anses

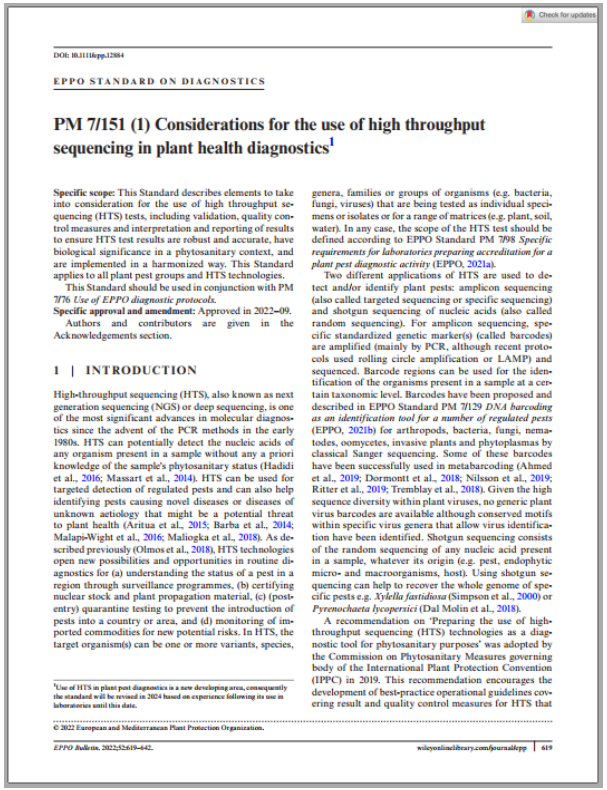
INTERPRETATION OF HTS RESULTS IN PLANT VIRUS DIAGNOSTICS: LESSONS LEARNED AND COMPARATIVE PRACTICES

FRENCH AGENCY FOR FOOD, ENVIRONMENTAL AND OCCUPATIONAL HEALTH & SAFETY

Mathieu Rolland

1 — Guidelines

PM 7/151 (1) Considerations for the use of high throughput sequencing in plant health diagnostics (& Valitest D2.2; Lebas B., Massart S.)



DOI: 10.1111/epp.12884

EPPO STANDARD ON DIAGNOSTICS

PM 7/151 (1) Considerations for the use of high throughput sequencing in plant health diagnostics[1]

Specific scope: This Standard describes elements to take into consideration for the use of high throughput sequencing (HTS) tests, including validation, quality control measures and interpretation and reporting of results to ensure HTS test results are robust and accurate, have biological significance in a phytosanitary context, and are implemented in a harmonized way. This Standard applies to all plant pest groups and HTS technologies.

This Standard should be used in conjunction with PM 7/76 *Use of EPPO diagnostic protocols*.

Specific approval and amendment: Approved in 2022–09.

Authors and contributors are given in the Acknowledgements section.

1 | INTRODUCTION

High-throughput sequencing (HTS), also known as next generation sequencing (NGS) or deep sequencing, is one of the most significant advances in molecular diagnostics since the advent of the PCR methods in the early 1980s. HTS can potentially detect the nucleic acids of any organism present in a sample without any a priori knowledge of the sample's phytosanitary status (Hadidi et al., 2016; Massart et al., 2014). HTS can be used for targeted detection of regulated pests and can also help identifying pests causing novel diseases or diseases of unknown aetiology that might be a potential threat to plant health (Aritua et al., 2015; Barba et al., 2014; Malapi-Wight et al., 2016; Maliogka et al., 2018). As described previously (Olmos et al., 2018), HTS technologies open new possibilities and opportunities in routine diagnostics for (a) understanding the status of a pest in a region through surveillance programmes, (b) certifying nuclear stock and plant propagation material, (c) (post-entry) quarantine testing to prevent the introduction of pests into a country or area, and (d) monitoring of imported commodities for new potential risks. In HTS, the target organism(s) can be one or more variants, species, genera, families or groups of organisms (e.g. bacteria, fungi, viruses) that are being tested as individual specimens or isolates or for a range of matrices (e.g. plant, soil, water). In any case, the scope of the HTS test should be defined according to EPPO Standard PM 7/98 *Specific requirements for laboratories preparing accreditation for a plant pest diagnostic activity* (EPPO, 2021a).

