This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 773139

Grant agreement N. 773139

DELIVERABLE N° 2.2

Title:

"Best practice" guidelines for validation and routine use of non-targeted techniques in diagnostic setting which could serve as a basis for a new EPPO Standard

Valitest

Validation of diagnostic tests to support plant health

Due date:	Month M30
Actual submission data	16-12-2020 (Month M32)
Start date of the project	01-05-2018
Deliverable lead contractor (organization name)	ULg
Participants (Partners short names)	ANSES, CREA, EPPO, FERA, GIORiN, NIB, ULg, UNITO, WBF, WR
Author(s) in alphabetical order	Lebas B., Massart S.
Contact for queries	sebastien.massart@uliege.be
VALITEST reviewers (in alphabetical order)	Adams I., Fox A., Gentit P., Kutjnak D., Mehle N., Petter F., Remenant B., Spadaro D., Trontin C.
External reviewers (in alphabetical order)	Al Rwahnih M., Baeyen S., Bilodeau G., Blouin A., Boonham N., Bruinsma M., Candresse T., Chandelier A., De Jonghe K., Gaafar Y., Haegemans A., Ho W., Hurtado-Gonzales O., Jonkers W., Kreuze J., Landa B., Leite Vicente C., Liefting L., Liu M., Maclot F., Malapi-Wight M., Maree H., Martoni F., Minafra A., Mollov D., Moreira A., Nakhla M., Piper A., Ponchart J., Rae R., Rivera Y., Rodoni B., Roenhorst A., Rollin J., Saldarelli P., Santala J., Souza-Richards R., Studholme D., Sultmanis S., van der Vlugt R., Tamisier L., Wetzel T., Ziebell H.
Level of dissemination	Public
Type of deliverable	Report

Abstract

These guidelines provide recommendations on the process of selection, development, validation and verification of high-throughput sequencing (HTS) tests and the quality assurance for their routine use as diagnostic tests in plant health laboratories. These guidelines are general to enable a broad application in all plant health fields with appropriate flexibility to account for the changes of technologies.

The guidelines cover all steps from sample collection to reporting including risk analysis, test selection/test development, test validation/verification, quality checks to ensure the validity of the results, confirmation of the detection and interpretation of biological relevance within a quality management system. Furthermore, these guidelines provide an overview of the HTS technologies and their applications in plant health.

Partners involved: ANSES, CREA, EPPO, FERA, GIORiN, NIB, ULg, UNITO, WBF, WR

TABLE OF CONTENTS

TERMS, ABBREVIATIONS AND DEFINITIONS 6
1 Purpose 6
2 Scope 6
3 Introduction 6
4 Overview of the HTS process in plant health diagnostics 9
5 General requirements 12
5.1 Facilities and environmental conditions 12
5.1.1 Laboratory infrastructure 12
5.1.2 IT infrastructure 13
5.2 Personnel 13
5.3 Quality management 14
5.4 Outsourcing 15
5.5 Managing modifications 15
5.6 Data management 16
5.6.1 Data backup and storage 16
5.6.2 Data transfer 16
5.6.3 Reference sequence database 16
6 Technical requirements 19
6.1 Scope of HTS tests 19
6.2 Special considerations for HTS tests 19
6.2.1 Laboratory component 20
6.2.1.1 Sampling 20
6.2.1.2 Sample handling 20
6.2.1.3 Nucleic acid extraction 21
6.2.1.4 Library preparation 22
6.2.1.5 Pooling of libraries 23
6.2.1.6 Sequencing platforms and methods 24
6.2.1.7 Contamination 24
6.2.2 Bioinformatic component 25
6.2.2.1 Analysis of raw reads 26
6.2.2.2 Identification of target(s) 30
6.2.2.3 Analysis of controls included in the HTS test 34
6.3 Risk analysis 35
6.3.1 Risk analysis for HTS test selection/test development 36

6.3.2 Risk analysis for validation/verification of HTS test 36
6.3.3 Risk analysis for ensuring the validity of results 44
6.4 Test selection 44
6.5 Test development and optimisation 45
6.6 Validation and verification of HTS test 46
6.6.1 Validation of HTS test 46
6.6.1.1 Specific considerations for analytical sensitivity 48
6.6.1.2 Specific considerations for analytical specificity 49
6.6.1.3 Other criteria: selectivity, repeatability and reproducibility 50
6.6.2 Verification of an existing validated HTS test 51
6.6.3 Impact of changes made to a validated HTS test 51
6.7 Ensuring the validity of results 52
6.7.1 Reference material 53
6.7.2 Internal and external quality checks 55
6.7.2.1 First line controls 55
6.7.2.2 Replicates 60
6.7.2.3 Second and third line controls and performance monitoring 60
7 Confirmation, biological interpretation and reporting 61
7.1 Confirmation of the identity of the pest(s) 61
7.2 Interpretation of the biological relevance of the identified target(s) 64
7.3 Reporting 66
7.3.1 General recommendations 66
7.3.2 Inconclusive results 66
7.3.3 Detection of unexpected organisms 67
7.3.4 Additional remarks and disclaimers 67
8 Conclusion 68
REFERENCES 69
APPENDIX 1 – Terms, abbreviations and definitions 84

TERMS, ABBREVIATIONS AND DEFINITIONS

The list of terms, abbreviations and definitions can be found in Appendix 1.

1 Purpose

High-throughput sequencing (HTS) is a powerful tool that enables the simultaneous sequencing of potentially all organisms present in a sample. Because of the growing interest in applying HTS for routine diagnostics in plant health laboratories, the purpose of this document is to provide recommendations to guide these laboratories on this complex process.

2 Scope

These guidelines on HTS have been developed to provide recommendations to guide plant health diagnostic laboratories on the selection, development and optimisation of HTS tests, on their validation and verification, and their routine application in plant pest diagnostics, including the use of internal and external quality checks, the interpretation and reporting of HTS test results.

They are relevant for plant health diagnostic laboratories that intend to routinely use HTS technologies for the detection and identification of any plant pest (e.g. arthropods, bacteria, fungi, nematodes, invasive plants, protozoan, viroids, viruses or weeds) from any type of matrices (e.g. pure microbial culture, plant tissue, soil, water) regardless of the type of HTS technology (e.g. amplicon sequencing, shotgun sequencing) and their application (e.g. surveillance programme, phytosanitary certification, crop protection).

3 Introduction

High-throughput sequencing (HTS), also known as next generation sequencing (NGS) or deep sequencing, is probably the most significant advance in molecular diagnostics since the advent of the PCR methods in the early 1980s (Mullis *et al*., 1986). With the exception of some HTS technologies such as amplicon sequencing, HTS can potentially, without any *a priori* knowledge on the sample infectious status, detect the nucleic acids of any organism, including variants and uncharacterized organisms present in a sample that might be potential threats to plant health (Hadidi *et al*., 2016; Massart *et al.*, 2014).

