Dear recommenders and reviewers,

We would like to thank you for reading carefully our manuscripts and for the comments provided. They allowed us to improve the document and we hereby provide a point-by-point response to them. All the lines are referred as lines from the original document and the revised version is provided with track changes.

Kind regards,

Sébastien Massart (on behalf of all co-authors)

Comments from recommender (Olivier Schumpp)

- L1091: ... "and Massart S." (?)

  Response : we have double checked all the bibliography and integrated this modification

- Diagram of figures 1 and 3: should there not be an exit arrow at the end of the verification and/or validation steps?

  Response: There is an exit arrow going to the grey box in the middle of the scheme (HTS test already used in the laboratory for pest diagnostic) by the right side. To make it more visible we have enlarged the arrows. On the other hand, both figures were somehow redundant and we decided to delete figure 3 from the document as it did not provide additional information compared to figure 1

Comments from Denis Altenbach

- General remark: Chapter 15 References needs careful reviewing and harmonization.

  Response: indeed, we have corrected all references after careful examination

- Risk analysis: Ihikawa diagram

For amplicon sequencing the number of PCR-Reactions (replicates) per sample and the percentage of the nucleic acid extract used for the analysis is essential. The probability of finding rare sequences increases or decreases depending on these factors. This issue could also be a critical step in the risk assessment of HTS approaches in certain cases and should probably be mentioned in this document.

Response: we have added the replicates item in the Amplicon sequencing branch of the diagram. We did not modify the main text but added this issue in the Supplementary material too.

- In Chapter 6.2. Analytic specificity you discuss the "desired taxonomic resolution": we think that the topic of required taxonomic resolution has to be addressed in parallel with the definition of the "intended use". Not all diagnostics need to go to the species level. The choice of genetic marker sequence may not enable to differentiate among closely related species, as correctly mentioned by the authors. Yet, using additional genetic marker sequences of other genes may so far not be helpful, as the corresponding reference data may be missing.

We guess that this topic is discussed in more details in Publication No. 12 (Lebas et al., EPPO Bull.). If not, this paragraph should be extended by a view sentences.

> **Response**: this is indeed a key element for the intended use and is not detailed extensively in the companion publication. So, we stated it clearly in the first paragraph of the chapter on intended use and HTS test selection (Lines 143-154). As analytical specificity should be aligned with the desired taxonomic resolution, we added a new sentence reminding it when starting the "Analytical specificity" chapter and modified the former first sentence by changing "species" to "required taxonomic resolution". We also added 4 lines in the second paragraph.

- For readers not so familiar with development and validation of diagnostic tests it would be nice to have a brief description of what is the difference between "reference material" and "controls".

> **Response**: It is indeed an important element. We stated in chapter 4.1 that reference material is useful at many steps, including as controls to monitor the performance of the test (lines 224-226). In lines 294-295 (chapter 5), we also stated that "Controls, corresponding to reference materials, are crucial in any diagnostic test, including HTS tests, as they provide essential traceability and validity in testing.". In addition, we referred to the EPPO standard PM7/147 for more information (lines 229-230). Although we could describe it better, we believe the current information might be enough and that the interested reader can consult the EPPO standard

- Line 166: ISO/IEC 17025:2017 should be named properly.

> **Response**: good point indeed, we modified it and changed also the format of ISO norms in lines 729, 731 and 1042.

<u>Comments from David Roquis</u>

Major comments

Two things I felt were missing in these guidelines are related to the choice of the HTS technology (and the associated sequencing kit) and the choice/optimization of the bioinformatic pipeline. HTS technologies have changed a lot in the past two decades, and not all of them are appropriate for every type of diagnostic tests. They have their inherent limits and I believe a short paragraph about that would be appropriate.

Response: We have added a small paragraph underlining these points in chapter 2 "Intended use and test selection" (Lines 157-161)

Same thing related to the bioinformatic treatment of the produced datasets. The choice of the tools, the selected parameters and the choice of the reference database (is it specific to some organisms? how well is it curated?) should be more developped. I understand that the manuscript is very generalist, but some tools will be more suited to detect and identify reads coming from bacteria than insects for example. Parameters should always be adjusted and reads QC and bioinformatics QC should always be performed for each experiment in order to properly assess the validity of the detection. Although it is somehow mentioned in parts 6.1 and 6.2, I felt it is not enough emphasized.

Response: You are right and these points have been addressed in depth in the companion publication which is by now publicly available from mid-August: https://onlinelibrary.wiley.com/doi/epdf/10.1111/epp.12863

Finally, again, I understand that the paper is very general, but I would have liked to have some very concrete guidelines to some questions. For example, chapter 6.1 discuss about the importance of the number of reads / reads ratio necessary to assess the presence/absence of an organism. It is indeed a very fundamental question, and I believe some case exemple could help the reader to decide what is good for him (what would be an optimal ratio range when working with bacteria in plant tissue matrix, or fungi in soil matrix, etc.). Something similar to Table 1 or 2, but on how to choose sensitivity / FDR treshold (for example).

Response: we agree threshold determination is really a key element. Nevertheless, proper threshold determination requires more research at this stage. Some considerations have been already addressed in the literature (see https://doi.org/10.3390/BIOLOGY11020263) and more publications on this topic will follow (at least for virus detection from main author's laboratory with the use of alien external control and adaptative threshold fixation depending on the detected reads on each batch).

Minor comments

99: trees
252: artificially

Response: both corrections done

253,257: I am always very skeptical with the use of artificially generated or simulated datasets. From my experience in bioinformatics, methods or bioinformatics tools optimized on simulated data tend to overperform on them, but very otfen completely underperform on real biological datasets (often causing a low sensitivity). Of course, simulated data can be used for validation of the pipeline, but it should never be the only reference datasets. Real curated biological datasets should always be used.

Response: we have added this suggestion “To complement biological material” and introduced the existence of semi-artificial datasets too on these lines

468: Although I understand the rationale here, I feel there are never too many controls and would always include water or non infected plants tissues as extra negative controls

Response: this is true but the cost of these controls is currently a strong challenge for laboratories. This challenge might be lower for amplicon sequencing as the library preparation is less expensive.

578: i.e. or e.g.

Response: done