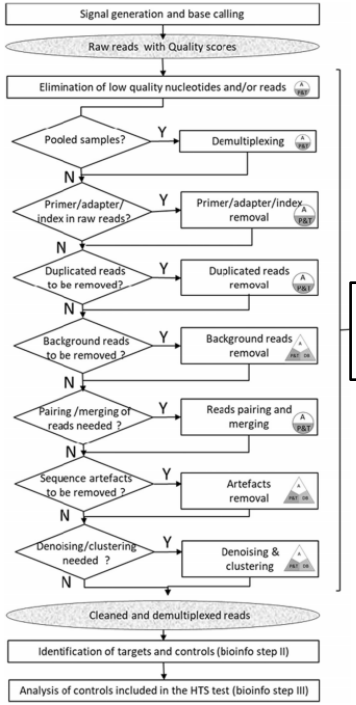
Two different applications of HTS are used to detect and/or identify plant pests: amplicon sequencing (also called targeted sequencing or specific sequencing) and shotgun sequencing of nucleic acids (also called random sequencing). For amplicon sequencing, specific standardized genetic marker(s) (called barcodes) are amplified (mainly by PCR, although recent protocols used rolling circle amplification or LAMP) and sequenced. Barcode regions can be used for the identification of the organisms present in a sample at a certain taxonomic level. Barcodes have been proposed and described in EPPO Standard PM 7/129 *DNA barcoding as an identification tool for a number of regulated pests* (EPPO, 2021b) for arthropods, bacteria, fungi, nematodes, oomycetes, invasive plants and phytoplasmas by classical Sanger sequencing. Some of these barcodes have been successfully used in metabarcoding (Ahmed et al., 2019; Dormontt et al., 2018; Nilsson et al., 2019; Ritter et al., 2019; Tremblay et al., 2018). Given the high sequence diversity within plant viruses, no generic plant virus barcodes are available although conserved motifs within specific virus genera that allow virus identification have been identified. Shotgun sequencing consists of the random sequencing of any nucleic acid present in a sample, whatever its origin (e.g. pest, endophytic micro- and macroorganisms, host). Using shotgun sequencing can help to recover the whole genome of specific pests e.g. *Xylella fastidiosa* (Simpson et al., 2000) or *Pyrenochaeta lycopersici* (Dal Molin et al., 2018).

A recommendation on 'Preparing the use of high-throughput sequencing (HTS) technologies as a diagnostic tool for phytosanitary purposes' was adopted by the Commission on Phytosanitary Measures governing body of the International Plant Protection Convention (IPPC) in 2019. This recommendation encourages the development of best-practice operational guidelines covering result and quality control measures for HTS that

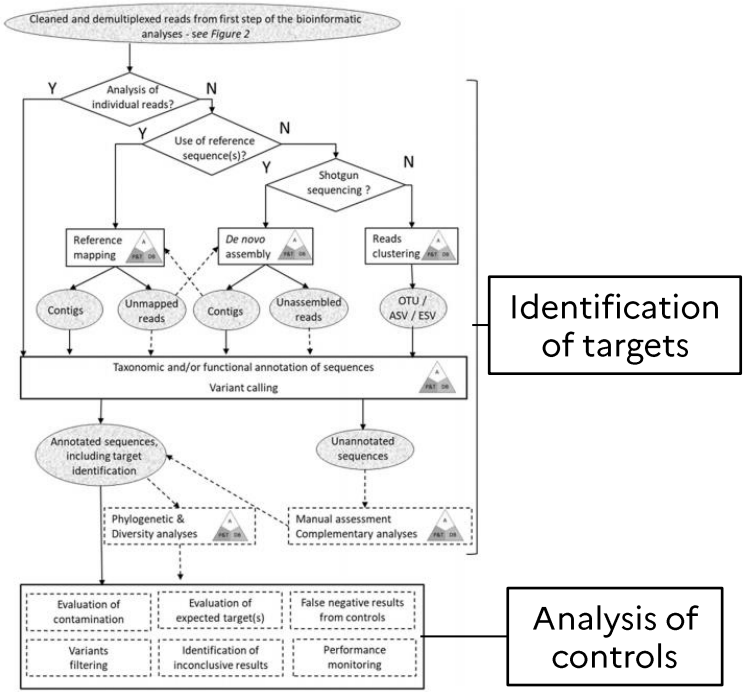
[1]Use of HTS in plant pest diagnostics is a new developing area, consequently the standard will be revised in 2024 based on experience following its use in laboratories until this date.