Application of HTS in plant pest diagnostics, surveillance and certification

One of the most frequent applications of HTS in plant pest diagnostics is the identification of pests causing novel diseases or diseases of unknown aetiology. Some relevant examples include the discovery of hundreds of previously uncharacterized viruses associated with symptomatic plants (Barba *et al*., 2014; Maliogka *et al*., 2018), novel strains of the bacteria *Xanthomonas* species strain Nyagatare (Aritua *et al*., 2015) or fungal pathogens as *Calonectria pseudonaviculata* (Malapi-Wight *et al*., 2016).

HTS technologies allowed a rapid increase of the number of sequenced pest genomes. For example, the first completed genome of the uncultured '*Candidatus* Liberibacter asiaticus' bacterium was obtained from DNA extracted from a single psyllid (Duan *et al*., 2009) and the genomes of four uncultured phytoplasmas belonging to the 16SrIII group (Vaccinium witches' broom phytoplasma, Italian clover phyllody phytoplasma strain MA, Poinsettia branch-inducing phytoplasma strain JR and Milk-weed yellows phytoplasma) were obtained through HTS technology (Palmano *et al*., 2012). The first complete genome of the bacterium *Xylella fastidiosa* using shotgun sequencing provided new insight in the pathogenicity mechanisms (Simpson *et al*., 2000). The first draft genome of the fungal pathogen *Pyrenochaeta lycopersici* was obtained from single-molecule real-time sequencing (Dal Molin *et al*., 2018) and a joint international collaboration has the objective of sequencing 1,000 fungal genome across the fungal tree of life using HTS technology in a five-year long project (https://mycocosm.jgi.doe.gov/mycocosm/home/1000-fungal-genomes). The sequencing data generated by metabarcoding of groups of organisms such as fungi, generate useful information for taxonomic purposes (Crous *et al*., 2015). Sequencing the genome of many isolates/specimens/strains of a pest provides a better view of its genetic diversity and, consequently, improves the design of primers and targeted diagnostic protocols (Adams *et al*., 2018). Indeed, the use of HTS sequencing data has facilitated the design of targeted diagnostic protocols for viruses, such as little cherry viruses (Katsiani *et al*., 2018), bacteria such as *Dickeya* spp. (Pritchard *et al*., 2016), *Pseudomonas coronafaciens* (An *et al*., 2015), fungi such as *Didymella pisi* (Owati *et al*., 2019), *Synchytrium endobioticum* (Bonants *et al*., 2015) and nematodes such as *Bursaphelenchus xylophilus* (Kikuchi *et al*., 2011).

HTS technology has also been used in surveillance programmes, monitoring and source tracking following a PCR-based approach called metabarcoding. The combined use of spore trapping and amplicon sequencing has been recently applied in various ecosystems and areas for surveillance studies of (i) airborne fungi and oomycetes, including plant pathogens (Abdelfattah *et al*., 2019; Aguayo *et al*., 2018; Chandelier *et al*., 2020; Nicolaisen *et al*., 2017; Núñez *et al*., 2017; Nilsson *et al*., 2019; Mbareche *et al*., 2020; Ovaskainen *et al*., 2020; Tremblay *et al*., 2018 and 2019), (ii) insects (Braukmann *et al*., 2019; Elbrecht *et al*., 2019; Piper *et al*., 2019), and (iii) plant species by pollen sampling through traps or honeybees (Bruni *et al*., 2015; Tremblay *et al*., 2019; Núñez *et al*., 2017). Metabarcoding has also been applied for the detection and identification of nematodes in soil to monitor soil biodiversity and ecosystem functioning (Herren *et al*., 2020; Waeyenberge *et al*., 2019). In addition, HTS technology using a shotgun sequencing protocol has even been used to detect plant viruses in waste water (Bačnik *et al*., 2020).

In phytosanitary certification schemes, HTS has been used as a generic method for pest detection in nuclear stock and plant propagation material of grapevine and fruit trees. Its performance was compared with traditional diagnostic tests mainly based on biological indexing, HTS outperforming standard bioassay in virus detection, speed and cost of analysis, and discovery of novel viruses (Al Rwahnih *et al*., 2015; Rott *et al*., 2017). HTS has also been used for the detection of viruses on sugarcane plants in quarantine and on *Ullucus tuberosus* during a plant health outbreak in United Kingdom and for the detection of bacteria on the re-emerging disease of oak, Acute oak declined in United Kingdom which is caused by a polymicrobial complex (Candresse *et al*., 2014; Denman *et al*., 2017; Fox *et al*., 2019). In addition, HTS for the detection of viable pant propagules has been trialled at an internal point of entry in USA (Whitehurst *et al*., 2020).

Potential for routine use of HTS in plant pest diagnostics

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 (for example genomic biosurveillance of forest pests, Hamelin and Roe, 2019; insect surveillance using metabarcoding, Piper *et al*., 2019; airborne inoculum of forest fungal pathogens using metabarcoding, Chandelier *et al*., 2020), (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. With the democratisation of sequencing through cost reductions, availability of effective and accessible platforms such as MinION (Oxford Nanopore Technologies), and the improvement in accessibility of bioinformatic tools in analyzing sequencing data, there is an increasing interest in applying these technologies for routine diagnostics including regulatory plant health diagnostics. To date, HTS has only been used in plant health diagnostics in specific situations in quarantine, during the investigation of a plant health outbreak or re-emerging disease (Candresse *et al*., 2014; Denman *et al*., 2017; Fox *et al*., 2019). HTS is currently mentioned in only one EPPO standard, PM 3/21 (2019) which describes inspection and tests for the detection of pests (bacteria, viroids and viruses) infecting tuber-forming *Solanum* species or hybrids imported for germplasm conservation, breeding or research purposes and in post-entry quarantine.

A need for HTS guidelines for plant pest diagnostics

A recommendation on preparing the use of HTS 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 highlightes the need for robust test design, test validation and quality assurance (FAO, 2019).

One of the main challenges for using HTS as a routine diagnostic test in plant health laboratories is the implementation of guidelines that are internationally recognized for the selection, development, validation and routine use of HTS-based tests. The performance criteria of HTS tests need to be defined before this technique can be used, and appropriate controls have to be put in place to ensure the validity of the results (Maree *et al*., 2018). A range of factors needs to be considered during the test development and optimisation such as establishing the quality metric thresholds and their acceptable range of values. Challenges related to the laboratory procedures and the bioinformatic analyses of the HTS-based test process need to be addressed (e.g. personnel, infrastructure, equipment) as well as the interpretation of the results and their biological significance (Olmos *et al*., 2018).

Baseline workshops for proposing the guidelines

The present guidelines were developed based on the IPPC recommendation following discussions between HTS experts during three meetings organized in the framework of the European COST Action DIVAS (Deep Investigation of Virus Associated Sequences), FA 1407 (Bari, November 2017; Brussels, February 2018; Liège, November 2018). During these meetings, HTS guidelines and recommendations from the human health area (e.g. Hébrant *et al*., 2017; Jennings *et al.,* 2017; Rehm *et al.,* 2013) were used to develop the bases of plant health guidelines. Other guidelines from different fields (e.g. animal health [OIE, 2018]) were also considered during the development of the present guidelines.

