenBank, but some entries are questionable. It has to be noted that at the date of the query no *ITS*1 sequences for *B. caryeae* and *B. pyrifoliae* (only one *ITS*2) were available.

An example for unreliable entries is the NCBI blast result for sample 38 (assigned value is *B. caryeae*):
First hit with this sample was *Bactrocera carambolae* "voucher" (accession number KJ544953.1 Guangzhou, P.R. China), followed by two hits for *B. dorsalis* with equal percent identity and score values (see Fig. 9). However, after a reference alignment with the specific *B. carambolae* insert, the sequence deposited lacked the specific insert (see Fig. 10). Therefore, the sample was determined as *B. dorsalis*, which finally was false positive.
No reference alignment could be performed for *B. caryeae* due to unavailability of *ITS* data.

Figure 9 - BLAST result in NCBIGenbank for *ITS*1 sequencing of sample 38. Three equal hits *B. carambolae* (1) and *B. dorsalis* (2) were obtained.

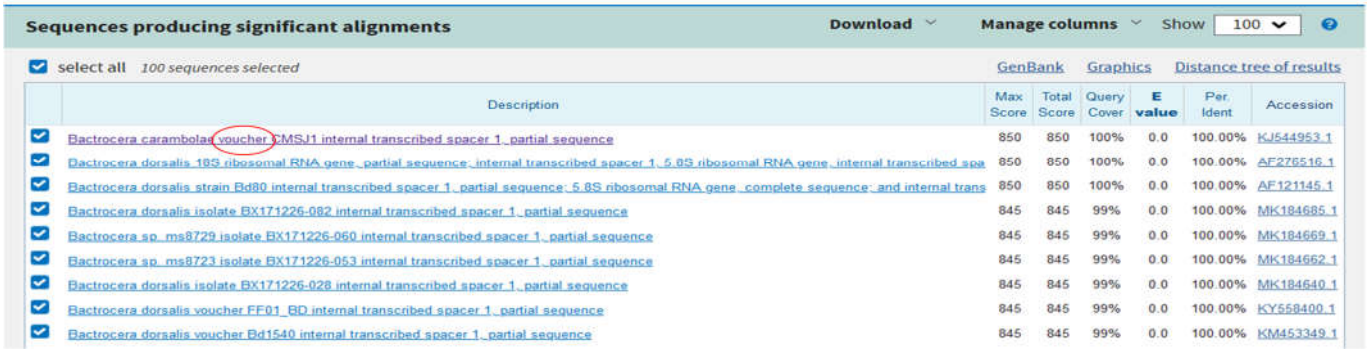


Figure 10 - *ITS*1 reference alignment: One each unambiguously identified *B. carambolae* (sample 11) and *B. dorsalis* (sample 08) from this study and sample 38 (ambiguous sample), and the NCBI accession numbers KJ544953.1 (deposited as *Bactrocera carambolae* "voucher") aligned to the reference sequence for *B. carambolae* KC446737.1 (IPPC standard). Sample 38 lacks the specific insert and was therefore assigned as *B. dorsalis* (false positive). NCBI accession number KJ544953.1 (deposited as *Bactrocera carambolae* "voucher") lacks the specific insert and misidentification is highly probable.

c. Specific deviations and issues in *ITS*1 sequencing

For the reliable discrimination of *B. carambolae* from *B. dorsalis* it is necessary to analyse the presence or absence of a 44bp insert near the *ITS*7 primer binding site. This insert is only present in *B. carambolae* and therefore specific for this species (IPPC 2019). However, there is an editing mistake in the sequence insert for *B. carambolae* displayed in the IPPC protocol (chapter 4.3.4). Here the insert consists only of 41bp (3 bases at the 5' end are missing) and should be corrected (Fig. 11).

Figure 11 - *ITS*1 Reference alignment: Several *B. carambolae* samples from this study and sequence of insert of *B. carambolae* according to the IPPC standard aligned to the reference sequence (accession Nb. KC446737.1). All of the aligned sequences contain the characteristic *B. carambolae* insert, except for sample 38 (assigned value *B. caryeae*). The red box indicates three base-pairs missing from the aligned insert in comparison to the other sequences.

Uncertainties in original assignment of specimens used in the validation sample set

In the case of one sample of the set (sample 23), the morphological and molecular identification results obtained did not allow to confirm the a priori assigned value of the sample. Sample 23 represents a very interesting and controversial case. The specimen, originally from the Philippines, was given to AGES as *Bactrocera occipitalis* and as such was included in the sample set of this validation study. However, molecular analyses indicate that the specimen is a *Bactrocera carambolae* (Fig. 12 and 13), possessing the 44bp insert that is unique and characteristic of this species (Fig. 13). In addition, identification by the three operators gave conflicting results (2 'negative results' and one 'not determined', see Appendix 4). To try to shed light on the case, a new morphological analysis of the specimen was conducted (Appendix 10). On the basis of this analysis, given the non-concordance of the results of the three operators, the specimen cannot be assigned with certainty to *B. occipitalis* on a morphological level. In cases like this it is necessary to identify the specimen as generically belonging to the *B. dorsalis* complex. To our knowledge, there are no data to date on the possible introgression of mitochondrial DNA from *B. carambolae* into specimens of B*. occipitalis*, which has been hypothesised in the case of intogression of *B. kandiensis* mitochondrial DNA into specimens of *B. dorsalis* (Schutze et al., 2015b).
For the purpose of this validation study, specimen 23 remains negative with respect to *B. dorsalis s.l.* and the uncertainty associated with its assigned value does not affect the performance characteristics.

Figure 12 - *COI* reference alignment: *COI* sequence of *B. carambolae* reference specimen (accession Nb. KC446059.1) aligned with sample 39 (unambiguous B. carambolae from this study) and sample 23. It is visible that they are identical on the *COI* gene locus.

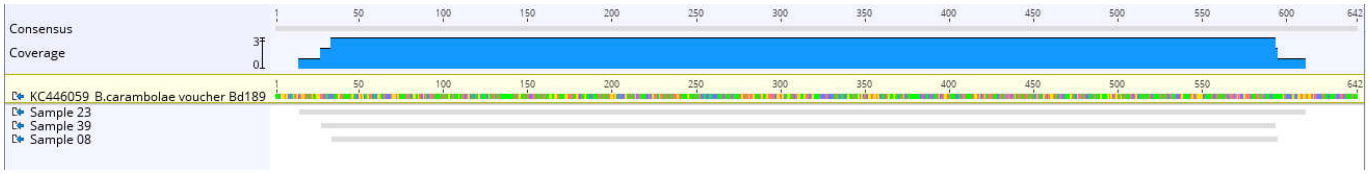
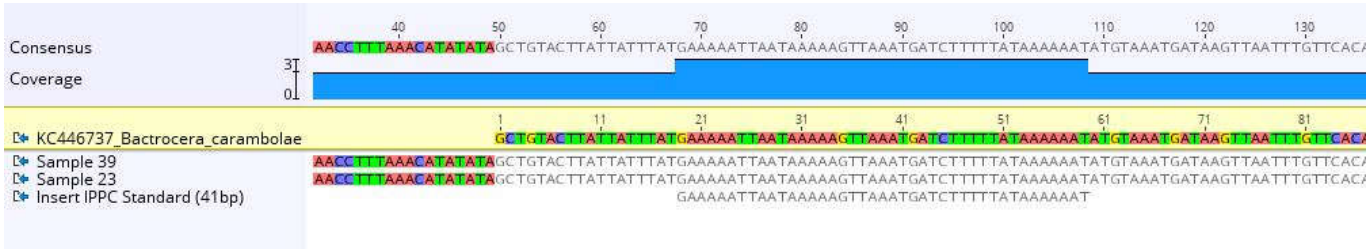


Figure 13 - *COI* reference alignment: Alignment of a *B. occipitalis* sequence mined from NCBI GenBank (accession Nb. KM023416.1) with sample 23. Sample 23 is different to *B. occipitalis* on the *COI* gene locus.

Figure 14 - *ITS* reference alignment: Sequences of sample 8, 39 and 23 aligned to the *ITS* sequence with the accession Nb. KC446737.1 (mined from NCBI GenBank according to the IPPC Standard) and the characteristic *B. carambolae* insert (IPPC Standard). It is visible that the sequences are identical to each other and do not differ from the described insert.



7. Discussion and conclusions

This study aimed at the validation of IPPC and EPPO diagnostic protocols for the morphological and molecular identification of *Bactrocera dorsalis s.l.*. The study has involved staff of the EURL for Insects and Mites from ANSES and AGES and the analytical activities have been carried out from May to August 2020. A main sample set of 40 Tephritidae specimens, including target and non-target species, has been used. Additionally, smaller sample sets have been prepared for validating the molecular tests. The drafting of the final report has taken a longer time due to necessary consultation with IPPC DP 29 authors and international Tephritidae experts.

Morphological diagnostic protocol

The morphological identification according to the diagnostic protocol IPPC DP 29 (IPPC, 2019) achieved the expected value of 100% for the validation criteria diagnostic specificity, diagnostic sensitivity, accuracy and repeatability. However, reproducibility obtained a value of 87,5% due to divergent results between the three operators performing the identification. After an in-depth cause analyses, it was assessed that these divergent results were originated from the description of some diagnostic characters in the protocol (key and/or table of characters) that may lead the user to misinterpretation and consequently to a wrong identification. To summarize,

- the term "mid-point" in the description of *Bactrocera occipitalis* costal band could be misleading as it refers to a point; it should be interpreted as a "mid-line" between wing venation R_{2+3} and R_{4+5} (clarification by Norman Barr and Camiel Doorenweerd);
- difference between the shape of costal band of *Bactrocera occipitalis* ("distinctly overlapping R_{2+3}") and *Bactrocera carambolae* ("slightly overlapping R_{2+3}") is not sufficiently clear from Figures 16 (a) and 16 (e);
- adverb "generally" referred to character "confluent costal band" in *Bactrocera dorsalis s.l.* is ambiguous and should be rather interpreted as "typically" confluent (clarification by Norman Barr and Camiel Doorenweerd);
- description of *Bactrocera carambolae* character "tergite 3" does not take into account possible variation ("non-continuous transverse band"), which on the contrary is documented in FruitFly ID Australia (Plant Health Australia, 2021).

Based on these results, the EURL recommends the use of the IPPC DP 29 (IPPC, 2019) to EU National Reference Laboratories for the morphological identification of *Batrocera dorsalis s.l.*. with some advice for the correct use of the diagnostic protocol:

- "mid-point" in the description of *Bactrocera occipitalis* costal band should be interpreted as a "mid-line" between wing venation R_{2+3} and R_{4+5} (see Fig. 2). Position in the document: page 11, Table 3 (4.2.3); page 12, Diagnostic key (4.2.4), couplet 3;
- in the Diagnostic key, couplet 3, decision between *B. occipitalis* and *B. carambolae* should be taken on the basis of all the diagnostic characters included in Table 3, not only on the basis of the shape of costal band. Position in the document: Diagnostic key (4.2.4), couplet 3, page 12;
- adverb "generally" referred to character "confluent costal band" in *Bactrocera dorsalis s.l.* should be rather interpreted as "typically" confluent. "Generally" may be misinterpreted as that a differently-shaped (e.g. overplapping) costal band is sometimes present in *B. dorsalis s.l.*. The diagnosis of *B. dorsalis* requires confluence of costal band. Position in the document: page 11, Table 3 (4.2.3);
- variation of the character "tergite 3 - with a narrow transverse black band across anterior margin (constituting a "T" pattern)" for *Bactrocera carambolae* should be considered, even if not mentioned in the document; a non-continuous transverse band on tergite 3 can sometimes be found (see Fig. 4 and 5).

In addition, it is very important to remind that the Diagnostic key serves as a first screening tool and final decision about the identification should rely on the careful examination of all the characters in Table 3 (possibly with the only exception of genitalia, see 4.2.3 in the DP 29). This is stated in the diagnostic protocol itself: "*An identification to one of the six species in the protocol requires the adult specimen to be examined for the characters provided in Table 3. This can be accomplished using the key in section 4.2.4 to screen specimens and then identification can be confirmed by comparing fly morphology to information in Table 3.*" "*If one or more characters are inconsistent between the specimen and the descriptions provided in Table 3, then the specimen cannot be diagnosed as one of these species"* and identification should be limited to ***Bactrocera dorsalis* complex**.

The present validation study has generated useful elements to improve the morphological part of the diagnostic protocol DP29. Therefore the authors of this report suggest the following points for improvement of the DP 29 to the IPPC bodies and the authors involved:

- a figure (i.e. Fig. 2) could be very useful to correctly interprete "mid-point" in the description of *Bactrocera occipitalis* costal band and showing that it is actually a "mid-line";

- if possible, a second diagnostic character could be very useful to distinguish between *Bactrocera occipitalis* and *Bactrocera carambolae* in couplet 3 of the Diagnostic key (e.g. dark markings on the abdomen?) as the character "costal band" alone hardly allows the discrimination between the two species (Fig. 16 (a) and (e) do not allow a certain interpretation of "distinctly" and "slightly" overlapping); however, this might be challenging since the alternative choice to *B. occipitalis* in couplet 3 leads to further couplets which consider three other species of the complex showing different features (*B. pyrifoliae, B. carambolae, B. dorsalis* s. l.);
- the adverb "generally" as referred to character "confluent costal band" in *Bactrocera dorsalis s.l.* should be preferably replaced by "typically";
- variation of the character "tergite 3" for *Bactrocera carambolae* should be mentioned in the protocol (non-continuous transverse band on tergite 3), as from FruitFly ID Australia (Plant Health Australia, 2021).