- Definitions
- Requirements (general and technical)
- Test
  - Selection
  - Development and optimisation
  - Validation and verification

PM 7/151 (1) Considerations for the use of high throughput sequencing in plant health diagnostics (& Valitest D2.2; Lebas B., Massart S.)



Analysis of raw reads



Identification of targets

Analysis of controls

Definition and calculation of indicators !

Interpretation rule ?

Addendum to PM 7/151: Examples of (HTS) tests for the detection and identification of viruses and viroids (NIVIP, Fera and NIB)

To declare a virus present:

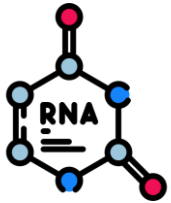
- Clustering analysis of the structurally annotated (near) complete virus and viroid sequences is used for (species) identification, according to the ICTV demarcation criteria

- Set a threshold to filter out false positive results [...] fraction of the reference sequence(s) covered by reads of the same virus exceeds 20% (note, this value needs to be defined in a validation of a test).

2 — Design and validation of an HTS test

Description of the test

Grinding



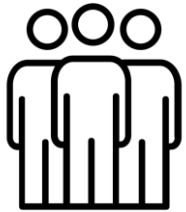
Total RNA extraction
(RNeasy)

Ribodepletion and
library preparation

Illumina sequencing 2x150bp

Bioinformatics
(eVIDanses)

- Fastp
- FastQC
- Spades
- Blast N & X
- Bowtie 2
- Krona



Results exploitation

Key parameters: scope of the test

The purpose of the test is to detect any known or unknown plant virus in a tested sample.

PM7-151 (1): Because HTS tests target a broad range of organisms, it is not possible to validate them for all possible combinations of organism, host or matrix. The validation of the HTS test should focus on key representatives of the targets/pests and use samples that mimic the concentration and composition of real samples expected to be tested

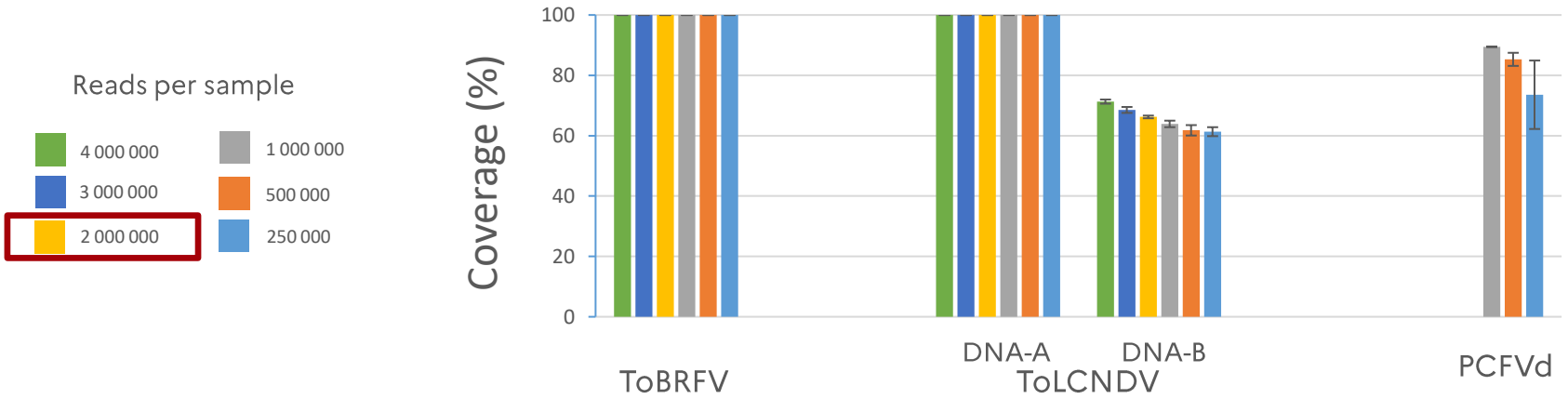
⇒ Validation limited to ToBRFV (RNA virus), ToLCNDV (DNA virus) and PCFVd (Viroid)

⇒ Accreditation (ongoing project) limited to the these viruses and viroids

Key parameters: number of reads (Gb) per sample

PM7-151 (1): The minimal number of reads per sample is determined during test development and can be re-evaluated during validation by the bioinformatic analysis. The generated reads for a sample can be rarefied by randomly selecting part of them. This rarefaction will generate subsamples of reads corresponding to variable lower sequencing depths. The bioinformatic analyses of all these subsamples will identify the sequencing depth(s) at which a target is no longer detected.