The present guidelines provide technical recommendations from sample handling to reporting of results with emphasis on aspects specific to HTS (e.g. data management, specialised personnel, bioinformatic analyses). They include recommendations on test selection, development and optimisation, validation and verification, internal, and external quality checks and interpretation of biological relevance, overarched by a quality management system.

The present guidelines have been developed, irrespective of the chemistry, instrumentation and software, and apply to any plant pest in any matrix. They have been designed to allow flexibility within this fast-evolving technology.

4 Overview of the HTS process in plant health diagnostics

Two different HTS technologies are mainly used to detect plant pests: sequencing of amplicon generated by PCR, RCA (e.g. metabarcoding) or any other relevant protocol (e.g. metabarcoding, amplicon sequencing) and shotgun sequencing of nucleic acids (also called random sequencing). The full HTS process can be divided into eight distinct steps (Figure 1), for which a short description is provided below. It is important to remember that for each step, a diversity of procedures have been developed and published and further improvements of the current protocols can reasonably be expected in the near future.

Step 1: Sampling. Sample requirements for HTS tests are similar to any other diagnostic test. The matrix to be sampled can contain multiple organisms (e.g. plant tissue harbouring microorganisms including pests, environmental samples, spore traps, insects with their microbiota) or can consist of isolated organisms (e.g. microbial colonies isolated on artificial media).

Step 2: Nucleic acid extraction. The nucleic acids can be genomic DNA or genomic RNA, total DNA or RNA, small RNAs, double-stranded RNAs. RNA extracts are usually DNase treated before being reverse transcribed into complementary DNAs.

Step 3: Library preparation. The nucleic acids are prepared in a format compatible with employed sequencing, for example amplicon sequencing (e.g. metabarcoding; step 3a of Figure 1) and shotgun sequencing of nucleic acids (step 3b of Figure 1). This process depends on the sequencing platform requirements. The library preparation produces a sufficient amount of nucleic acids of appropriate size that are flanked with adapters (oligonucleotide sequences) required for sequencing.

For **amplicon sequencing** (also called targeted sequencing or specific sequencing), specific genomic regions are amplified (mainly by PCR, although recent protocols used rolling circle amplification or LAMP) and sequenced. In this case, the primers used are usually composed of, at the 3'-end, the sequence complementary to the target sequence allowing the amplification of the target region and, at the 5'-end, an adapter sequence specifically designed by the sequencing platform provider Another option is to first perform a PCR with the target primers followed by a PCR with overlapping primers containing an adapter or an adapter ligation step. Alternatively, the amplified products (for example: long amplified regions) can be fragmented and further shotgun sequenced as previously described.

For **shotgun sequencing**, library preparation protocols consist of classical molecular techniques, including shearing (sonication) or digestion (restriction enzymes or chemical lysis) of nucleic acids, end-repair, ligation of oligonucleotides (adapters). These adapters correspond to oligonucleotides designed by the sequencing platform provider and are mandatory to allow the sequencing reaction to proceed. These steps can be complemented by a PCR amplification. Alternatively, random hybridization and amplification using degenerated oligonucleotides or use of transposases can be applied to fragment nucleic acids (van Opijnen and Camilli, 2013; Wilcox *et al*., 2018). Appropriate adapters are then ligated to one or both ends of the sample's fragmented nucleic acids. A reverse-transcription step is also applied when starting from RNA although direct RNA sequencing protocols are under development (see MinION of Oxford Nanopore Technologies).

Optional Target enrichment or selection (opt. 1 on Figure 1). Optionally, the targeted nucleic acids can be enriched during the nucleic acid extraction or the library preparation process. For shotgun sequencing, the target enrichment or selection can be done by removing untargeted sequences (for example removal of plant rRNA, called ribodepletion, enriching dsRNA by nuclease digestion) or by using oligonucleotides specific to the targets, sometimes also referred to as probe capture (Adams and Fox, 2016; Gaafar and Ziebell, 2020; Maliogka *et al*., 2018).

Optional pooling (opt. 2 on Figure 1). Depending on the library preparation protocol, there is the option to add a unique identifier, called index (also known as barcode, tag, molecular Identifier (MID)) to each sample to allow the sequencing of a pool of samples, called multiplexing. Analysis of pooled samples in a single sequencing run allows a reduction of the costs of the analysis per sample. The index is a short sequence of oligonucleotides flanked with the adapter added during the library preparation and is unique to each sample. This index is sequenced along with the target and allows each sequence to be linked to the appropriate sample in the pool. Indexes can be added to one side or to both sides (dual indexing) of the fragment.

