High-throughput sequencing for the diagnostic of plant pathologies and identification of pests: recommendations and challenges

Olivier Schumpp based on reviews by Denise Altenbach and David Roquis

A recommendation of:

Open Access

Guidelines for the reliable use of high throughput sequencing technologies to detect plant pathogens and pests

Published: 2022-10-07

Submitted: 13 June 2022, Recommended: 12 September 2022

Recommendation

High-throughput sequencing (HTS) has revealed an incredible diversity of microorganisms in ecosystems and is also changing the monitoring of macroorganism biodiversity (Deiner et al. 2017; Piper et al. 2019).

The diagnostic of plant pathogens and the identification of pests is gradually integrating the use of these techniques, but there are still obstacles. Most of them are related to the reliability of these analyses, which have long been considered insufficient because of their dependence on a succession of sophisticated operations involving parameters that are sometimes difficult to adapt to complex matrices or certain diagnostic contexts. The need to validate HTS approaches is gradually being highlighted in recent work but remains poorly documented (Bester et al. 2022).

In this paper, a large community of experts presents and discusses the key steps for optimal control of HTS performance and reliability in a diagnostic context (Massart et al. 2022). It also addresses the issue of costs. The article provides recommendations that closely combine the quality control requirements commonly used in conventional diagnostics with newer or HTS-specific control elements and concepts that are not yet widely used. It discusses the value of these for the use of the various techniques currently covered by the terms "High Throughput Sequencing" in diagnostic activities. The elements presented are intended to limit false positive or false negative results but will also optimise the interpretation of contentious results close to the limits of analytical sensitivity or unexpected results, both of which appear to be frequent when using HTS.

Furthermore, the need for risk analysis, verification and validation of methods is well illustrated with numerous examples for each of the steps considered crucial to ensure reliable use of HTS. The clear contextualisation of the proposals made by the authors complements and clarifies the need for user expertise according to the experimental objectives. Some unanswered questions that will require further development and validation are also presented.

This article should benefit a large audience including researchers with some level of expertise in HTS but unfamiliar with the recent concepts of controls common in the diagnostic world as well as scientists with strong diagnostic expertise but less at ease with the numerous and complex procedures associated with HTS.

References

Bester R, Steyn C, Breytenbach JHJ, de Bruyn R, Cook G, Maree HJ (2022) Reproducibility and Sensitivity of High-Throughput Sequencing (HTS)-Based Detection of Citrus Tristeza Virus and Three Citrus Viroids. Plants, 11, 1939. https://doi.org/10.3390/plants11151939

Deiner K, Bik HM, Mächler E, Seymour M, Lacoursière-Roussel A, Altermatt F, Creer S, Bista I, Lodge DM, de Vere N, Pfrender ME, Bernatchez L (2017) Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. Molecular Ecology, 26, 5872–5895. https://doi.org/10.1111/mec.14350

Massart, S et al. (2022) Guidelines for the reliable use of high throughput sequencing technologies to detect plant pathogens and pests. Zenodo, 6637519, ver. 3 peer-reviewed and recommended by Peer Community in Infections. https://doi.org/10.5281/zenodo.6637519

Piper AM, Batovska J, Cogan NOI, Weiss J, Cunningham JP, Rodoni BC, Blacket MJ (2019) Prospects and challenges of implementing DNA metabarcoding for high-throughput insect surveillance. GigaScience, 8, giz092. https://doi.org/10.1093/gigascience/giz092

Reviews
Toggle reviews

Evaluation round #1
DOI or URL of the preprint: *https://doi.org/10.5281/zenodo.6637518*

Version of the preprint: 1

Author's Reply, 06 Sep 2022

Download author's replyDownload tracked changes file
Dear reviewers and recommender,

We would like to thank you for the review of the document. All your comments and suggestions have been addressed in the attached document and the text edited accordingly.

Some co-authors also made slight corrections in the document (also visible through track change)

Kind regards,

Sébastien Massart (on behalf of the co-authors)

Decision by Olivier Schumpp, 06 Aug 2022

Dear authors,

Thank you for submitting your manuscript to PCI Infection. I particularly enjoyed reading it. The manuscript was evaluated by 2 reviewers and both found it to be of great interest and high quality. However, both raised a number of points which I invite you to address before a final decision is made.

For my side, I have two additional minor comments:

- L1091: ... "and Massart S." (?)
- Diagram of figures 1 and 3: should there not be an exit arrow at the end of the verification and/or validation steps?

I look forward to receiving your revised version.

Best regards,

Olivier Schumpp

Reviewed by Denise Altenbach, 22 Jul 2022

The authors present guidelines for using high-throughput sequencing to detect plant pathogens and pests, as they elaborated in the EU-funded project VALITEST. In the words of the authors, it includes "all the key phases to ensure reliable use of HTS technologies". Together with a companion paper (Reference section: No. 12 Lebas et al., EPPO Bull.) where the authors describe "the steps of the laboratory and bioinformatics components", agricultural diagnosticians have, for the first time, a comprehensive yet concise resource for setting up and running HTS based diagnostics in the strictly regulated plant health sector.

This manuscript is well structured and written and it contains excellent examples to enhance understanding and diagrams and decision trees to visualize the workflows. We have hence only very minor comments:

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

Many references are incomplete and their format are not always the same

Examples:

Ref 12: how to look it up?

Ref 24: submitted paper

Ref 35, 38: no page numbers. Adding the "doi.no" would be greatly appreciated

- 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.

- 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.

-     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".

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

This publication is of great interest for the plant health community. We greatly recommend its publication.

Denise Altenbach and Laure Apothéloz

Dr. Denise Altenbach

Head of Group Molecular Diagnostics of Regulated Plant Pests

Federal Department of Economic Affairs, Education and Research EAER

Agroscope

Methods Development and Analytics

Reviewed by David Roquis, 20 Jul 2022

Dear authors:

I have read the manuscript with great interest. Overall, I found the manuscript to be of an excellent quality, and ready for publication. I only have a few comments/suggestions that may help to improve the manuscript.

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.

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.

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).

Minor comments

99: trees
252: artifically

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.

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

578: i.e. or e.g.