Coverage according to the number of reads considered after bioinformatic subsampling



Key parameters: database

GenBank:
$3.9\ 10^{19}$ b

Core_nt:
used for BLAST

In-house
« Phytovirus »
base:
$3.7\ 10^{8}$ bases

⇩

Draw up a list of viruses suspected of being present

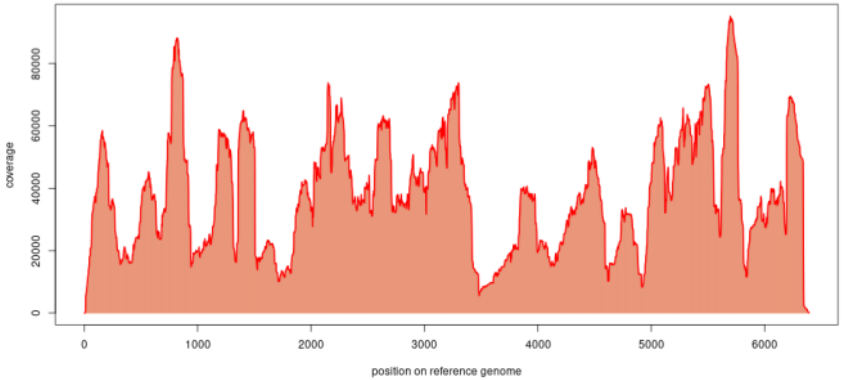
Selection of currated genomes (Qbank)

⇩

Confirm the presence and identity

Interpretation

Reads of a ToBRFV contaminated sample mapped on ToBRFV



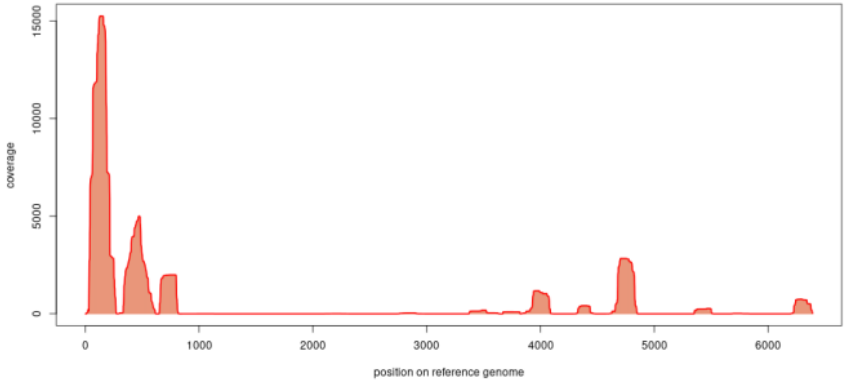
⇒ Complete coverage is ideal

⇒ But, by requiring a near-complete genome, there is a risk of missing viruses present at a lower concentration.

⇒ Which indicators ?
Which threshold ?

Interpretation

Reads of tomato mosaic virus (ToMV) contaminated sample mapped on ToBRFV



=> Tobamoviruses (81% identity)