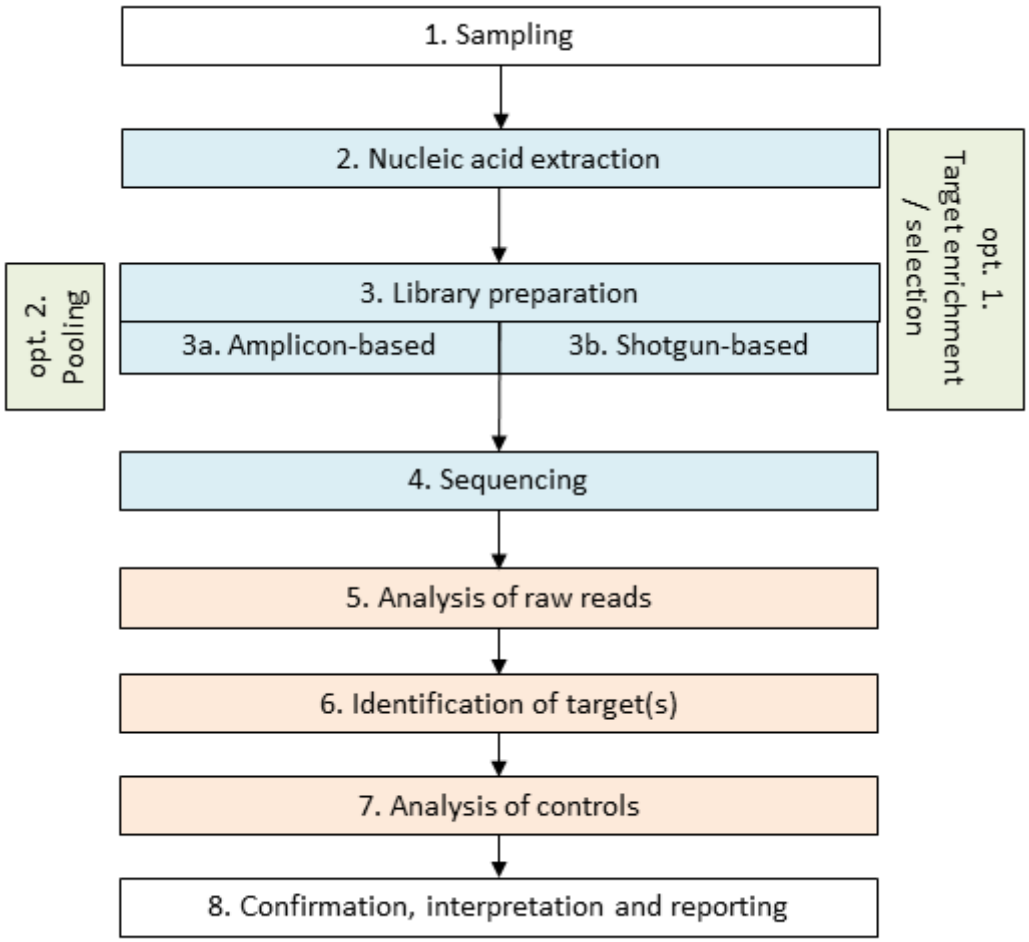

Figure 1. Scheme representing the eight main steps of HTS technologies used in plant health diagnostics. The laboratory steps are highlighted in blue and the bioinformatic steps in orange. Two optional steps are included in green: opt. 1. the target enrichment or selection and opt. 2. the pooling of samples (for both types of library preparation).

Step 4: Sequencing. Currently, there is a limited number of sequencing platforms that have been widely commercialized and the most important ones are 454 pyrosequencing (Roche, discontinued from 2016), Ion Torrent (Thermo Fisher Scientific), Illumina sequencing by synthesis (Illumina), PacBio single molecule real-time (SMRT) (Pacific Biosciences) and nanopore sequencing (Oxford Nanopore Technologies). They have already been described in detail and reviewed in several publications (Nilsson *et al*., 2019; Maljkovic Berry *et al*., 2020; Pervaiz *et al*., 2017; Reuter *et al*., 2015).

The sequence data generated by the HTS platforms are delivered together with their corresponding quality scores, usually in FASTQ format. This is a text-based format for storing both the nucleotide sequence data and the corresponding quality score for each base. Both are encoded with a single ASCII character for brevity (Cock *et al*., 2010).

Step 5: Analysis of raw reads. This bioinformatic step consists of several operations including the quality control of the generated sequences (allowing the elimination of low quality sequences and nucleotides or potential artefacts), the (optional) elimination of adapter, index and primer sequences. In the case of pooled samples, demultiplexing enables the correct assignment of the generated sequences to each sample. After the first step of the bioinformatic analyses, the remaining sequences are grouped by samples and are ready for target identification (i.e. the second step of bioinformatic analyses). Optionally, some additional analyses can be performed to reduce the data, for example merging forward and reverse reads based on the overlapping region (if present), or removing duplicate reads.

Step 6: Identification of target(s). This bioinformatic step aims to identify sequences, also called sequence annotation. Sequence annotation can be either taxonomic (e.g. giving a taxonomic position) or functional (e.g. coding region, intron, promoter, miRNA, lncRNA, transposon, repeated sequences) depending on the intended use of the HTS test. Target identification always relies on the comparison with existing annotated sequences in a database. It can be performed in different ways: (i) on the individual reads (read annotation or read classification), (ii) after *de novo* assembly of the reads into contigs or (iii) after mapping of the reads on reference sequences (reference assembly) or a combination of these. In metabarcoding, reads are grouped into representative bins or clusters called operational taxonomic units (OTUs) or amplicon sequence variants (ASV) and then compared against sequence reference database(s) to identify the target(s). Alternatively, reads with artefacts introduced during PCR amplification (noisy sequences, e.g. nucleotide substitutions, length variation, chimeras) can be removed before the OTU clustering (this is called denoising). Furthermore, when the reads have been assembled *de novo* or mapped on a reference sequence or grouped into OTUs, variants, corresponding to single nucleotide polymorphism (SNP), insertion and deletion of nucleotides (indels) or to the integration or deletion of genes (structural variants [SV]), can be identified.

Step 7: Analysis of controls. This analysis aims to verify that all the controls included in the HTS run produced the expected results in order to identify and eliminate potential false positive and/or negative results before moving on to step 8. Some controls, like the positive control, can be analysed before processing the samples to validate the identification of all the expected targets but the complete analysis of controls, including the evaluation of inter-samples contamination relies on the target identification.

Step 8: Target confirmation, interpretation and reporting. The last step of the HTS test consists of (i) the confirmation of the identity of the target(s) detected in the sample(s), (ii) the interpretation of the biological and phytosanitary relevance of the target(s) identified in a sample (in particular for uncharacterized organisms) and (iii) the reporting of the results of the HTS test.

5 General requirements

5.1 Facilities and environmental conditions

Recommendation: **The laboratory should have, or have access to, appropriate facilities and information technology (IT) infrastructures to perform HTS tests.**

5.1.1 Laboratory infrastructure

Contamination in HTS tests is particularly problematic because of the multiple handling steps and use of many different reagents in the sample preparation process. A few contaminating reads can easily be detected among millions of generated reads because of the very low analytical sensitivity of the technology even lower than PCR-based techniques that need more copies of the target for its detection. Further details on the sources of contamination and the monitoring of the level of contamination using first line controls can be found in section 6.2.1.7 – Contamination. To minimise the risk of contamination, the laboratory should have a clear workflow following the HTS process with the following laboratory practices:

• Dedicated equipment (including pipettes) and appropriate molecular grade reagents and consumables should be used in each work area. A dedicated biological hood with UV light and the air fan off can be used to work in a more protected environment free of DNA/RNA from exogenous sources, lowering the risk of contamination and allowing to clean and sterilise the working station more thoroughly.

• Dedicated laboratory coats should preferably be used in each work area (at least a specific coat for master mix preparation) and gloves should be worn.

• Tubes containing amplified products should not be opened within work areas used for nucleic acid extraction or master mix/reaction mixture preparation. If possible, they should be opened under extraction fume hoods within the amplification analysis area.

The implementation of HTS technologies requires a molecular laboratory to carry out nucleic acid extraction, library preparation, nucleic acid amplification and sequencing. The recommendations for dedicated PCR work areas as described in EPPO standard PM 7/98 (2019) are applicable to HTS tests and should be followed. Noteworthy, library preparation step can include several nucleic acid amplifications and the forward flow principle should be maintained all the time.

There should be dedicated work areas for:

(a) nucleic acid extraction,
(b) library preparation (amplicon or shotgun),
    (b1) preparation of master mix,
    (b2) addition of sample to the master mix,
    (b3) enzymatic reaction among which nucleic acid amplification,
    (b4) analysis of amplification products and
(c) sequencing.

Some steps of the laboratory protocol, for example the nucleic acid extraction, library preparation, running the sequencing instrument or the bioinformatic analyses might be outsourced to external sequencing service and bioinformatics providers (see section 5.4 - Outsourcing).