Molecular diagnostic protocols

This validation study aimed to generate the performance characteristics including molecular identification of *B. dorsalis* using DNA barcoding (*COI*) as well as *ITS*1 sequencing.

According to the IPPC standard (2019) **molecular methods alone are not recommended** for the identification of the six economically most relevant *B. dorsalis* complex species. However, the IPPC protocol recommends molecular tools for the discrimination of *B. dorsalis* and *B. carambolae* specimens after morphological determination (*ITS*1 sequencing).

In routine diagnosis, especially when dealing with larvae e.g. in the frame of import control, molecular identification is sometimes the only available method and therefore the EPPO DNA barcoding standard (EPPO, 2021b) was also validated. Hence, *COI* barcoding was applied as first line identification.

The experience gathered in this study was that for samples from which reasonable DNA quality could be extracted, the identification was quite straightforward for all samples not belonging to the *B. dorsalis* complex. All these samples could be identified at least at genus level using EPPO barcoding standard only. All samples belonging to the *B. dorsalis* complex were subsequently analysed with *ITS*1-sequencing.

It has to be considered that molecular identification via sequence analysis is a multistep process (DNA extraction, PCR, sequencing and sequence analysis). Performance characteristics were elaborated for PCR and sequence analysis steps.

***COI* sequencing:** The performance characteristics of the diagnostic specificity and accuracy displayed that sequencing of the *COI* locus cannot fully discriminate all listed species. Nevertheless, the obtained values (96% for the LCO1490/HCO2198 primer set and 95% for the LepF/LepR primer set) were higher than the expected values, which had been calculated as 77%. Samples where no amplicons could be generated at all were excluded, whereas lacking amplicons in one of the two primer sets were assigned as negative deviations. All *B. dorsalis* assigned samples from different geographic origins could be correctly identified. In regards to the exclusivity cross-reactions could be observed with one sample (sample 16), which was assigned to *B. kandiensis* (*ITS1*) and misidentified as *B. dorsalis* on the *COI* locus.
The results of this study also showed that the analytical sensitivity of both primer sets (0.325ng/µl) was below the expected value (4ng/µl). However, the value for the repeatability of LCO1490/HCO2198 primer set was lower than the expected 100%.

***ITS*1 sequencing** was proven to be a valid confirmatory tool for *B. dorsalis* and *B. carambolae*. *B. carambolae* could be identified and clearly distinguished from *B. dorsalis* complex.
Most of the performance characteristics met the expected values except the values for the diagnostic specificity and accuracy which were below 100%. One sample (sample 38) lacked the *ITS*1 insert and was misidentified as *B. dorsalis s.l.*. According to the IPPC standard the lack of the insertion and a match to *B. dorsalis s.l.* cannot exclude other species in the *B. dorsalis* complex. However, in our study, barcoding in combination with *ITS*1 sequencing could accurately identify *B. dorsalis s.l.* in all cases, if the results on both loci (*COI* and *ITS*1) were congruent. If the results deviated between the loci (sample 38 and 16), *ITS*1 sequencing was more reliable. Contrary to the IPPC protocol, this study shows that molecular identification of *B. kandiensis* with *COI* can be confirmed by *ITS*1 sequencing. We assume that this could also be the case for *B. occipitalis*, but due to the lack of further specimens of this species, this could not be confirmed in this study. The possible suitability of the species identification of *B. caryeae* and *B. pyrifoliae* based on *ITS*1 could not be evaluated since no *ITS*1 sequences are available in the databases.

Several **critical issues** during this validation, which need to be addressed.

Firstly, DNA quality is important for the success of subsequent sequence analysis. This is highly dependent on the quality and the yield of the sample tissue. In some cases DNA quality was not suitable for a successful molecular identification process, even upon repeated extractions (singular legs and/or non-destructive DNA extractions from specimens from collections).

According to the EPPO DNA barcoding standard, when identifying unknown samples via barcoding, the choice of sequence database has a great impact on the results. Different databases utilize a different combination of nucleotide similarity, tree clustering et cetera, with varying focus on the similarity, query cover and the like.

Availability of sequence data in NCBI GenBank, Bold and EPPO Q-Bank differed greatly and affected the results. This has to be taken into account when using these databases in routine diagnosis. If for example no sequence data are available in the database this can lead to a false result. In the case of sample 16, the lack of sequences for *B. kandiensis* in EPPO Q-Bank at the time of query led to an incorrect barcoding result, in addition to the poor DNA quality of this sample. The different ways to display the results is also noteable: Bold database provides a preliminary result and indicates inconclusive results, contrary to NCBI GenBank and EPPO Q-Bank databases, which depict only hits.

Furthermore, the reliability of the deposited accessions is not always given (e.g. *B. carambolae* accessions like Nb. KJ544953.1 shown in this study or Nb. KF998794.1 according to Manger *et al.* 2017). This might be due to the reason that voucher specimens for generating the barcodes have been wrongly identified (Manger *et al.* 2017).

In the case of one sample of the set (sample 23), the morphological and molecular identification results diverged. This sample originated in the Philippines. The original assignment was *Bactrocera occipitalis*, nevertheless, in this validation study some uncertainties occurred during the morphological identification process. The assigned value remains unclear. This does not influence the performance characteristics of this validation study, as the true negative stands correct.
During the molecular validation for sample 23, a clear deviation from the assigned value (B*. occipitalis*) could be recorded. DNA barcoding (*COI*) as well as sequencing of the *ITS*1 unambiguously resulted in *B. carambolae*. This included the presence of the 44bp insert near the *ITS*7 primer binding site that is described in the IPPC standard and could be observed in this study for *B. carambolae* only.

Recommendation / Conclusion

The choice of the DNA extraction procedure is a very important first step when applying sequence-based molecular methods (https://eurl-insects-mites.anses.fr/en/minisite/insects-and-mites/approved-reagents).
If only molecular methods are used for identification, it is recommended to perform the diagnosis stepwise. In a first step, DNA barcoding should be used for discrimination of species not included in the *B. dorsalis* complex. *ITS*1 sequencing can be applied as a confirmatory step, to discriminate *B. carambolae* from all other species of the *B. dorsalis* complex and to possibly increase the resolution within the complex.

Sanger sequence analysis requires adequately proficient operators and the employment of multiple online resources. Caution is necessary when evaluating the hits achieved in various databases, as single sequences might be questionable (e.g. so-called voucher sequences) and the lack of sequences for some species leads to false hits altogether. In addition to database alignment, we therefore recommend to perform a reference sequence alignment.

A follow up study, including newly generated *ITS*1 sequences of *B. caryeae, B. occipitalis* and *B. pyrifoliae* is planned. The aim is to generate approved sequence data (*COI* and *ITS*1) on to date underrepresented economically relevant *Bactrocera* species from the complex and make them available, e.g. via EPPO Q-Bank.

Date: 15 October 2021

Phillipe Reynaud
EURL Director

Helga Reisenzein
EURL Deputy Director

Acknowledgements

The team wants to thank Marc de Meyer (Royal Museum for Central Africa, Tervuren, Belgium), Norman Barr (United States Department of Agriculture, United States of America), Luc Leblanc (University of Idaho, United States of America) and Camiel Doorenweerd (University of Hawai'i at Mānoa, United States of America) for their valuable advice as international experts on the genus *Bactrocera*, which was an essential contribution to this study.

References

Boykin, L.M., Schutze, M.K., Krosch, M.N., Chomič, A., Chapman, T.A., Englezou, A., Armstrong, K.F., Clarke, A.R., Hailstones, D. and Cameron, S.L. (2014). Multi-gene phylogenetic analysis of south-east Asian pest members of the *Bactrocera dorsalis* species complex (Diptera: Tephritidae) does not support current taxonomy. *Journal of Applied Entomology*, 138: 235-253.

CABI (2021). Datasheet *Bactrocera dorsalis* (Oriental fruit fly). Available online. https://www.cabi.org/isc/datasheet/17685#tocontributors

Chen, P. & Hui, Y. (2007). [Advances in taxonomy of *Bactrocera dorsalis* complex.] *Kunchong Zhishi*, 44(1): 41–47 (in Chinese).

Clarke, A.R., Armstrong, K.F., Carmichael, A.E., Milne, J.R., Raghu, S., Roderick, G.K. & Yeates, D.K. (2005). Invasive phytophagous pests arising through a recent tropical evolutionary radiation: The *Bactrocera dorsalis* complex of fruit flies. *Annual Review of Entomology*, 50: 293–319.

Doorenweerd, C., Leblanc, L., Norrbom, A. L., 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.

Drew, R.A.I. & Romig, M.C. (2016). *Keys to the tropical fruit flies (Tephritidae: Dacinae) of southeast Asia*. Wallingford, UK, CABI. 496 pp.

Egartner, A., Lethmayer, C., Gottsberger, R. A., & Blümel, S. (2019). Survey on *Bactrocera* spp.(Tephritidae, Diptera) in Austria. *Bulletin OEPP/EPPO Bulletin*, 49(3): 578-584.

EPPO (2014). EPPO standards PM 7/122 (1) Guidelines for the organization of interlaboratory comparisons by plant pest diagnostic laboratories. *Bulletin OEPP/EPPO Bulletin*, 44(3): 390-399.

EPPO (2018a). EPPO standards PM 7/76 (5) Diagnostics. Use of EPPO Diagnostic Standards. *Bulletin OEPP/EPPO Bulletin* 48 (3): 373–377.

EPPO (2019). EPPO standards PM 7/98 (4) Diagnostics. Specifics requirements for laboratories preparing accreditation for plant pest diagnostic activity. *Bulletin OEPP/EPPO Bulletin* 49 (3): 530-563.

EPPO (2021a) *Bactrocera dorsalis*. EPPO datasheets on pests recommended for regulation. Available online. https://gd.eppo.int

EPPO (2021b). EPPO standards PM 7/129 (2) DNA barcoding as an identification tool for a number of regulated pests. *Bulletin OEPP/EPPO Bulletin*, 51 (1): 100–143.

Folmer O, Black M, Hoeh W, Lutz R & Vrijenhoek R (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine, Biology and Biotechnology* 3: 294–299.

Hajibabaei, M., Janzen, D. H., Burns, J. M., Hallwachs, W., & Hebert, P. D. (2006). DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences*, 103(4): 968-971.

IPPC (2019). ISPM 27 Diagnostic protocols for regulated pests DP 29: *Bactrocera dorsalis.*

Leblanc, L., San Jose, M., Barr, N. & Rubinoff, D. (2015). A phylogenetic assessment of the polyphyletic nature and intraspecific color polymorphism in the *Bactrocera dorsalis* complex (Diptera, Tephritidae). *ZooKeys*, 540: 339–367.

Manger, A., Behere, G.T., Firake, D.M., Sharma, B., Deshmukh, N.A., Firake, P.D. *et. al.* (2017). Genetic characterization of *Bactrocera* fruit flies (Diperta:Tephritidae) from Northeastern India based on DNA barcodes. *Mitochondrial DNA Part A*, 29(5), 792-799.

Plant Health Australia (2021). FruitFly Identification Australia (https://fruitflyidentification.org.au/).

Schutze, M.K., Aketarawong, N., Amornsak, W., Armstrong, K.F., Augustinos, A.A., Barr, N., Bo, W. *et al.* (2015a). 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*, 40: 456-471.

Schutze, M. K., Mahmood, K., Pavasovic, A. N. A., Bo, W., Newman, J., Clarke, A. R., ... & Cameron, S. L. (2015b). One and the same: integrative taxonomic evidence that *Bactrocera invadens* (Diptera: Tephritidae) is the same species as the Oriental fruit fly B*actrocera dorsalis*. *Systematic Entomology*,40(2): 472-486.

Schutze, M.K., Bourtzis, K., Cameron, S.L., Clarke, A.R., De Meyer, M., Hee, A.K.W., Hendrichs, J., Krosch, M.N. & Mwatawala, M. (2017). Integrative taxonomy versus taxonomic authority without peer review: The case of the Oriental fruit fly, *Bactrocera dorsali*s (Tephritidae). *Systematic Entomology*, 42: 609–620.

White, I. M., & Elson-Harris, M. M. (1992). *Fruit flies of economic significance: their identification and bionomics*. Wallingford, UK, CABI. 601 pp.