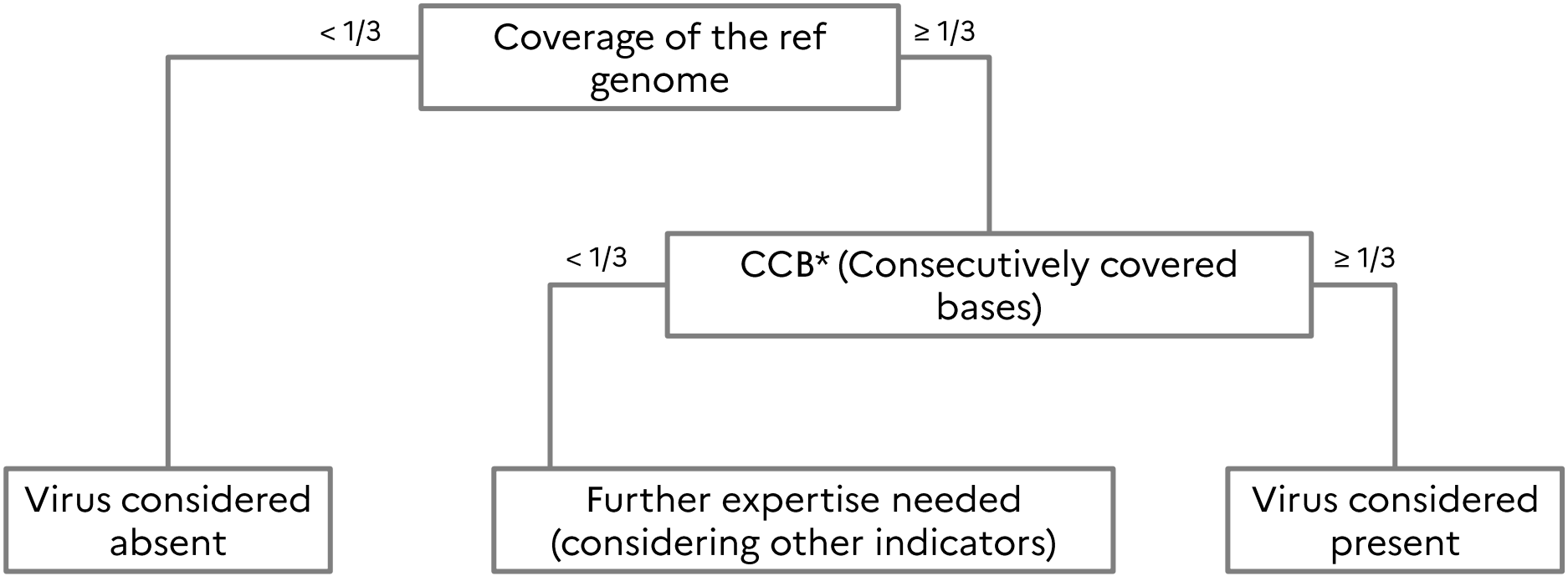
- Mean Read Depth (MRD): 631
- Coverage: 50,59%

- Consecutively covered bases (CCB): percentage of the reference genome represented by the longest continuously covered region

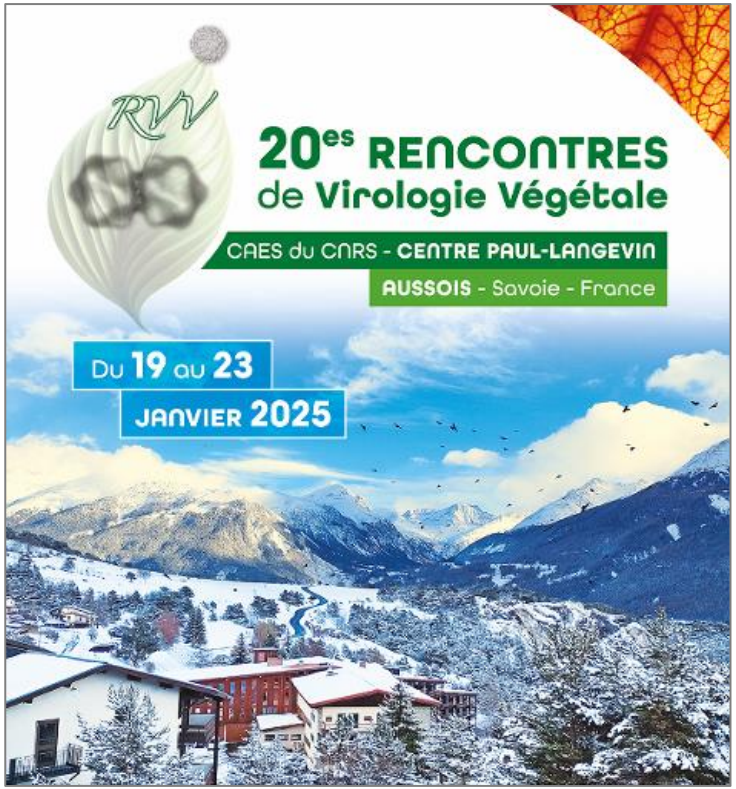
CCB: 9,56%

Interpretation

Scheme that seems adapted to our conditions



3 — Comparative practices



- Diversity of the approaches
- Passionate debates on the interpretation of the HTS results

Diversity of interpretation strategies in laboratories

Illumina
RNASeq
1 million reads per sample

In-house data cleaning pipeline
Interpretation: suspicion from a 200b sequence

Up to →

Illumina
RNASeq
2x25 millions reads per sample

Interpretation rules:
- Threshold for the nb of reads
- Coverage > 80 %
- Identity > 85 %

Conclusions

PM7/151 & addendum: very valuable documents

No turn-key solution, setting up an HTS test remains puzzling

Variable interpretation practices

Still room for harmonization