5.1.2 IT infrastructure

The implementation of HTS requires significant investment in information technology and bioinformatics. Large files (up to a few gigabytes per sample) are generated and need to be stored and properly backed-up on remote internal and/or external servers for a duration of time that meets customer and legal requirements (see section 5.6.1 – Data backup and storage). For example, a current (in 2020) sequencing run using Illumina technology will generate 25 and 400 millions of reads on the MiSeq and NovaSeq platforms, respectively. These outputs, corresponding to the sequencing of several samples multiplexed, represent a volume of 7,5 and 120 GB, respectively. It is important to mention that these numbers are representative of the current technologies and will evolve in the future. A laboratory planning to implement HTS testing should investigate on the latest capacity of the technology for getting an appropriate IT infrastructure.

Transferring large data files from servers to data analysis computers requires a fast network (see section 5.6.2 – Data transfer) and machines with high computational power for running the bioinformatic pipeline (Olmos *et al*., 2018). Laboratories that do not have extensive data analysis capabilities can outsource the bioinformatic data analysis to external facilities or rent the computational power and storage space on commercially available computer clusters (see section 5.4 - Outsourcing).

The IT infrastructure configuration for storage should take into account the expected number of samples, volume of data per sample (including raw generated reads, intermediate data files and final results), the legal or commercial obligations related to data, the maintenance and data back-up. Further aspects concern the operating system environment (e.g. Windows, MacOS, Linux) and computing power or server required to run the bioinformatic software and bioinformatic personnel to process and analyse the data for timely delivery of results.

5.2 Personnel

Recommendation: **The laboratory should have personnel proven to be competent for the overall HTS process including laboratory and bioinformatic components and for the biological interpretation of the data**.

The use of HTS technologies requires trained personnel with expertise for each step of the process. As with any other molecular diagnostic test, only qualified and trained personnel can process the samples. For the sequence data analysis, specific expertise in IT infrastructure (see section 5.1.2 – IT infrastructure) and bioinformatics (see below) is needed. EPPO's recommendations on the competence and expertise of personnel should be followed: EPPO standard PM 7/84 (2018) for laboratories' basic requirements and EPPO standard PM7/98 (2019) for laboratories preparing for ISO 17025 accreditation. It should be noted that some specific expertise can be outsourced (e.g. development and installation of a pipeline; see section 5.4 - Outsourcing).

The bioinformatic analyses is specific to HTS tests. It requires trained personnel able to run the bioinformatic pipeline correctly (installation, development, validation, routine use and regular update of the software and databases). In addition, relevant scientific expertise is needed for the appropriate interpretation of the data and/or results and for the decision on possible follow-up actions (e.g. confirmatory testing, see section 7.1).

Relevant scientific expertise is also required for the interpretation of results (avoiding reporting of false positive and/or false negative results that could potentially have significant consequences) and of their biological relevance (see section 7.2 – Interpretation of the biological relevance of the identified target(s)). The importance of scientific expertise has been highlighted during a proficiency test on the interpretation of the significance of food-borne pathogens in a simulated dataset (organized by the COMPARE network, http://www.compare-europe.eu/; Brinkmann *et al*., 2019) and on the sequence analysis strategies of small RNAs of plant viruses (as part of the European COST Action FA1407 (DIVAS); Massart *et al*., 2019).

5.3 Quality management

Recommendation: **The laboratory should have a quality management system in place to ensure its consistent operation in performing HTS tests and traceability throughout the process.**

The laboratory should have a quality management system in place (including a documentation system) for quality assurance purposes for example, allowing to trace back the origin of samples or of contamination. The documentation system should describe all the procedures required to perform an HTS test from sampling to reporting, including the laboratory component, the IT infrastructure (e.g. software version), bioinformatic pipeline (e.g. software versions and settings with details on all the parameters, scripts and sequence databases version) and data (e.g. input and output files for each sub-step of the bioinformatic pipeline). The documentation system should also contain procedures on the operation of critical instruments (e.g. sequencing machine) and bioinformatic pipeline. Such requirements are illustrated in Aziz *et al*. (2015), Hébrant *et al*. (2018) and Roy *et al*. (2018).

The procedures should be detailed enough to ensure consistent implementation of HTS tests, including the bioinformatic component. The laboratory should ensure that the procedures are kept up to date and that the current versions are readily available to personnel (EPPO PM 7/98, 2019).

As part of the quality system, the laboratory should keep records of:

- test development documents (when relevant),
- validation/verification reports (when appropriate including those after changes have been made to an HTS test),
- test run reports (including the values of relevant quality metrics, version of software/pipeline/databases),
- diagnostic reports (see EPPO PM 7/77, 2019),
- sample information (see EPPO PM 7/77, 2019),
- administrative information (see EPPO PM 7/77, 2019),
- maintenance and calibration certificates of equipment,
- critical kits/reagents (e.g. lot number, expiration dates of reagents),
- personnel competency and training.

The records should be readily accessible and kept for a minimum period of 5 years as recommended by EPPO standard PM 7/77 (2019), unless the national requirements specify otherwise.

5.4 Outsourcing

Recommendation: **The laboratory should check the quality performance of the outsourced services at appropriate intervals.**

Laboratories can outsource some parts of an HTS test (e.g. nucleic acid extraction, library preparation, sequencing services, bioinformatic services). The outsourcing or subcontracting of some parts of a diagnostic test is included in the ISO 17025 standard (2017) in paragraph 6.6. For plant health diagnostics, EPPO standard PM 7/130 (2016) states that "*Subcontracting of testing by the authorized laboratory is permitted if documented in the quality system manual (EPPO PM 7/84, 2007) and approved by the NPPO* [National Plant Protection Organization]. *The requirements of this standard* [i.e. EPPO PM 7/130, 2016] *also apply to the subcontracting laboratory*."

It is recommended to select a provider that has as at least the same level of quality assurance management as the diagnostic laboratory, ideally with an official accreditation or certification. The assessment of outsourcing should be part of the risk analysis (see section 6.3 – Risk analysis). The outsourced services should be regularly monitored to ensure that the provider performs as expected. For example, the laboratory can demonstrate that outsourcing does not negatively influence the reliability of the results by monitoring the results of sending and analysing blind samples (see section 6.7 – Ensuring the validity of results).

A procedure for the selection and evaluation of external services should be developed and applied by the laboratory. It should be noted that although some parts of the HTS process can be outsourced, the laboratory is responsible for the interpretation and reporting of the results (Hébrant *et al.,* 2018).

5.5 Managing modifications

Recommendation: **The laboratory should have a procedure for monitoring, implementing, and documenting modifications of reagents, kits, sequencing chemistries, instruments and bioinformatic pipelines**.

HTS technologies and protocols evolves quickly, both for the laboratory and bioinformatic components. This leads to the rapid obsolescence of protocols, sequence databases and bioinformatic software. Their lifespan is therefore shorter than most existing diagnostic tests used by a diagnostic laboratory.