Appendix 1 – Composition of the sample set and codification

Sample codification	New codification	Country of collection	Identification	Notes
0700634_4	**1**	Sri Lanka	*Bactrocera kandiensis*	ANSES extra samples
F20049	**2**	Thailand (Saraburi)	*Bactrocera dorsalis*	AGES
F20050	**3**	USA/Florida	*Anastrepha suspensa*	AGES
0502118_2	**4**	French Guyana	*Bactrocera carambolae*	ANSES sample set
1800896	**5**	Mali	*Bactrocera dorsalis*	ANSES sample set; only 1 specimens
F20043	**6**	Malaysia (Selangor)	*Bactrocera carambolae*	AGES
F20051	**7**	USA/Florida	*Anastrepha suspensa*	AGES
1002478_1	**8**	French Guyana	*Bactrocera carambolae*	ANSES sample set
1800894_1	**9**	Laos	*Bactrocera dorsalis*	ANSES sample set
0700634_3	**10**	Sri Lanka	*Bactrocera kandiensis*	ANSES extra samples
1800897	**11**	Mali	*Bactrocera dorsalis*	ANSES sample set; only 1 specimens
1800894_2	**12**	Laos	*Bactrocera dorsalis*	ANSES sample set
0700634_5	**13**	Sri Lanka	*Bactrocera kandiensis*	ANSES extra samples
1901279_1	**14**	Vietnam	*Bactrocera dorsalis*	ANSES sample set
1901277_2	**15**	Sri Lanka	*Zeugodacus cucurbitae*	ANSES sample set
0700634_2	**16**	Sri Lanka	*Bactrocera kandiensis*	ANSES extra samples
1301340_2	**17**	Senegal	*Bactrocera dorsalis*	ANSES sample set
F20048	**18**	Thailand (Saraburi)	*Bactrocera dorsalis*	AGES
1901277_1	**19**	Sri Lanka	*Zeugodacus cucurbitae*	ANSES sample set
1901064	**20**	France	*Bactrocera oleae*	ANSES sample set
F20047	**21**	Vietnam	*Bactrocera pyrifoliae*	AGES
1002478_2	**22**	French Guyana	*Bactrocera carambolae*	ANSES sample set
F20042	**23**	Philippines	*Bactrocera occipitalis*	AGES
1901279_4	**24**	Vietnam	*Bactrocera dorsalis*	ANSES sample set
F20053	**25**	Mexico/Tapachula /Chiapas	*Anastrepha obliqua*	AGES
1500326_1	**26**	Reunion	*Dacus ciliatus*	ANSES sample set
0502118_1	**27**	French Guyana	*Bactrocera carambolae*	ANSES sample set
1401020_1	**28**	Sri Lanka	*Dacus ciliatus*	ANSES sample set
2000042	**29**	Taiwan	*Bactrocera dorsalis*	ANSES sample set; only 1 specimens
1301340_1	**30**	Senegal	*Bactrocera dorsalis*	ANSES sample set
0700634_1	**31**	Sri Lanka	*Bactrocera kandiensis*	ANSES extra samples
1901279_2	**32**	Vietnam	*Bactrocera dorsalis*	ANSES sample set
F20046	**33**	India (Kerala)	*Bactrocera caryeae*	AGES
1800894_3	**34**	Laos	*Bactrocera dorsalis*	ANSES sample set
F20052	**35**	Mexico/Tapachula /Chiapas	*Anastrepha obliqua*	AGES
1901279_3	**36**	Vietnam	*Bactrocera dorsalis*	ANSES sample set
1901549	**37**	France	*Bactrocera oleae*	ANSES sample set
F20045	**38**	India (Kerala)	*Bactrocera caryeae*	AGES
1002478_3	**39**	French Guyana	*Bactrocera carambolae*	ANSES sample set
F20041	**40**	Philippines	*Bactrocera occipitalis*	AGES

Appendix 2 - Check lists for the morphological analysis

Operator		Date	

A combination of characters to diagnose the *Bactrocera dorsalis* complex (modified from Table 3, IPPC ISPM 27 DP29: *Bactrocera dorsalis)*

Morphological character		**Sample code**									
Head	Face yellow with distinct facial spots present (Figures 9(a), 9(b), 12)										
Scutum	Colour mostly black to mostly red-brown (inter-regionally variable) (Figure 13)										
	Lateral vittae present (Figure 11) and yellowish (Figures 13 and 14)										
	Medial vittae absent (Figure 11)										
Scutellum	Yellowish colour (Figures 1 and 13)										
	With a dark basal band (Figures 11 and 1)										
	Never with other dark patterns (Figure 13)										
Femora	Entirely or mostly fulvous (reddish-yellow or tawny) colour but may possess dark patterns particularly on and around apices (Figure 15)										
Wing	Cells bc and c hyaline (colourless) or, at most, with an extremely pale tint (Figures 10 and 16)										
	Without dense microtrichia covering cells bc and c (Figure 10)										
	Costal band narrow (never confluent with R4+5) (Figure 10)										
	Narrow anal streak present (diagonal marking that is above anal lobe) (Figures 10 and 16)										
Abdomen	With a “T” pattern on tergites 3–5 (Figures 7(a) and 17)										

Comments / Results B. dorsalis complex confirmed? Y / N										

Diagnostic key to six economically important species belonging to the *Bactrocera dorsalis* complex (adult) (modified from key 4.2.4, IPPC ISPM 27 DP29: *Bactrocera dorsalis*)

Key for 6 species from the *Bactrocera dorsalis* complex (adult)		go to (mark the decision; note any comments)										
Morphological character		**Sample code**										
1	**Postpronotal lobe** yellow with dark anteromedial corner (Figures 19(b) and (d))	…	…	…	…	…	…	…	…	…	…	2
	Postpronotal lobe entirely yellow (Figures 19(a), (c), (e), (f))	…	…	…	…	…	…	…	…	…	…	3
2	**Scutum** entirely black (Figure 13(b)), abdominal **tergites 3–5** with broad black dorsolateral markings (Figures 17(b) & 18(b)); **lateral vittae** very narrow (Figure 4(b))	… … …	… … …	… … …	… … …	… … …	… … …	… … …	… … …	… … …	… … …	***B. caryeae***
	Scutum mostly black (Figure 13(d)), abdominal **tergites 3–5** with "T" pattern and **tergites 4–5** with very narrow anterolateral black marking (Figures 17(d) and 18(d)); **lateral vittae** narrow (Figure 4(d))	… … … …	… … … …	… … … …	… … … …	… … … …	… … … …	… … … …	… … … …	… … … …	… … … …	***B. kandiensis***
3	**Costal band** distinctly overlapping R2+3 and expanding broadly around apex of wing reaching mid-point between R2+3 & R4+5 (Figure 16(e))	…	…	…	…	…	…	…	…	…	…	***B. occipitalis***
	Costal band widening slightly (Figure 16(c)) to moderately (Figure 16(a)) around apex of wing	…	…	…	…	…	…	…	…	…	…	4
4	Abdominal **tergites 3–5** with broad black dorsolateral markings (Figures 17(f) and 18(f))	…	…	…	…	…	…	…	…	…	…	***B. pyrifoliae***
	Abdominal **tergites 3–5** without broad black dorsolateral markings	…	…	…	…	…	…	…	…	…	…	5

5	**Costal band** slightly overlapping R2+3, moderately broad around apex of wing (Figure 16(a)); abdominal **tergite 3** with a narrow transverse black band across anterior margin (constituting a "T"pattern), widening to cover lateral margins; **tergite 4** with rectangular (occasionally triangular) anterolateral or narrow lateral black markings; **tergites 3–5** with medium-width medial longitudinal black stripe (Figures 17(a) and 18(a))	… … … …	… … … …	… … … …	… … … …	… … … …	… … … …	… … … …	… … … …	… … … …	… … … …	***B. carambolae***
	Costal band confluent with R2+3, narrow to moderately broad around apex of wing (Figure 16(c)); abdominal **tergite 3** exhibits variations from black band across anterior margin (constituting a "T" pattern) to broad lateral bands, **tergite 4** without markings or with anterolateral or narrow lateral black margins (occasionally rectangular), **tergite 5** without markings or with anterolateral black markings (Figures 17(c) and 18(c))	… … … …	… … … …	… … … …	… … … …	… … … …	… … … …	… … … …	… … … …	… … … …	… … … …	***B. dorsalis s.l.***
	Comments / Results											

Diagnostic morphological characters of adult fruit flies of two economically important species of the *Bactrocera dorsalis* complex (modified from Table 3, IPPC ISPM 27 DP29: *Bactrocera dorsalis*)

		Sample code											**Sample code**									
Structure	***B. dorsalis s.l.***											***B. carambolae***										
Facial spot	**Medium to large**, circular to oval (interregionally variable)											**Medium**-sized, oval										
Tergites III-V	With **narrow to medium** width medial longitudinal black stripe											With **medium-width** medial longitudinal black stripe										
T III	**Exhibits variations** from **transverse black band** across anterior margin (constituting a "T" pattern) to **broad lateral bands**											With a **narrow transverse black band** across anterior margin (constituting a "T" pattern) widening to cover lateral margins										
T IV	**Without** any markings **or** with **anterolateral black markings** (occasionally **rectangular** in shape)											With **rectangular anterolateral** (occasionally triangular) **black** markings										
T V	**Without** any markings **or with anterolateral black** markings											**With anterolateral black** markings										
Scutum colour	Black to red–brown (inter or intra-regionally variable)											Dull black										

Postpronotal lobe	Entirely yellow											Entirely yellow										
Anterior margin of anepisternal stripe	Reaching midway between anterior margin of notopleuron and anterior npl. bristle; **straight to convex** (anterior margin)											Reaching midway between anterior margin of notopleuron and anterior npl. bristle; **convex** (anterior margin)										
Basal band of scutellum	Narrow											Narrow										
Lateral vittae	**Narrow** to **broad** (inter-regionally variable), parallel-sided, ending at or just behind ia. Bristles											**Broad,** parallel-sided, ending at or behind ia. Bristles										
Costal band	Narrow, **generally confluent with R2+3** (inter- or intra-regionally variable), **narrow to moderately broad** around apex of wing											Narrow, **slightly overlapping R2+3, moderately broad** around apex of wing										
Femora	Generally fulvous, <u>occasionally</u> with a small dark marking on outer surface of fore femora (inter-regionally variable)											Fulvous, **generally with a large elongate oval black** marking on outer surface of fore femora										

Summary Results sheet for the morphological test IPPC ISPM 27 DP29: *Bactrocera dorsalis*

Operator	
Stereomicroscope	

Sample code	Identification result	Date of analysis	Notes
01			
02			
03			
04			
05			
06			
07			
08			
09			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			

22			
23			
24			
25			
26			
27			
28			
29			
30			
31			
32			
33			
34			
35			
36			
37			
38			
39			
40			

Appendix 3 – Specifications and parameters for the molecular tests

Specification of the PCR Assay 1 (*COI*)

Name of the primer incl. sequence, literature reference, fragment length in bp:

LepF: 5'- ATTCAACCAATCATAAAGATATTGG-3'
LepR: 5'- TAAACTTCTGGATGTCCAAAAAAAATCA-3'

Literature: Hajibabaei et al., 2006: DNA barcodes distinguish species of tropical Lepidoptera, PNAS _ January 24, 2006 _ vol. 103_ no. 4, 968-971

Fragment length: 658bp

PCR - Parameters:

Thermocyler used: Biometra T3000 Thermal cycler

Mastermix: 5x HOT FIREPol® Master Mix, Solis Biodyne:

Composition:		**Final concentration:**
	Volume per reaction µl	
Water	6	
Mastermix	2	1x
Primer1:	0,5	0,5µM
Primer2:	0,5	0,5µM
∑	9	
DNA	1	

PCR conditions:

	°C	**Duration (min., sec.)**	**Nr. of Cycles**
Start	95	15 min	1
Denaturation	95	45 sec	5
Annealing	44	45 sec	
Extension	72	45 sec	
Denaturation	95	45 sec	35
Annealing	49	45 sec	
Extension	72	45 sec	
Final extension	72	7 min	1
Cooling	15	∞	

Specification of the PCR Assay 2 (*COI*)

Name of the primer incl. sequence, literature reference, fragment length in bp:

LCO1490: 5'- GGTCAACAAATCATAAAGATATTGG-3'
HCO2198: 5'- TAAACTTCAGGGTGACCAAAAAATCA-3'

Literature: Folmer O, Black M, Hoeh W, Lutz R & Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine, Biology and Biotechnology 3, 294–299.

Fragment length: 709bp

PCR - Parameters:

Thermocyler used: Biometra T3000 Thermal cycler

Mastermix: 5x HOT FIREPol® Master Mix, Solis Biodyne:

Composition:		**Final concentration:**
	Volume per reaction µl	
Water	6	
Mastermix	2	1x
Primer1:	0,5	0,5µM
Primer2:	0,5	0,5µM
∑	9	
DNA	1	

PCR conditions:

	°C	**Duration (min., sec.)**	**Nr. of Cycles**
Start	95	15 min	1
Denaturation	95	30 sec	5
Annealing	45	30 sec	
Extension	72	1 min	
Denaturation	95	30 sec	35
Annealing	51	1 min	
Extension	72	1 min	
Final extension	72	10 min	1
Cooling	15	∞	

Specification of the PCR Assay 3 (*ITS*1)

*ITS*6: 5'- AGC CGA GTG ATC CAC CGC T-3'
*ITS*7: 5'- GAA TTT CGC ATA CAT TGT AT-3'

Boykin et al., (2014); Armstrong and Cameron, (2000)

Fragment length: 499–543bp (the amplicon size varies for species and individuals) *B. carambolae* seem to have an additional insert of 44bp compared to *B. dorsalis*

PCR - Parameters:

Thermocyler used: Biometra T3000 Thermal cycler

Mastermix: 5x HOT FIREPol® Master Mix, Solis Biodyne:

Composition:		**Final concentration:**
	Volume per reaction µl	
Water	6	
Mastermix	2	1x
Primer1:	0,5	0,5µM
Primer2:	0,5	0,5µM
∑	9	
DNA	1	

PCR conditions:

	°C	**Duration (min., sec.)**	**Nr. of Cycles**
Start	95	15 min	1
Denaturation	95	30 sec	40
Annealing	55	30 sec	
Extension	72	30 sec	
Final extension	72	5 min	1
Cooling	15	∞	

Appendix 4 – Summary Results sheets with the results from the three operators

Operator 1	
Instrument	ZEISS Stemi 2000-C
Date of analysis/identification	29/05/20 – 03/06/20

Sample number	Analysis/Identification	Notes	Expected result	Assigned value
1	**Negative**	*B. kandiensis*	Negative	*Bactrocera kandiensis*
2	**Positive**	*B. dorsalis*	Positive	*Bactrocera dorsalis*
3	**Negative**	*Anastrepha* sp.	Negative	*Anastrepha suspensa*
4	**Negative**	*B. carambolae* ?	Negative	*Bactrocera carambolae*
5	**Positive**	*B. dorsalis*	Positive	*Bactrocera dorsalis*
6	**Negative**	*B. carambolae*	Negative	*Bactrocera carambolae*
7	**Negative**	*Anastrepha* sp.	Negative	*Anastrepha suspensa*
8	**Positive**	*B. dorsalis*	Negative	*Bactrocera carambolae*
9	**Positive**	*B. dorsalis*	Positive	*Bactrocera dorsalis*
10	**Negative**	*B. kandiensis*	Negative	*Bactrocera kandiensis*
11	**Positive**	*B. dorsalis*	Positive	*Bactrocera dorsalis*
12	**Positive**	*B. dorsalis*	Positive	*Bactrocera dorsalis*
13	**Negative**	*B. kandiensis*	Negative	*Bactrocera kandiensis*
14	**Positive**	*B. dorsalis*	Positive	*Bactrocera dorsalis*
15	**Negative**	*Dacus* sp. ?	Negative	*Zeugodacus cucurbitae*
16	**Negative**	*B. kandiensis* ?	Negative	*Bactrocera kandiensis*
17	**Positive**	*B. dorsalis*	Positive	*Bactrocera dorsalis*
18	**Positive**	*B. dorsalis*	Positive	*Bactrocera dorsalis*
19	**Negative**	*Dacus* sp. ?	Negative	*Zeugodacus cucurbitae*
20	**Negative**	Not *B. dorsalis* complex	Negative	*Bactrocera oleae*
21	**Negative**	*B. pyrifoliae* ?	Negative	*Bactrocera pyrifoliae*
22	**Not determined**	*B. dorsalis* or *B. carambolae* ?	Negative	*Bactrocera carambolae*
23	**Negative**	*Bactrocera occipitalis ?*	Negative	*Bactrocera occipitalis*
24	**Positive**	*B. dorsalis*	Positive	*Bactrocera dorsalis*
25	**Negative**	*Anastrepha* sp.	Negative	*Anastrepha obliqua*
26	**Negative**	*Dacus* sp. ?	Negative	*Dacus ciliatus*
27	**Not determined**	*B. dorsalis* or *B. carambolae* ?	Negative	*Bactrocera carambolae*
28	**Negative**	*Dacus* sp.	Negative	*Dacus ciliatus*
29	**Positive**	*B. dorsalis*	Positive	*Bactrocera dorsalis*
30	**Positive**	*B. dorsalis*	Positive	*Bactrocera dorsalis*
31	**Negative**	*B. kandiensis*	Negative	*Bactrocera kandiensis*
32	**Positive**	*B. dorsalis*	Positive	*Bactrocera dorsalis*
33	**Negative**	*B. caryeae*	Negative	*Bactrocera caryeae*
34	**Positive**	*B. dorsalis*	Positive	*Bactrocera dorsalis*
35	**Negative**	*Anastrepha* sp.	Negative	*Anastrepha obliqua*
36	**Not determined**	*B. dorsalis* or *B. carambolae* ?	Positive	*Bactrocera dorsalis*
37	**Negative**	Not *B. dorsalis* complex	Negative	*Bactrocera oleae*
38	**Negative**	*B. caryeae*	Negative	*Bactrocera caryeae*
39	**Not determined**	*B. dorsalis* or *B. carambolae* ?	Negative	*Bactrocera carambolae*
40	**Negative**	*B. carambolae*	Negative	*Bactrocera occipitalis*

Operator 2	
Instrument	ZEISS Stemi 508
Date of analysis/identification	08/06/20 – 18/06/20

Sample number	Analysis/Identification	Notes	Expected result	Assigned value
1	**Negative**	/	Negative	*Bactrocera kandiensis*
2	**Positive**	/	Positive	*Bactrocera dorsalis*
3	**Negative**	/	Negative	*Anastrepha suspensa*
4	**Positive**	Specimen characters match also with *B. carambolae*	Negative	*Bactrocera carambolae*
5	**Positive**	/	Positive	*Bactrocera dorsalis*
6	**Negative**	/	Negative	*Bactrocera carambolae*
7	**Negative**	/	Negative	*Anastrepha suspensa*
8	**Positive**	T pattern on T3: transverse band not continuous	Negative	*Bactrocera carambolae*
9	**Positive**	/	Positive	*Bactrocera dorsalis*
10	**Negative**	/	Negative	*Bactrocera kandiensis*
11	**Positive**	/	Positive	*Bactrocera dorsalis*
12	**Positive**	/	Positive	*Bactrocera dorsalis*
13	**Negative**	/	Negative	*Bactrocera kandiensis*
14	**Positive**	/	Positive	*Bactrocera dorsalis*
15	**Negative**	/	Negative	*Zeugodacus cucurbitae*
16	**Negative**	Several characters ambiguous, however NEGATIVE because of anepisternal stripe and markings on femora	Negative	*Bactrocera kandiensis*
17	**Positive**	/	Positive	*Bactrocera dorsalis*
18	**Negative**	Postpronotal lobe is NOT entirely yellow (T3 transversal band not continuous)	Positive	*Bactrocera dorsalis*
19	**Negative**	/	Negative	*Zeugodacus cucurbitae*
20	**Negative**	/	Negative	*Bactrocera oleae*
21	**Negative**	/	Negative	*Bactrocera pyrifoliae*
22	**Positive**	T pattern on T3: transverse band not continuous	Negative	*Bactrocera carambolae*
23	**Not determined**	Thorax is partly covered by a layer. Medial longitudinal stripe ambiguous	Negative	*Bactrocera occipitalis*
24	**Positive**	/	Positive	*Bactrocera dorsalis*
25	**Negative**	/	Negative	*Anastrepha obliqua*
26	**Negative**	/	Negative	*Dacus ciliatus*
27	**Positive**	/	Negative	*Bactrocera carambolae*
28	**Negative**	/	Negative	*Dacus ciliatus*
29	**Positive**	/	Positive	*Bactrocera dorsalis*
30	**Positive**	T pattern on T3: transverse band not continuous	Positive	*Bactrocera dorsalis*
31	**Negative**	Specimen not in good condition	Negative	*Bactrocera kandiensis*
32	**Positive**	Absence of medial vittae not visible	Positive	*Bactrocera dorsalis*
33	**Negative**	/	Negative	*Bactrocera caryeae*
34	**Positive**	/	Positive	*Bactrocera dorsalis*
35	**Negative**	/	Negative	*Anastrepha obliqua*

36	**Positive**	/	Positive	*Bactrocera dorsalis*
37	**Negative**	/	Negative	*Bactrocera oleae*
38	**Negative**	/	Negative	*Bactrocera caryeae*
39	**Positive**	T pattern on T3: transverse band not continuous	Negative	*Bactrocera carambolae*
40	**Positive**	/	Negative	*Bactrocera occipitalis*

Operator 3	
Instrument	LEICA M205 c
Date of analysis/identification	20/07/20 – 23/07/20 n_1 28/07/20 – 30/07/20 n_2 04/08/20 n_3

Sample number	Analysis/Identification_1	Analysis/Identification_2	Analysis/Identification_3	Notes	Expected result	Assigned value
1	**Negative**	**Negative**	**Negative**	*B. kandiensis*	Negative	*Bactrocera kandiensis*
2	**Positive**	**Positive**	**Positive**	*Bactrocera dorsalis s.l.*	Positive	*Bactrocera dorsalis*
3	**Negative**	**Negative**	**Negative**	*Anastrepha* sp.	Negative	*Anastrepha suspensa*
4	**Negative**	**Negative**	**Negative**	*B. carambolae*	Negative	*Bactrocera carambolae*
5	**Positive**	**Positive**	**Positive**	*Bactrocera dorsalis s.l.*	Positive	*Bactrocera dorsalis*
6	**Negative**	**Negative**	**Negative**	*B. carambolae*	Negative	*Bactrocera carambolae*
7	**Negative**	**Negative**	**Negative**	*Anastrepha* sp.	Negative	*Anastrepha suspensa*
8	**Negative**	**Negative**	**Negative**	*B. carambolae*	Negative	*Bactrocera carambolae*
9	**Positive**	**Positive**	**Positive**	*Bactrocera dorsalis s.l.*	Positive	*Bactrocera dorsalis*
10	**Negative**	**Negative**	**Negative**	*B. kandiensis*	Negative	*Bactrocera kandiensis*
11	**Positive**	**Positive**	**Positive**	*Bactrocera dorsalis s.l.*	Positive	*Bactrocera dorsalis*
12	**Positive**	**Positive**	**Positive**	*Bactrocera dorsalis s.l.*	Positive	*Bactrocera dorsalis*
13	**Negative**	**Negative**	**Negative**	*B. kandiensis*	Negative	*Bactrocera kandiensis*
14	**Positive**	**Positive**	**Positive**	*Bactrocera dorsalis s.l.*	Positive	*Bactrocera dorsalis*
15	**Negative**	**Negative**	**Negative**	/	Negative	*Zeugodacus cucurbitae*
16	**Negative**	**Negative**	**Negative**	*B. kandiensis*	Negative	*Bactrocera kandiensis*
17	**Positive**	**Positive**	**Positive**	*Bactrocera dorsalis s.l.*	Positive	*Bactrocera dorsalis*
18	**Positive**	**Positive**	**Positive**	*Bactrocera dorsalis s.l.*	Positive	*Bactrocera dorsalis*
19	**Negative**	**Negative**	**Negative**	/	Negative	*Zeugodacus cucurbitae*
20	**Negative**	**Negative**	**Negative**	/	Negative	*Bactrocera oleae*
21	**Negative**	**Negative**	**Negative**	*B. pyrifoliae*	Negative	*Bactrocera pyrifoliae*
22	**Negative**	**Negative**	**Negative**	*B. carambolae*	Negative	*Bactrocera carambolae*
23	**Negative**	**Negative**	**Negative**	*B. occipitalis*	Negative	*Bactrocera occipitalis*
24	**Positive**	**Positive**	**Positive**	*Bactrocera dorsalis s.l.*	Positive	*Bactrocera dorsalis*
25	**Negative**	**Negative**	**Negative**	*Anastrepha* sp.	Negative	*Anastrepha obliqua*
26	**Negative**	**Negative**	**Negative**	*Dacus ?*	Negative	*Dacus ciliatus*
27	**Negative**	**Negative**	**Negative**	*B. carambolae*	Negative	*Bactrocera carambolae*
28	**Negative**	**Negative**	**Negative**	/	Negative	*Dacus ciliatus*
29	**Positive**	**Positive**	**Positive**	*Bactrocera dorsalis s.l.*	Positive	*Bactrocera dorsalis*
30	**Positive**	**Positive**	**Positive**	*Bactrocera dorsalis s.l.*	Positive	*Bactrocera dorsalis*
31	**Negative**	**Negative**	**Negative**	*B. kandiensis*	Negative	*Bactrocera kandiensis*
32	**Positive**	**Positive**	**Positive**	*Bactrocera dorsalis s.l.*	Positive	*Bactrocera dorsalis*

33	**Negative**	**Negative**	**Negative**	*B. caryeae*	Negative	*Bactrocera caryeae*
34	**Positive**	**Positive**	**Positive**	*Bactrocera dorsalis s.l.*	Positive	*Bactrocera dorsalis*
35	**Negative**	**Negative**	**Negative**	*Anastrepha* sp.	Negative	*Anastrepha obliqua*
36	**Positive**	**Positive**	**Positive**	*Bactrocera dorsalis s.l.*	Positive	*Bactrocera dorsalis*
37	**Negative**	**Negative**	**Negative**	/	Negative	*Bactrocera oleae*
38	**Negative**	**Negative**	**Negative**	*B. caryeae*	Negative	*Bactrocera caryeae*
39	**Negative**	**Negative**	**Negative**	*B. carambolae*	Negative	*Bactrocera carambolae*
40	**Negative**	**Negative**	**Negative**	*B. occipitalis*	Negative	*Bactrocera occipitalis*

Appendix 5 – Calculation of performance characteristics - morphological protocol

Sensitivity, specificity, accuracy :

Diagnostic sensitivity, specificity and accuracy is assessed on the basis of the analysis of the whole set carried out by operator 3 (ANSES)

Operator_3_R1

Diagnostic sensitivity = true positives/(true positives + false negatives)

Diagnostic specificity = true negatives/(true negatives + false positives)

		Expected result	
		positive	negative
Operator result	positive	14	0
	negative	0	26

Sensitivity	**100**
Specificity	**100**
Accuracy	**100**

Repeatability : Operator_3_R1, Operator_3_R2, Operator_3_R3

Repeatability is assessed on the basis of the analysis of the whole set carried out by operator 3 (ANSES) (three repetitions of analysis).