Modification of the laboratory protocols, due to new versions of kits or newly developed kits for library preparation or sequencing should be documented. The laboratory should make efforts to keep track of any change based on a procedure on monitoring, implementing and documenting the modifications.

Regarding bioinformatic analyses, pipelines are composed of several sub-steps that use different software applications with specific algorithms and parameters adapted to the intended use of the HTS test. The laboratory should keep track of software versions and updates/upgrades with parameter settings and keep records of changes to the underlying operating systems which might affect how pipelines and tools work (e.g. integrate a Log system to track all versions in the bioinformatic pipeline).

In response to the constantly evolving technology, these guidelines have been designed to allow the modification of the laboratory kits or bioinformatic pipelines under certain conditions. EPPO standard

PM 7/98 (2019) recommends that an expert judgement should be made as to whether the update to a validated HTS test requires validation or verification (see section 6.6.1 – Validation of HTS test and section 6.6.2 – Verification of an existing validated HTS test). This evaluation should be documented.

5.6 Data management

5.6.1 Data backup and storage

Recommendation: **The laboratory should backup and safely store data generated during each HTS test.**

Data can be stored on the laboratory's own data storage system or on a cloud-based computing resource, internally or externally. When data are stored on an external cloud-based computing system, the laboratory should be aware of the local legislation on data protection, especially when dealing with official testing and quarantine pests. The laboratory should have a documented procedure on how and where files generated during sequencing and bioinformatic analyses should be stored to ensure the integrity and confidentiality of the data, including a backup, ideally on a server at another location (Hébrant *et al*., 2018; Roy *et al*., 2018). The laboratory should decide which files (i.e. input files, intermediates files and output files) should be kept and for how long. EPPO recommends a minimum retention period for records of 5 years or longer in the case the national requirements specify otherwise (EPPO PM 7/77, 2019).

5.6.2 Data transfer

Recommendation: **The laboratory should have access to a network allowing a safe, easy and complete transfer of data.**

Large data files are generated during each HTS run. These data sets need to be easily transferable within the IT infrastructure of the laboratory and, if relevant, from the sequencing provider to the laboratory (see section 5.1.2 – IT infrastructure). The network should be secured to ensure the integrity and confidentiality of the data. The laboratory should have a procedure for data transfer, including that any issue related to data transfer should be recorded if they can influence the test results (Aziz *et al*., 2015; Hébrant *et al*., 2018). Numerous algorithms have been developed for such goal (md5sum is one example).

5.6.3 Reference sequence database

Recommendation: **The laboratory should use appropriate sequence databases to analyse the sequence data generated by the HTS test.**

Sequence databases used for reference can be incomplete, contain errors and their content is constantly evolving because of scientific discoveries and changes in taxonomy of pests. The inappropriate selection of sequence database(s) can lead to incorrect results. An inter-laboratory comparison of 21 plant virology laboratories, each using a different bioinformatic pipeline, revealed that the database accuracy and completeness is critical for the identification of the target(s) and most

importantly for the identification of uncharacterized organisms (Massart *et al*., 2019). In this paper, a novel member of the family *Tymoviridae* was detected only by two out of 21 participants. In addition, wrong species identification of a nepovirus was reported because of the incompleteness of the sequence database used for reference (absence of the target(s) sequences and/or incorrectly annotated sequences). Similar observations have also been reported for the metabarcoding of insects and fungi (Nilsson *et al*., 2019; Piper *et al*., 2019) as well as in food safety for the detection of pathogens (Brinkmann *et al*., 2019).

Sequence databases can be available publicly (see box below) or can be developed and maintained by the laboratory (i.e. in-house sequence databases). The selection of sequence database(s) is important for a correct taxonomic assignment (see section 6.2.2.2 – Identification of target(s)), since a reference database containing too few or too many sequences can give false negative or positive results, depending on the taxonomic classification method used. Their selection depends on the intended use of the HTS test and should be clearly defined during the test selection and development (see sections 6.4 – Test selection and 6.5 – Test development and optimisation). When the focus of the HTS test is on a limited range of known pests, a curated database can be created with sequences of high quality that are accurately annotated and not redundant. However, when searching for uncharacterized or unexpected organisms, a more extensive and less curated database might provide a much better chance for their discovery than a well curated database with a limited number of entries. (Lambert *et al*., 2018; Piper *et al.,* 2019).

Sequence databases should be tested for their ability to identify at least the expected target(s) including the key target(s) for HST tests, with the database preferably prepared from documented reference material (for example, vouchered specimens; see section 6.7.1 – reference material). Sequencing of reference material, especially from reference collections, is critical for a correct diagnosis. This highlights the importance of specimens accurately identified taxonomically as a key factor in the compilation of curated sequence databases, avoiding that morphological or phenotypic misidentification lead to incorrectly named sequences on database (e.g. Taylor and Martoni, 2019). Sequence databases '*should be kept up to date and readily available*' (EPPO PM 7/98, 2019). As for computer softwares, information on sequence databases used during the analysis should be documented for traceability purposes. The information to document should include, but not be limited to, version number, date of download, original source, location. It is important to note the version used for the databases; sometimes names of organisms disappeared from one version to the other. Also the laboratory needs to make sure the target organisms are still part of the databases.

Sequence databases (by alphabetical order) currently (in 2020) being used:

BOLD: www.barcodinglife.org (barcode of DNA species)

EMBL-EBI: https://www.ebi.ac.uk/services (wide range of sequences including plant pests)

ENA: https://www.ebi.ac.uk/ena/browser/home (wide range of sequences including plant pests)

EPPO-Q-bank: https://qbank.eppo.int/ (plant pests)

EzBioCloud: https://www.ezbiocloud.net/ (bacteria and archaeal)

GenBank: www.ncbi.nlm.nih.gov (wide range of sequences including plant pests)

Genome Taxonomy Database: https://gtdb.ecogenomic.org/ (bacteria and archaeal)

GreenGenes: https://greengenes.secondgenome.com/ (bacteria and archaeal)

InsectBase: http://www.insect-genome.com (insects)

NEMBASE4: http://www.nematodes.org/nembase4/ (nematodes)

NemaGene: http://nematode.net/NemaGene/.(nematodes)

SILVA: https://www.arb-silva.de/ (ribosomal RNA sequence data)

UNITE: https://unite.ut.ee/ (eukaryotic nuclear ribosomal ITS region)

WormBase: https://wormbase.org/#012-34-5 (nematodes)

The laboratory should endeavour to upload sequence(s) (partial or (near to) complete genome sequences, variants sequences) with biological information when available, to an online database such as the Barcode of Life Data System (BOLD; Ratnasingham and Hebert, 2007) for metabarcoding data or the EPPO Q-bank database (https://qbank.eppo.int/) so the sequence(s) is(are) permanently linked to a voucher specimen [it can be morphological when a non-destructive DNA extraction is used (see section 6.2.1.3 – Nucleic acid extraction)]. Whenever possible, a voucher specimen should be kept by the laboratory and/or stored in a depository such as the International collection of microorganisms from plants (ICMP, https://www.landcareresearch.co.nz/tools-and-resources/collections/icmp-culture-collection/).