Operator_3_R1, Operator_3_R2, Operator_3_R3

Expressed as % level of agreement among repetitions by Operator 3

Sample code	Repetitions	Operator3_ R1	Operator3_ R2	Operator3_ R3	Agreement	Disagreement	Level of agreement
1	3	Negative	Negative	Negative	3	0	100
2	3	Positive	Positive	Positive	3	0	100
3	3	Negative	Negative	Negative	3	0	100
4	3	Negative	Negative	Negative	3	0	100
5	3	Positive	Positive	Positive	3	0	100
6	3	Negative	Negative	Negative	3	0	100
7	3	Negative	Negative	Negative	3	0	100
8	3	Negative	Negative	Negative	3	0	100
9	3	Positive	Positive	Positive	3	0	100
10	3	Negative	Negative	Negative	3	0	100
11	3	Positive	Positive	Positive	3	0	100
12	3	Positive	Positive	Positive	3	0	100
13	3	Negative	Negative	Negative	3	0	100
14	3	Positive	Positive	Positive	3	0	100
15	3	Negative	Negative	Negative	3	0	100
16	3	Negative	Negative	Negative	3	0	100
17	3	Positive	Positive	Positive	3	0	100
18	3	Positive	Positive	Positive	3	0	100
19	3	Negative	Negative	Negative	3	0	100
20	3	Negative	Negative	Negative	3	0	100
21	3	Negative	Negative	Negative	3	0	100
22	3	Negative	Negative	Negative	3	0	100
23	3	Negative	Negative	Negative	3	0	100
24	3	Positive	Positive	Positive	3	0	100
25	3	Negative	Negative	Negative	3	0	100
26	3	Negative	Negative	Negative	3	0	100
27	3	Negative	Negative	Negative	3	0	100
28	3	Negative	Negative	Negative	3	0	100
29	3	Positive	Positive	Positive	3	0	100
30	3	Positive	Positive	Positive	3	0	100
31	3	Negative	Negative	Negative	3	0	100
32	3	Positive	Positive	Positive	3	0	100
33	3	Negative	Negative	Negative	3	0	100
34	3	Positive	Positive	Positive	3	0	100
35	3	Negative	Negative	Negative	3	0	100
36	3	Positive	Positive	Positive	3	0	100
37	3	Negative	Negative	Negative	3	0	100
38	3	Negative	Negative	Negative	3	0	100
39	3	Negative	Negative	Negative	3	0	100
40	3	Negative	Negative	Negative	3	0	100
	120				120	0	100

Repeatability **100**

Reproducibility : Operator_1, Operator_2, Operator_3_R1

Reproducibility is assessed on the basis of the analysis of the whole set carried out by operator 1, 2 (AGES) and 3 (ANSES) (first of the three repetitions of analysis).

Operator_1, Operator_2, Operator_3_R1

Expressed as % level of agreement among repetitions by the three Operators

Sample code	Operator3_R1	Operator1	Operator2	Repetitions	Agreement	Disagreement	Level of agreement
1	Negative	Negative	Negative	3	3	0	100,0
2	Positive	Positive	Positive	3	3	0	100,0
3	Negative	Negative	Negative	3	3	0	100,0
4	Negative	Negative	Positive	3	2	1	66,7
5	Positive	Positive	Positive	3	3	0	100,0
6	Negative	Negative	Negative	3	3	0	100,0
7	Negative	Negative	Negative	3	3	0	100,0
8	Negative	Positive	Positive	3	2	1	66,7
9	Positive	Positive	Positive	3	3	0	100,0
10	Negative	Negative	Negative	3	3	0	100,0
11	Positive	Positive	Positive	3	3	0	100,0
12	Positive	Positive	Positive	3	3	0	100,0
13	Negative	Negative	Negative	3	3	0	100,0
14	Positive	Positive	Positive	3	3	0	100,0
15	Negative	Negative	Negative	3	3	0	100,0
16	Negative	Negative	Negative	3	3	0	100,0
17	Positive	Positive	Positive	3	3	0	100,0
18	Positive	Positive	Negative	3	2	1	66,7
19	Negative	Negative	Negative	3	3	0	100,0
20	Negative	Negative	Negative	3	3	0	100,0
21	Negative	Negative	Negative	3	3	0	100,0
22	Negative	Not determined	Positive	3	1	2	0,0
23	Negative	Negative	Not determined	3	2	1	66,7
24	Positive	Positive	Positive	3	3	0	100,0
25	Negative	Negative	Negative	3	3	0	100,0
26	Negative	Negative	Negative	3	3	0	100,0
27	Negative	Not determined	Positive	3	1	2	0,0
28	Negative	Negative	Negative	3	3	0	100,0
29	Positive	Positive	Positive	3	3	0	100,0
30	Positive	Positive	Positive	3	3	0	100,0
31	Negative	Negative	Negative	3	3	0	100,0
32	Positive	Positive	Positive	3	3	0	100,0
33	Negative	Negative	Negative	3	3	0	100,0
34	Positive	Positive	Positive	3	3	0	100,0
35	Negative	Negative	Negative	3	3	0	100,0
36	Positive	Not determined	Positive	3	2	1	66,7
37	Negative	Negative	Negative	3	3	0	100,0
38	Negative	Negative	Negative	3	3	0	100,0
39	Negative	Not determined	Positive	3	1	2	0,0
40	Negative	Negative	Positive	3	2	1	66,7
				120	105	15	**87,5**

Reproducibility **87,5**

Appendix 6 – *In silico* testing of analytical specificity with DNA barcoding primer sets

In silico testing of analytical specificity by a database alignment (NCBI GenBank) was performed with the DNA barcoding primer sets (LCO1490/HCO2198 and LepF/LepR). The search set was limited to Tephritidae. The *ITS*1 primer set (*ITS*7/*ITS*6) was aligned in the same manner with the search being limited to *Bactrocera*. The results showed suitabilitiy of both primer sets (see Fig. 15-20) for identification of several *Bactrocera* spp., although we have to state that both barcoding and *ITS*1 sequencing are generic tests including targets and non-targets.

Distance trees of results from BLAST search were created with organism search set to Tephritidae with single primers (LepF, LepR, LCO1490, HCO2198, *ITS*7 and *ITS*6)