Uploading sequences to public sequence databases will assist the scientific community in their quest of identification of organisms. The more complete sequence databases are, the greater the likelihood to detect and identify organisms. This is particularly important for the detection and identification of (yet) uncharacterized organisms. Their availability will be useful for research studies such as studies on newly discovered organisms or genetic diversity studies. The link between the sequence(s) and the biological information will enable future characterisation (e.g. taxonomic description), morphological re-assessments and, consequently, the possibility to correct misidentified sequences and specimens. A bottleneck is to ensure that the contract with the customer or the national legislation allow the sharing of sequence data.

6 Technical requirements

There are different steps a diagnostic laboratory should follow when using HTS technologies in routine diagnostics. Before implementing an HTS test on routine samples, the scope of the test should be defined (see section 6.1) and the type of the test selected (see section 6.4). A newly developed test should be validated (see section 6.6.1) and a test described in an official standard or in a scientific publication with known performance characteristics should be verified (see section 6.6.2). After its validation or verification, the HTS test can be used in routine diagnostic but when changes that can potentially negatively affect its performance characteristics are made to the protocol, it should be verified or validated again (see section 6.6.3).

A risk analysis (see section 6.3) on the different steps of the HTS test and its validation/verification should be carried out after the scope has been defined. The risk analysis will be the basis for establishing the critical parameters and quality checks of the HTS test for routine analyses. The risk analysis should be updated regularly (e.g. change in the level of and/or type of risks) depending on the results of the quality checks during the development, validation or verification phases and during the routine implementation of the test. Due to the specificities of HTS tests as compared to other molecular methods, special considerations for the analytical sensitivity, analytical specificity and selectivity should be addressed during the different phases presented in section 6.2.

6.1 Scope of HTS tests

Recommendation: **The laboratory should define the scope of the HTS test before conducting the risk analysis.**

The scope of the test should be defined as recommended by EPPO: *e.g. detection and/or identification of organism x in matrix y by HTS test z* (EPPO PM 7/98, 2019). 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 for a range of matrices (e.g. plant, soil, water). A clear definition of the use of the HTS test and the matrix tested is important for the risk analysis (see section 6.3). For example, in post-entry quarantine testing, the detection and/or identification by shotgun sequencing of viroids and viruses infecting tuber-forming *Solanum* species imported for germplasm conservation, breeding or research purposes. Another example is the surveillance of insects, bacteria or fungi collected from traps using amplicon sequencing (Aguayo *et al*., 2018; Núñez *et al*., 2017; Piper *et al*., 2019).

6.2 Special considerations for HTS tests

Special considerations related to the use of HTS tests are provided in this section. These considerations are proposed for each step of the HTS test and have been divided in two components related to the laboratory and bioinformatic analyses, respectively.

These considerations should be taken into account for every phase: test selection, test development and optimisation, test validation or verification, or routine diagnostics. These considerations can also feed the risk analysis (see section 6.3).

6.2.1 Laboratory component

The laboratory component of HTS tests consists of several steps, and spans from sampling to sequencing. Each step should be selected, developed and optimized and validated for its intended use before it can be used in routine testing. After the validation or verification of an HTS test, its performance should be monitored using appropriate controls during its routine use (see section 6.7 – Ensuring the validity of results).

6.2.1.1 Sampling

Recommendation: **The sampling protocol should be appropriate to the matrix and pest(s) targeted by the HTS test.**

Although the laboratory may not be involved in sampling, it should recommend a procedure for sampling. The type of samples (e.g. different plant parts) and the season of sampling can affect the results of a diagnostic test, including HTS tests (e.g. organisms not detected) (Prezelj *et al.*, 2013; Dr. Martha Malapi-Wight, USDA-APHIS, pers. comms, May 2020). The recommendations on sampling prescribed in section 5.7 of EPPO standard PM 7/98 (2019) are applicable to HTS tests and should be followed. Specific recommendations are available for inspection procedures (EPPO standards PM3 series, https://www.eppo.int/RESOURCES/eppo_standards/pm3_procedures) and for surveillance procedures (https://www.eppo.int/RESOURCES/eppo_standards/pm9_control_systems). Methodologies for sampling of consignments can be found in ISPM 31 (https://www.ippc.int/static/media/files/publication/en/2016/11/ISPM_31_2008_Sampling_of_consignments_EN.pdf).

The laboratory should have a procedure describing the type of material (e.g. expected tissue for plants), the amount of material needed, the number of samples and when relevant, the season of sampling, and the requirements for sampling symptomatic and/or asymptomatic material. The procedure should also define how to deal with samples that do not meet these criteria (Hébrant *et al*., 2018).

Some sampling procedures do not require any supervision from an operator and can be considered automated or semi-automated. This is the case for some insect traps (i.e. pitfall traps and suction traps) and some fungal traps (i.e. spore traps) that are left unsupervised for days or even weeks. In such instances, the need for preservation of DNA throughout the sampling phase should be taken into consideration. Examples of DNA preservatives include different concentrations of ethanol (Marquina *et al.*, 2020) and propylene glycol (Robinson *et al*., 2020).

6.2.1.2 Sample handling

Recommendation: **Sample handling should ensure sample integrity and suitability for the HTS test.**

As with any diagnostic test, the quality of samples can affect the results of HTS tests (e.g. organisms not detected) (Hébrant *et al*., 2018). The recommendations on sample handling prescribed in section 5.8 of EPPO standard PM 7/98 (2019) are applicable to HTS tests and should be followed. Sample handling includes subsampling, traceability of samples, sample preservation between collection and

laboratory receipt (e.g. insects preserved in glycol/ethanol), transportation to the laboratory (e.g. cold-chain box containers, plastic bags to avoid dehydration), samples condition on receipt, storage, aliquoting, retention, and disposal. The laboratory should have a procedure on sample handling (EPPO PM 7/98, 2019).

6.2.1.3 Nucleic acid extraction

Recommendation: **The nucleic acid extraction protocol should deliver nucleic acids of appropriate quality and quantity for the HTS test. Minimal thresholds of quality and quantity should be established and respected.**

The quality (in terms of purity and integrity) and quantity (i.e. ng/µl) of nucleic acids is important as they can affect the results of an HTS test. The extraction allows for the removal of inhibitors that can negatively impact the test result. The selection of the extraction method depends on the type and size of the target(s) expected to be detected by the HTS test (e.g. DNA versus RNA genomes, genomes of a few hundreds of nucleotides to megabases) and the type of matrix from which the nucleic acids are extracted (e.g. plant parts: seed, leaf, stem, purified cultures, soil, water, insects). For example, different types of nucleic acids can be extracted from plant viruses such as small interfering RNAs (siRNA), double-stranded RNA (dsRNA), total DNA and/or RNA, and virion-associated nucleic acids purified from virus-like particles (VANA) (Gaafar and Ziebell, 2020; Maliogka *et al*., 2018; Pecman *et al*., 2017; Visser *et al*., 2016). The composition of the matrix can also affect the extraction of nucleic acids as demonstrated in a study comparing DNA extraction methods of plant-associated bacterial communities from soil and different plant species tested by amplicon sequencing (Giangacomo *et al*., 2020). Several extraction protocols can be evaluated to determine the most suitable method(s) for the intended test use during selection and/or validation and further application in routine testing.