Figure 15 - Distance tree of results from BLAST search for LepF primer

Figure 16 - Distance tree of results from BLAST search for LepR primer

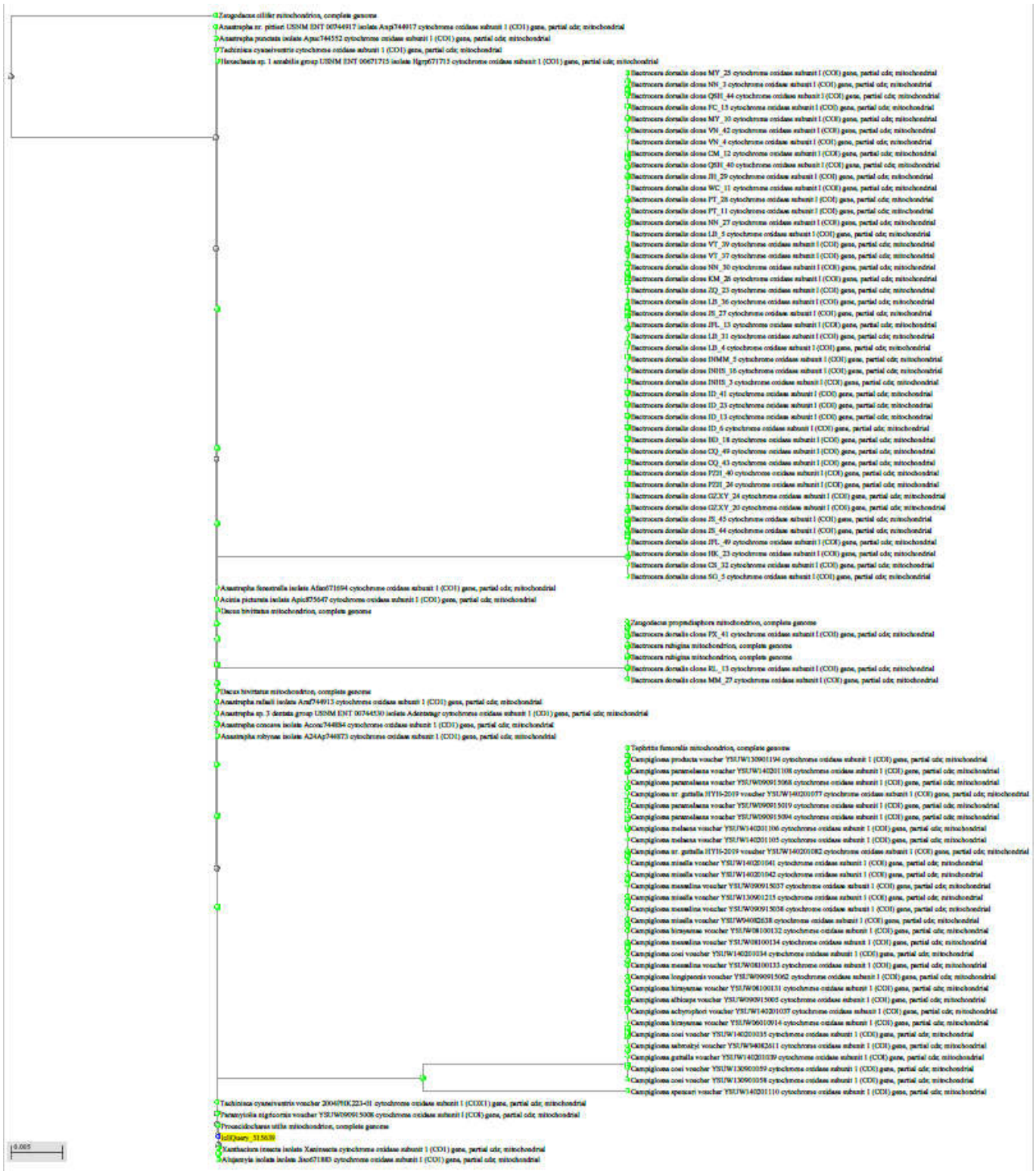


Figure 17 - Distance tree of results from BLAST search for LCO1490 primer

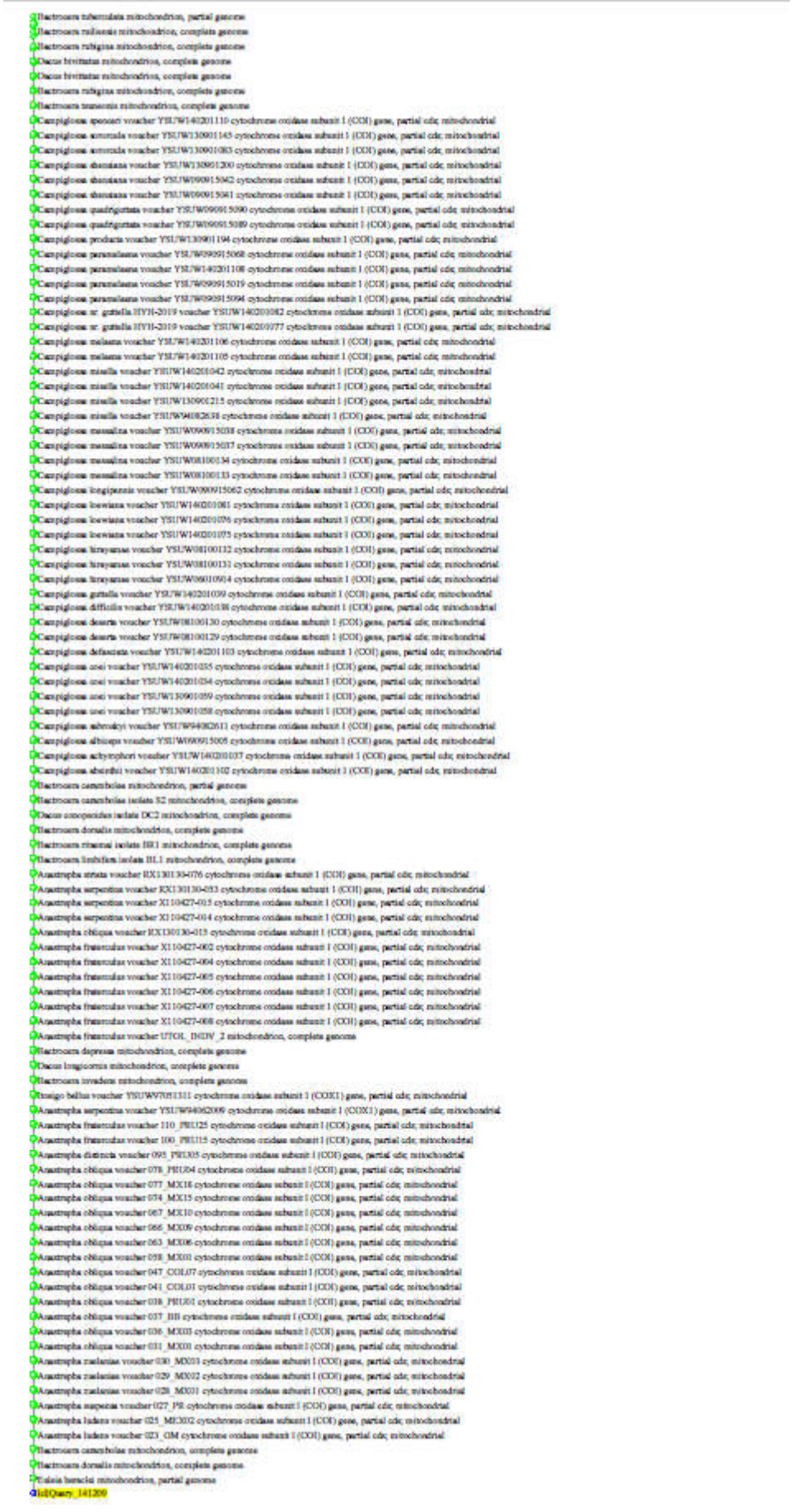


Figure 18 - Distance tree of results from BLAST search for HCO2198 primer

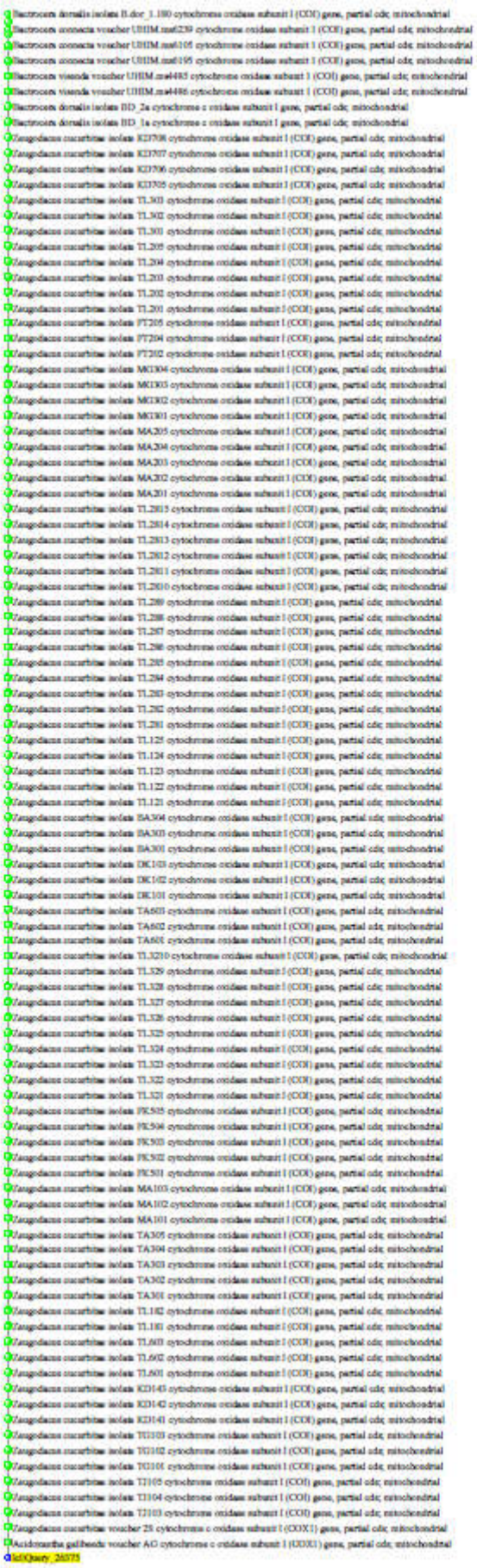


Figure 19 - Distance tree of results from BLAST search for *ITS7* primer

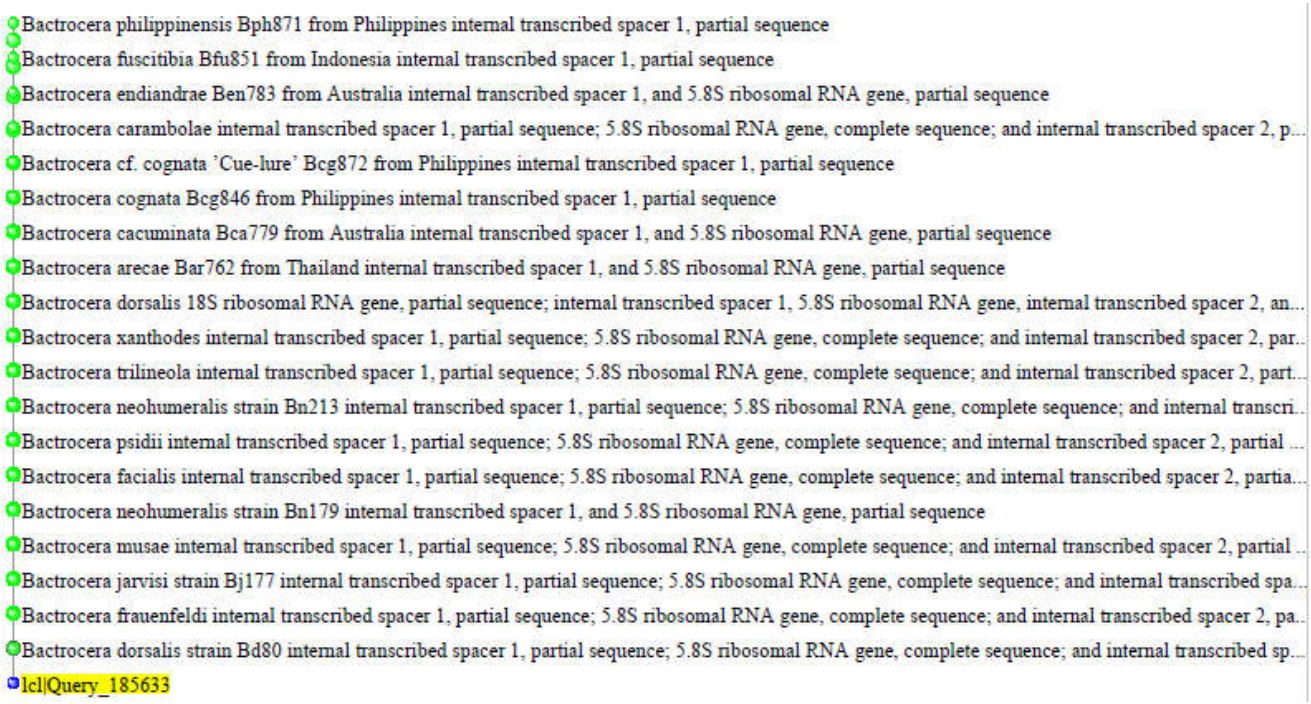
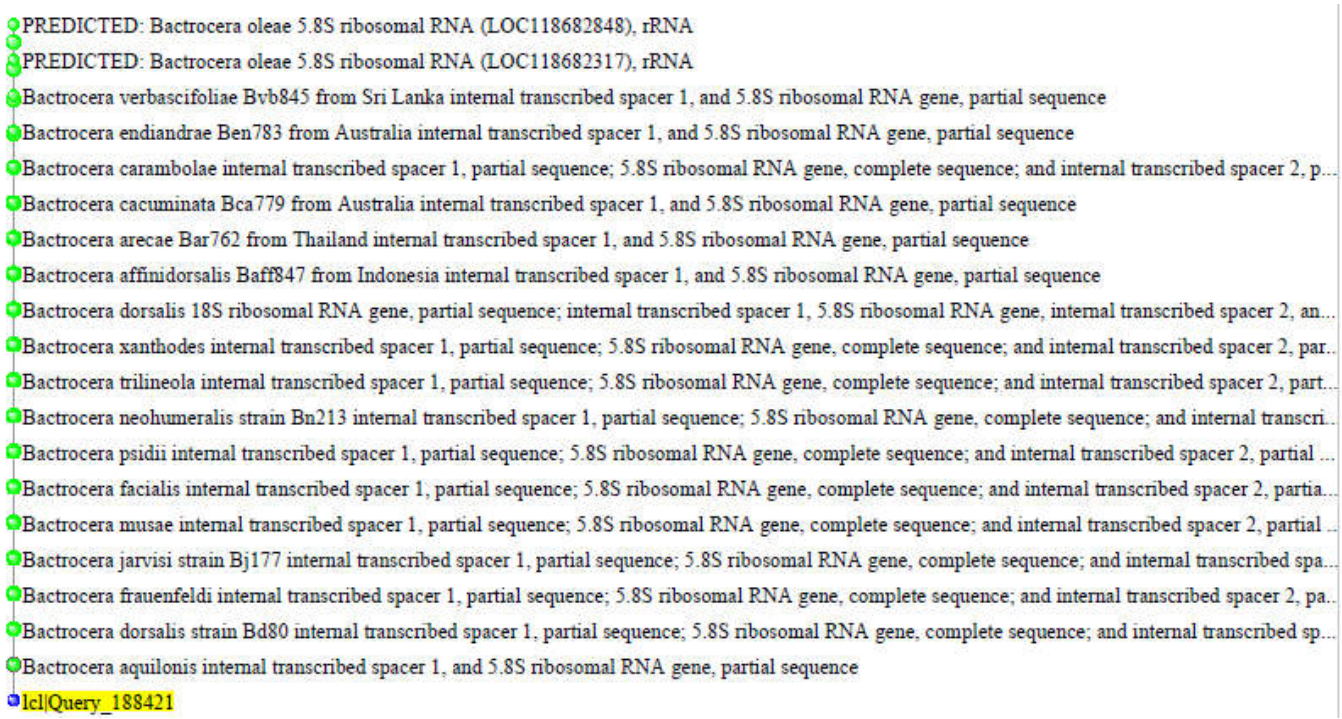


Figure 20 - Distance tree of results from BLAST search for *ITS6* primer



Appendix 7 – Results diagnosic specificity with DNA barcoding primer sets and *ITS*1 sequencing

Table 10: Results diagnostic specificity with DNA barcoding primer sets and *ITS*1 sequencing

	EPPO PM7/129 (LCO1490/HCO2198)		EPPO PM7/129 (LepF/LepR)		IPPC 27:DG29 (*ITS6*/*ITS7*)		Final	
Sample Nb.	**Result**	**Note**	**Result**	**Note**	**Result**	**Note**	**Expected result**	**Assigned value**
1	Negative	*B. kandiensis*	Negative	*B. kandiensis*	Negative	*B. kandiensis*	Negative	*Bactrocera kandiensis*
2	Positive	*B. dorsalis*	Positive	*B. dorsalis*	Positive	*B. dorsalis*	Positive	*Bactrocera dorsalis*
3	Negative	*Anastrepha fraterculus*	Negative	*Anastrepha sp.*	Negative	Not tested	Negative	*Anastrepha suspensa*
4	Negative	*B. carambolae*	Negative	*B. carambolae*	Negative	*B. carambolae*	Negative	*Bactrocera carambolae*
5	Positive	*B. dorsalis*	Positive	*B. dorsalis*	Positive	*B. dorsalis*	Positive	*Bactrocera dorsalis*
6	Negative	*B. carambolae*		No amplicon	Negative	*B. carambolae*	Negative	*Bactrocera carambolae*
7	Negative	*Anastrepha suspensa*	Negative	*Anastrepha sp.*	Negative	*Not tested.*	Negative	*Anastrepha suspensa*
8	Negative	*B. carambolae*	Negative	*B. carambolae*	Negative	*B. carambolae*	Negative	*Bactrocera carambolae*
9	Positive	*B. dorsalis*	Positive	*B. dorsalis*	Positive	*B. dorsalis*	Positive	*Bactrocera dorsalis*
10	Negative	*B. kandiensis*	Negative	*B. kandiensis*	Negative	*B. kandiensis*	Negative	*Bactrocera kandiensis*
11	Positive	*B. dorsalis*	Positive	*B. dorsalis*	Positive	*B. dorsalis*	Positive	*Bactrocera dorsalis*
12	Positive	*B. dorsalis*	Positive	*B. dorsalis*	Positive	*B. dorsalis*	Positive	*Bactrocera dorsalis*
13	Negative	*B. kandiensis*	Negative	*B. kandiensis*	Negative	*B. kandiensis*	Negative	*Bactrocera kandiensis*
14	Positive	*B. dorsalis*	Positive	*B. dorsalis*	Positive	*B. dorsalis*	Positive	*Bactrocera dorsalis*
15	Negative	*Zeugodacus cucurbitae*	Negative	*Zeugodacus cucurbitae*	Negative	Not tested	Negative	*Zeugodacus cucurbitae*
16	Positive	*B. dorsalis*	Positive	*B. dorsalis*	Negative	*B. kandiensis*	Negative	*Bactrocera kandiensis*
17	Positive	*B. dorsalis*	Positive	*B. dorsalis*	Positive	*B. dorsalis*	Positive	*Bactrocera dorsalis*
18	Positive	*B. dorsalis*	Positive	*B. dorsalis*	Positive	*B. dorsalis*	Positive	*Bactrocera dorsalis*
19	Negative	*Zeugodacus cucurbitae*	Negative	*Zeugodacus cucurbitae*	Negative	Not tested	Negative	*Zeugodacus cucurbitae*
20	Negative	*B. oleae*	Negative	*B. oleae*	Negative	Not tested	Negative	*Bactrocera oleae*
21		No amplicon		No amplicon		No amplicon	Negative	*Bactrocera pyrifoliae*
22	Negative	*B. carambolae*	Negative	*B. carambolae*	Negative	*B. carambolae*	Negative	*Bactrocera carambolae*
23	Negative	*B. carambolae*	Negative	*B. carambolae*	Negative	*B. carambolae*	Negative	*Bactrocera occipitalis*
24	Positive	*B. dorsalis*	Positive	*B. dorsalis*	Positive	*B. dorsalis*	Positive	*Bactrocera dorsalis*
25	Negative	*Anastrepha sp.*	Negative	*Anastrepha sp.*	Negative	Not tested	Negative	*Anastrepha obliqua*
26	Negative	*Dacus ciliatus*		No amplicon	Negative	Not tested	Negative	*Dacus ciliatus*
27	Negative	*B. carambolae*	Negative	*B. carambolae*	Negative	*B. carambolae*	Negative	*Bactrocera carambolae*
28	Negative	*Dacus ciliatus*	Negative	*Dacus ciliatus*	Negative	*Not tested.*	Negative	*Dacus ciliatus*
29	Positive	*B. dorsalis*	Positive	*B. dorsalis*	Positive	*B. dorsalis*	Positive	*Bactrocera dorsalis*
30	Positive	*B. dorsalis*	Positive	*B. dorsalis*	Positive	*B. dorsalis*	Positive	*Bactrocera dorsalis*
31	Negative	*B. kandiensis*	Negative	*B. kandiensis*	Negative	*B. kandiensis*	Negative	*Bactrocera kandiensis*
32	Positive	*B. dorsalis*	Positive	*B. dorsalis*	Positive	*B. dorsalis*	Positive	*Bactrocera dorsalis*
33	Negative	*B. carambolae*	Negative	*B. carambolae*		No consensus	Negative	*Bactrocera caryeae*
34	Positive	*B. dorsalis*	Positive	*B. dorsalis*	Positive	*B. dorsalis*	Positive	*Bactrocera dorsalis*
35	Negative	*Anastrepha sp.*	Negative	*Anastrepha sp.*	Negative	Not tested	Negative	*Anastrepha obliqua*
36	Positive	*B. dorsalis*	Positive	*B. dorsalis*	Positive	*B. dorsalis*	Positive	*Bactrocera dorsalis*
37	Negative	*B. oleae*	Negative	*B. oleae*	Negative	Not tested	Negative	*Bactrocera oleae*
38	Negative	*B. carambolae*	Negative	*B. carambolae*	Positive	*B. dorsalis*	Negative	*Bactrocera caryeae*
39	Negative	*B. carambolae*	Negative	*B. carambolae*	Negative	*B. carambolae*	Negative	*Bactrocera carambolae*
40		No amplicon		No amplicon	Negative	*B. occipitalis*	Negative	*Bactrocera occipitalis*

Appendix 8 – Summary Results sheets for analytical sensitivity, repeatability and reproducibility – molecular tests

Sample panel:
Sample 333/20: 1 adult specimen of *B. dorsalis*
Sample 334/20: 1 leg of *B. dorsalis*
Sample 335/20: 1 larvae of *B. dorsalis*
Sample 336/20: 1 pupa of *B. dorsalis*

Three experimental replicates were performed with this sample panel.

Measurement of DNA concentration

Quantity of DNA was determined using the Thermo Scientific Nanodrop 2000 Spectrophotometer, samples were measured three times (technical replicates), the mean and the standard deviation were calculated (Table 11)

Results for analytical sensitivity (DNA barcoding and *ITS*1 sequencing)

4 samples with one adult or part of the individuals were prepared in different dilutions (1:10, 1:100, 1:1000; 1:10.000; 1:100.000, 1:1.000.000). Specifications and parameters for the molecular tests are provided in Appendix 3. Amplicons at the detection limit and the last dilution step before the detection limit were sent for SANGER sequencing (Table 11). The quality of sequences was assessed by the length of the consensus sequences and % of high quality bases (%HQ), see Table 12.

Table 11: – Extracted DNA concentration and PCR sensistivity for *B. dorsalis* sample panel used for sensitivity testing

		DNA Concentration [ng/µl]				EPPO PM7/129		IPPC ISPM 27:DP 29
Sample Nb. & Developmental stage of *B. dorsalis*	**Dilution**	**Repetition 1**	**Repetition 2**	**Repetition 3**	**Mean ± SD**	**Amplicon (LCO1490/HCO2198)**	**Amplicon (LepF/lepR)**	**Amplicon (ITS6/ITS7)**
333/20 (adult)	Undiluted	202.9	203.4	206.6	204.3 ±2 .00			
	1:10	21.2	21.7	21.3	21.4 ± 0.24	Strong	Strong	Strong
	1:100	1.9	1.9	1.7	1.8 ± 0.12	Strong*	Strong	Strong
	1:1.000	N/A	N/A	N/A		Weak*	Strong*	Strong
	1:10.000	N/A	N/A	N/A		Negative	Negative	Strong
	1:100.000	N/A	N/A	N/A		Negative	Weak	Strong*
334/20 (leg)	Undiluted	141.4	142.1	141.9	141.8 ± 0.36			
	1:10	15.1	15.3	15.1	15.2 ± 0.12	Strong	Strong	Strong
	1:100	1.1	1.7	1.1	1.3 ± 0.35	Strong*	Strong	Strong
	1:1.000	N/A	N/A	N/A		Weak*	Strong	Strong
	1:10.000	N/A	N/A	N/A		Negative	Strong	Strong
	1:100.000	N/A	N/A	N/A		Negative	Strong*	Strong*
335/20 (larva)	Undiluted	387.9	390.7	387.2	388.6 ± 1.85			
	1:10	41.3	41.8	41.7	41.6 ± 0.26	Strong	Strong	Strong
	1:100	4.3	3.9	4.5	4.2 ± 0.31	Strong*	Strong	Strong
	1:1.000	N/A	N/A	N/A		Strong*	Strong	Strong
	1:10.000	N/A	N/A	N/A		Negative	Strong	Strong
	1:100.000	N/A	N/A	N/A		Negative	Strong*	Strong*
336/20 (pupa)	Undiluted	500.4	500.4	501.7	500.8 ± 0.75			
	1:10	53.4	54.4	54.0	53.9 ± 0.50	Strong	Strong	Strong
	1:100	5.9	5.5	5.7	5.7 ± 0.20	Strong*	Strong	Strong
	1:1.000	N/A	N/A	N/A		Strong*	Strong	Strong
	1:10.000	N/A	N/A	N/A		Negative	Strong*	Strong
	1:100.000	N/A	N/A	N/A		Negative	Weak*	Strong*

N/A: not validly measurable
*Sequenced amplicons

Table 12: Sequence quality criteria for *B. dorsalis* sample panel used for sensitivity testing

Test	Sample Nb. & Developmental stage of *B. dorsalis*	Dilution	Approx. Consensus Length (bp)	High Quality (HQ%) of Consensus	Calculated DNA Concentration [ng/µl]		
EPPO PM7/129 (LCO1490/ HCO2198)	1 (adult)	1:1.000	562	100	0.18	Mean	0.325
	8 (leg)	1:1.000	562	100	0.13		
	15 (larvae)	1:1.000	573	100	0.42		
	22 (pupa)	1:1.000	567	100	0.57		
EPPO PM7/129 (LepF/LepR)	1 (adult)	1:1.000	582	100	0.18	Mean	0.325
	8 (leg)	1:1.000	579	100	0.13		
	15 (larvae)	1:1.000	581		0.42		
	22 (pupa)	1:1.000	584	100	0.057		
IPPC27:DG26 (*ITS*6/*ITS*7)	1 (adult)	1:100.000	406	100	0.0018	Mean	0.00325
	8 (leg)	1:100.000	426	100	0.0013		
	15 (larvae)	1:100.000	411	100	0.0042		
	22 (pupa)	1:100.000	405	100	0.0057		

Results for repeatability (DNA barcoding and *ITS*1 sequencing)

Three replicates of *B. dorsalis* (adult – dilutions) were analysed with 3 technical repetitions.

The sample panel was analysed with three dilution steps and each with three technical repetitions. The results were summarized in table 13.

Table 13: Amplicon generation for DNA barcoding and *ITS*1 sequencing PCR repeatability test

			Amplicon production		
Test	**Sample Nb. & Developmental stage of B. dorsalis**	**Dilution**	**Repetition 1**	**Repetition 2**	**Repetition 3**
EPPO PM7/129 (LCO1490/HCO2198)	1 (adult)	1:100	Strong	Strong	Strong
		1:1.000	Strong	Strong	Weak
		1:10.000	Negative	Negative	Negative
	8 (leg)	1:100	Strong	Strong	Strong
		1:1.000	Strong	Strong	Strong
		1:10.000	Weak	Weak	Strong
	15 (larvae)	1:100	Strong	Strong	Strong
		1:1.000	Strong	Strong	Strong
		1:10.000	Weak	Strong	Strong
	22 (pupa)	1:100	Strong	Strong	Negative
		1:1.000	Negative	Negative	Strong
		1:10.000	Strong	Strong	Strong
EPPO PM7/129 (LepF/LepR)	1 (adult)	1:100	Strong	Strong	Strong
		1:1.000	Strong	Strong	Strong
		1:10.000	Strong	Strong	Strong
	8 (leg)	1:100	Strong	Strong	Strong
		1:1.000	Strong	Strong	Strong
		1:10.000	Strong	Strong	Strong
	15 (larvae)	1:100	Strong	Strong	Strong
		1:1.000	Strong	Strong	Strong
		1:10.000	Strong	Strong	Strong
	22 (pupa)	1:100	Strong	Strong	Strong
		1:1.000	Strong	Strong	Strong
		1:10.000	Strong	Strong	Strong
IPPC 27:DG29 (*ITS*6/7)	1 (adult)	1:100	Strong	Strong	Strong
		1:1.000	Strong	Strong	Strong
		1:10.000	Strong	Strong	Strong
	8 (leg)	1:100	Strong	Strong	Strong
		1:1.000	Strong	Strong	Strong
		1:10.000	Strong	Strong	Strong
	15 (larvae)	1:100	Strong	Strong	Strong
		1:1.000	Strong	Strong	Strong
		1:10.000	Strong	Strong	Strong
	22 (pupa)	1:100	Strong	Strong	Strong
		1:1.000	Strong	Strong	Strong
		1:10.000	Strong	Strong	Strong

Results for reproducibility (DNA barcoding and *ITS*1 sequencing)

Table 14: Sample panel

Target	Non target	Origin
B. dorsalis adult	/	Thailand /Saraburi
B. dorsalis larva	/	Thailand /Saraburi
B. dorsalis pupa	/	Thailand /Saraburi
	Bactrocera correcta larva	India
	Bactrocera carambolae adult	French Guyana
	Bactrocera latifrons larva	Thailand

Testing reproducibility of the PCR tests:

The tests were performed with three technical replicates and under different conditions (two operators on different days and using different thermocycler machines). The results are shown in tables 15 and 16.

Table 15: Reproducibility of the PCR tests operator 1

Operator:	Pohn
Date of performance:	22.02.2021
Thermocycler machine:	BiometraT3000 (I)

	EPPO PM 7/129 (LCO1490/HCO2198)			EPPO PM 7/129 (LepF/LepR)			IPPC 27:DG29 (*ITS6*/*ITS7*)		
Species & Sample Nb.	**Repetition 1**	**Repetition 2**	**Repetition 3**	**Repetition 1**	**Repetition 2**	**Repetition 3**	**Repetition 1**	**Repetition 2**	**Repetition 3**
Bactrocera dorsalis (adult), 333/20	Amplicon*	Amplicon	Amplicon	Amplicon*	Amplicon	Amplicon	Amplicon*	Amplicon	Amplicon
Bactrocera *dorsalis* (adult), 335/20	Amplicon*	Amplicon	Amplicon	Amplicon*	Amplicon	Amplicon	Amplicon*	Amplicon	Amplicon
Bactrocera dorsalis (adult), 336/20	Amplicon*	Amplicon	Amplicon	Amplicon*	Amplicon	Amplicon	Amplicon*	Amplicon	Amplicon
Bactrocera correcta (larva), 158/21	Amplicon*	Amplicon	Amplicon	Amplicon*	Amplicon	Amplicon	Amplicon*	Amplicon	Amplicon
Bactrocera carambola e (adult), 8	Amplicon*	Amplicon	Amplicon	Amplicon*	Amplicon	Amplicon	Amplicon*	Amplicon	Amplicon
Bactrocera latifrons (larva), 867/20c	Amplicon*	Amplicon	Amplicon	Amplicon*	Amplicon	Amplicon	Amplicon*	Amplicon	Amplicon

*Sequenced

Table 16: reproducibility of the PCR tests operator 2

Operator:	Heiss
Date of performance:	21.02.2021
Thermocycler machine:	BiometraT3000 (II)

	EPPO PM7/129 (LCO1490/HCO2198)			**EPPO PM7/129 (LepF/LepR)**			**IPPC 27:DG29 (*ITS6*/*ITS7*)**		
Species & Sample Nb.	**Repetition 1**	**Repetition 2**	**Repetition 3**	**Repetition 1**	**Repetition 2**	**Repetition 3**	**Repetition 1**	**Repetition 2**	**Repetition 3**
Bactrocera dorsalis (adult), 333/20	Amplicon*	Amplicon	Amplicon	Amplicon*	Amplicon	Amplicon	Amplicon*	Amplicon	Amplicon
Bactrocera dorsalis (adult), 335/20	Amplicon*	Amplicon	Amplicon	Amplicon*	Amplicon	Amplicon	Amplicon*	Amplicon	Amplicon
Bactrocera dorsalis (adult), 336/20	Amplicon*	Amplicon	Amplicon	Amplicon*	Amplicon	Amplicon	Amplicon*	Amplicon	Amplicon
Bactrocera correcta (larva), 158/21	Amplicon*	Amplicon	Amplicon	Amplicon*	Amplicon	Amplicon	Amplicon*	Amplicon	Amplicon
Bactrocera carambola e (adult),8	Amplicon*	Amplicon	Amplicon	Amplicon*	Amplicon	Amplicon	Amplicon*	Amplicon	Amplicon
Bactrocera latifrons (larva), 867/20c	Amplicon*	Amplicon	Amplicon	Amplicon*	Amplicon	Amplicon	Amplicon*	Amplicon	Amplicon

*Sequenced

Testing reproducibility of the SANGER sequence analysis:

The reproducibility of the SANGER sequence analysis was tested with the same sample panel. The sequence analysis was performed by two operators on different days. The alignment of the consensus sequence will be performed in three different data bases (NCBI GenBank, Bold, Q-Bank). Tables 17 and 18 depict the results of reproducibility.

Table 17: Reproducibility of the SANGER sequence analysis operator 1

Operator:	Pohn
Date of performance:	23.02.2021
Software:	Geneious prime® 10.1.3

	EPPO PM7/129 (LCO1490/HCO2198)			**EPPO PM7/129 (LepF/LepR)**			**IPPC 27:DG29 (*ITS6*/*ITS7*)**
Species & Sample Nb.	**NCBI GenBank**	**Bold**	**Q-Bank**	**Species & Sample Nb.**	**NCBI GenBank**	**Bold**	**Q-Bank**
Bactrocera dorsalis (adult), 333/20	*B. dorsalis*	*B. dorsalis*	*B. dorsalis*	*Bactrocera dorsalis* (adult), 333/20	*B. dorsalis*	*B. dorsalis*	*B. dorsalis*
Bactrocera dorsalis (adult), 335/20	*B. dorsalis*	*B. dorsalis*	*B. dorsalis*	*Bactrocera dorsalis* (adult), 335/20	*B. dorsalis*	*B. dorsalis*	*B. dorsalis*
Bactrocera dorsalis (adult), 336/20	*B. dorsalis*	*B. dorsalis*	*B. dorsalis*	*Bactrocera dorsalis* (adult), 336/20	*B. dorsalis*	*B. dorsalis*	*B. dorsalis*
Bactrocera correcta (larva), 158/21	*B. correcta*	*B. correcta*	*B. correcta*	*Bactrocera correcta* (larva), 158/21	*B. correcta*	*B. correcta*	*B. correcta*
Bactrocera carambolae (adult),8	*B. carambolae*	*B. carambolae*	*B. carambolae*	*Bactrocera carambolae* (adult),8	*B. carambolae*	*B. carambolae*	*B. carambolae*
Bactrocera latifrons (larva), 867/20c	*B. latifrons*	*B. latifrons*	*B. latifrons*	*Bactrocera latifrons* (larva), 867/20c	*B. latifrons*	*B. latifrons*	No sequence in database

Table 18: Reproducibility of the SANGER sequence analysis operator 2

Operator:	Gottsberger
Date of performance:	26.02.2021
Software:	Geneious prime® 10.1.3

	EPPO PM7/129 (LCO1490/HCO2198)			**EPPO PM7/129 (LepF/LepR)**			**IPPC 27:DG29 (*ITS6*/*ITS7*)**
Species & Sample Nb.	**NCBI GenBank**	**Bold**	**Q-Bank**	**NCBI GenBank**	**Bold**	**Q-Bank**	**NCBI GenBank**
Bactrocera dorsalis (adult), 333/20	*B. dorsalis*	*B. dorsalis*	*B. dorsalis*	*B. dorsalis*	*B. dorsalis*	*B. dorsalis*	*B. dorsalis*
Bactrocera dorsalis (adult), 335/20	*B. dorsalis*	*B. dorsalis*	*B. dorsalis*	*B. dorsalis*	*B. dorsalis*	*B. dorsalis*	*B. dorsalis*
Bactrocera dorsalis (adult), 336/20	*B. dorsalis*	*B. dorsalis*	*B. dorsalis*	*B. dorsalis*	*B. dorsalis*	*B. dorsalis*	*B. dorsalis*
Bactrocera correcta (larva), 158/21	*B. correcta*	*B. correcta*	*B. correcta*	*B. correcta*	*B. correcta*	*B. correcta*	*B. correcta*
Bactrocera carambolae (adult),8	*B. carambolae*	*B. carambolae*	*B. carambolae*	*B. carambolae*	*B. carambolae*	*B. carambolae*	*B. carambolae*
Bactrocera latifrons (larva), 867/20c	*B. latifrons*	*B. latifrons*	*B. latifrons*	*B. latifrons*	*B. latifrons*	*B. latifrons*	No sequence in database

Appendix 9 – Calculations for the performance characteristics - molecular tests

Appendix 9 shows the calculations for the performance characteristics.

Table 19: Calculations of the applicable performance characteristics (sensitivity, specificity and accuracy) for all three primer sets.

Target Species	**Criteria**	**EPPO PM7/129 (LCO1490/HCO2198)**	**EPPO PM7/129 (LepF/LepR)**	**IPPC 27:DG29 (*ITS7*/*ITS6*)**
Bactrocera dorsalis	Number of Positive Agreements	14	14	14
	Number of Negative Agreements	23	21	13
	Number of Negative Deviations	0	0	0
	Number of Positive Deviations	1	1	1
	Sensitivity	**100**	**100**	**100**
	Specificity	**96**	**95**	**93**
	Accuracy	**97**	**97**	**96**

Appendix 10 - Morphological analysis of sample 23

Due to the conflicting results obtained from the molecular analysis with respect to the assigned value of sample 23 (*Bactrocera occipitalis*), an in-depth morphological analysis was conducted by the three operators involved in the morphological validation study. High resolution pictures of sample 23 and *B. occipitalis* and *B. carambolae* specimens included in the sample set were taken to support conclusions (see Fig. 21). Pictures from FruitFly ID Australia were also checked (https://fruitflyidentification.org.au/), in addition to figures from DP 29 (IPPC, 2019).

Sample 23 was donated to the AGES collection (recodification F20042).
Data of sample 23 are the following.

- origin: Philippines, UPLB campus, rainforest area (mixed vegetation, trap catch)
- sampling date: 08.01.2000
- leg. and/or det. (no indication of the name's role available): G. Quimio

Characters that allow the discrimination of *B. occipitalis* and *B. carambolae* (IPPC, 2019) are resumed in table 20, together with the comments and the general final opinion of operators.

Table 20: When not otherwise indicated, referring to figures and table 3 means in the DP29 (IPPC, 2019)

Structure	*B. carambolae*	*B. occipitalis*	Operator 1	Operator 2	Operator 3	General opinion of operators
Tergites III-V	With **medium-width** medial longitudinal black stripe	With **very broad** medial longitudinal black stripe	*There is a (very?) broad medial longitudinal black stripe on tergites 3–5*	*The broadness of the medial longitudinal stripe on T3-5 was not fully clear to me*	*Medial longitudinal black stripe looks broad/ very broad*	Not definitively clear if it is broad or very broad (see also Fruit Fly ID Australia - abdomen variation CAR002 in *B. carambolae*)
T III	With a **narrow transverse black band** across anterior margin (constituting a "T" pattern) widening to cover lateral margins	With a **narrow transverse black band** across anterior margin widening to cover lateral margins	*Figures match higher with B. occipitalis, but no difference to* B. *carambolae in description*	*Figures match higher with B. occipitalis, but no difference to B. carambolae in description*	*Considering Fig. 17 and 18, higher matching with (e), B. occipitalis. However variations are possible, see FruitFly ID Australia*	Even if a higher match with *B. occipitalis* Figures in DP 29, no clear differences arise from the description in Table 3. See also Fruit Fly ID Australia - abdomen variation CAR002 in *B. carambolae* and abdomen variation OCC008 in *B. occipitalis*
T IV	With **rectangular anterolateral** (occasionally triangular) **black** markings	Exhibits variations from **anterolateral black** markings to **broad lateral bands**	*Pictures match higher with B. occipitalis, but no difference to B. carambolae in description*	*Pictures match higher with B. occipitalis, but no difference to B. carambolae in description*	*Considering Fig. 17 and 18 in IPPC, 2019, higher match with (e), B. occipitalis. However variations are possible, see FruitFly ID Australia*	
T V	**With anterolateral black** markings	With **broad lateral black bands** that cover lateral margins	*Pictures match higher with B. occipitalis, but no difference to B. carambolae in description*	*Pictures match higher with B. occipitalis, but no difference to B. carambolae in description*	*Considering Fig. 17 and 18, higher matching with (e), B. occipitalis. However variations are possible, see FruitFly ID Australia*	

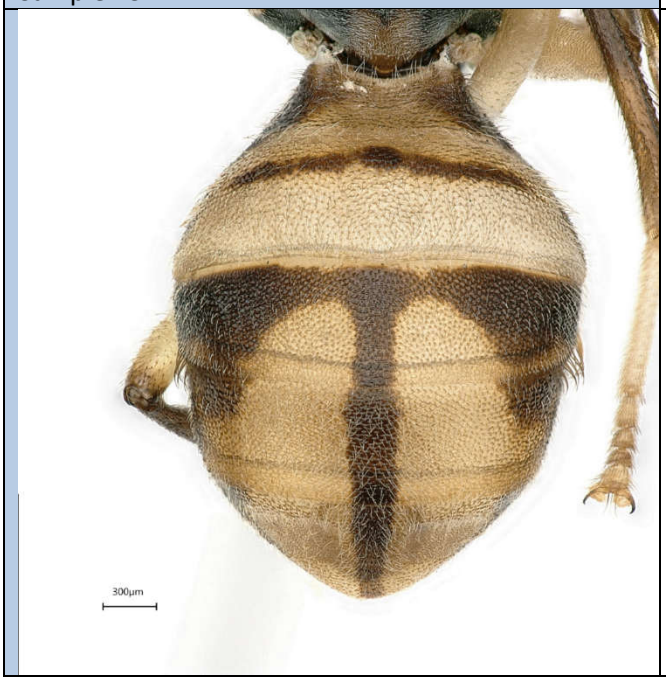
Lateral vittae	**Broad,** parallel-sided, ending at or behind ia. Bristles	**Broad, parallel- or subparallel-sided**; either ending at ia. bristles or (in some specimens) ending behind ia. Bristles	*Hard to see after DNA extraction; vittae look subparallel sided; Considering Fig. 13, higher matching with (e), B. occipitalis*	*Hard to see after DNA extraction; vittae look subparallel sided; Considering Fig. 13, higher matching with (e), B. occipitalis*	*Hard to see after DNA extraction; vittae look subparallel sided; Considering Fig. 13, higher matching with (e), B. occipitalis*	Hard to see after DNA extraction; vittae look subparallel sided; Considering Fig. 13, higher matching with (e), *B. occipitalis*
Costal band	Narrow, **slightly overlapping R2+3, moderately broad** around apex of wing	Narrow, **distinctly overlapping R2+3, broad around apex of wing** extending to mid-point between R_{2+3} and R_{4+5}	*The costal band overlaps only "slightly" but not until the "mid-point between R_{2+3} and R_{4+5}".*	*I crossed out the costal band character at B. occipitalis ('distinctly') and noted: ~slightly overlapping, moderately broad around the apex*	*Considering Fig. 16, higher matching with (e), B. occipitalis.*	Costal band doesn't seem to be "distinctly" overlapping R_{2+3}, but rather "slightly" overlapping. It is noted that the difference between "slightly" and "distinctly" overlapping is not clear from Fig. 16 (a) and (e)
Operators independent conclusion			*B. occipitalis - but with uncertainty.*	*Not determined. Excluded B. occipitalis ('distinctly') and B. dorsalis ('confluent') in the key*	*Bactrocera occipitalis with a certain degree of uncertainty*	***Bactrocera dorsalis* complex**

The general conclusion after the comments of the operators is that sample 23 cannot be morphologically identified as *Bactrocera occipitalis* with a sufficient degree of certainty. Even if the shape of lateral vitte looks subparallel-sided, other key characters to distinguish between *B. occipitalis* and *B. carambolae* do not lead to a clear, undoubtful identification. In addition, figures from FruitFly ID Australia website about variations from the typical abdomen appearance for the two species add a further level of uncertainty with respect to DP 29.

In the light of this morphological examination and the results from the molecular analysis (*Bactrocera carambolae*), sample 23 should be only identified as belonging to the ***Bactrocera dorsalis* complex**. It should be recalled that for regulamentary purposes, the identification to the level of "complex" is already largely sufficient to EU National Plant Protection Organisations to trigger adequate phytosanitary measures, like, for example, the destruction of an infested lot at an EU entry point.

Figure 21

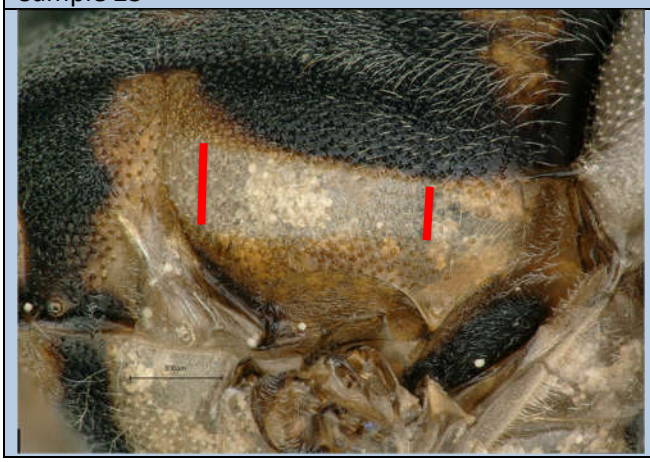
Tergites III-V

Sample 23	Sample 6 (*B. carambolae*)
	

Tergites III-V

Sample 23	Sample 6 (*B. carambolae*)
	

Lateral vittae

Sample 23	Sample 6 (*B. carambolae*)
	

Costal band

Sample 23	Sample 27 (*B. carambolae*)
	
Sample 40 (*B. occipitalis*)	Sampe 4 (*B. carambolae*)