A large range of in-house protocols and commercial kits are available, and their selection will depend on the biological material to be analysed. Specific adaptations of a protocol may be needed for specific organisms/matrices (e.g. nematodes Non-destructive DNA extraction may be preferred for macroorganisms to preserve morphological voucher specimens as it has been applied to freshwater invertebrate samples (Carew *et al*., 2018) and terrestrial arthropods (Nielsen *et al*., 2019). This is particularly important in the case of uncharacterised macroorganisms (e.g. insects) where preservation of a morphological voucher specimen would enable to permanently link the DNA sequence to a morphological sample (Piper *et al*., 2019). Furthermore, DNA from a voucher specimen can be re-extracted when additional genetic information is required, such as a longer DNA sequence (Carew *et al*., 2018).

Based on experience, each laboratory has its own preference in terms of extraction protocol(s). In most cases, a protocol extracting nucleic acids with a purity and integrity satisfactory for PCR or real-time PCR (preceded by reverse-transcription for RNA extracts) should be suitable for HTS tests, particularly if amplicon-based. However, some library preparation protocols require a higher nucleic acid integrity and minimal concentration. This is often the case with long-read HTS technologies. The quality of extracted nucleic acids can be checked by several methods (see section 6.3 – Risk analysis) and minimal thresholds should be defined [for example, minimum, average or maximum fragment length, minimal acceptable purity indicator(s)]. The target organisms should always be kept in mind while selecting a suitable extraction procedure. Target organisms with thick cell walls such as gram-positive bacteria, or with cuticles such as insects and nematodes, might require extra steps during nucleic acid extraction

to lyse the cells (for example sonication or enzymatic protein lysis) (Nielsen *et al*., 2019; Waeyenberge *et al*., 2019; Wesolowska-Andersen *et al*., 2014).

The concentration of target(s) in a sample can be very low in some types of matrices. A low concentration of target(s) in background sequences (e.g. host, non-relevant organisms present in the sample) may result in failure to detect the target(s). The nucleic acid extraction protocol may therefore include a target enrichment or selection step to improve the analytical sensitivity of the HTS test. For example, in water samples, the enrichment of the target(s) is essential (Mehle *et al*., 2018).

The selection of the enrichment protocol depends on the target genome (e.g. ssRNA, dsRNA, total RNA, circular DNA for viruses), its physical properties (e.g. viroid naked RNA, encapsidated viral RNA/DNA, DNA of bacteria and fungi protected by a cell wall), the matrix (e.g. plants, soil, water). For plant samples, viral particle enrichment by ultracentrifugation of tissue sap homogenized in a buffer before nucleic acids extraction, depletion of ribosomal RNA (rRNA) from total RNA or enrichment of dsRNA by cellulose affinity chromatography with or without additional nuclease treatment are three examples of protocols that can improve the sensitivity (Adams and Fox, 2016). Rolling circle amplification is also frequently used as an enrichment procedure when targeting DNA viruses with circular genomes (Johne *et al*., 2009).

6.2.1.4 Library preparation

Recommendation: **The library preparation protocol should be suitable for the sequencing platform. Minimal quality and quantity thresholds should be identified and respected.**

The selection of the protocol for library preparation depends on the HTS process used. It should be determined during the test selection/test development and tested and optimized when relevant (see sections 6.4 – Test selection and 6.5 – Test development and optimisation).

For **shotgun sequencing**, several protocols are available. They depend on the sequencing technology and are often provided as kit(s) with all the reagents included. Their selection will depend on technical criteria (e.g. minimum required quantity and integrity of the extracted nucleic acid and expected proportion of target nucleic acid), the time needed, the required staff, the costs of reagents and consumables. The enrichment of target nucleic acids can also be carried out during library preparation. It can be based on a size selection or on the use of specific oligonucleotides to either eliminate non-target nucleic acids (like ribosomal RNA in plant samples) or to specifically select the target nucleic acids. For example, it has been shown that the removal of plant ribosomal RNA by specific oligonucleotides resulted in a 10-fold enrichment of viral sequences (Adams and Fox, 2016).

For **amplicon sequencing** which relies usually on a PCR step, special care should be taken for the selection of primers to make sure the target organisms can be amplified with these primers, as demonstrated in a study of the fungal microbiome of higher plants by Scibetta *et al*. (2018). A high fidelity polymerase should preferably be used to minimise the amplification biases due to the miss-incorporation of nucleotides, Budowle *et al*., 2014; McInerney *et al*., 2014). The number of PCR cycles should be selected to ensure the PCR is still in the exponential phase; this is usually 25-30 cycles for a quantitative metabarcoding test. PCR amplification in metabarcoding targets a small region of the genome, the barcode, generally corresponding to partial genes. Barcode regions allow the detection and identification of the targeted organisms. Barcodes have been proposed and described in EPPO standard PM7/129 (2016) for arthropods, bacteria, fungi, nematodes, oomycetes, invasive plants,

phytophthoras and phytoplasmas by classical Sanger sequencing. Some of these barcodes have been successfully used in metabarcoding (Ahmed *et al*., 2019; Dormontt *et al*., 2019; 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.

6.2.1.5 Pooling of libraries

Recommendation: **The pooling level of libraries should be adapted to the intended use of the HTS test and its required analytical sensitivity.**

Several libraries can be pooled together to reduce the sequencing costs. During library preparation, nucleic acids extracted from each sample are tagged with an index incorporated in the sequencing adapters so that each obtained sequence can be traced back to the original sample (Budowle *et al*., 2014; Piper *et al*., 2019). The process of pooling increases the risk of misassignment of reads to a sample due to cross contamination of tagging that can occur during library preparation and sequencing (i.e. index-hopping) or between sequencing runs (i.e. inter-run contamination when identical indexes are used in successive runs) (Galan *et al*., 2016; Kircher *et al*., 2011; van der Valk *et al*., 2018). Index misassignments can also occur during the demultiplexing step due to sequencing errors on indexes, which depends on the sequencing technology. The risk is increased when high sequencing depths are obtained with pooled libraries (Budowle *et al.,* 2014; Massart *et al*., 2019). Sample misassignments can be reduced on the Illumina platform by using dual indexes (Kircher *et al*., 2011) and almost abolished by using unique dual indexes which increases the bioinformatic power in catching index-hopping (MacConaill *et al*., 2018). Another option is to use indexes that are sufficiently long and different, so that their identification is robust and tolerates several sequencing errors. Nevertheless, these options can only limit the problem of index switching as they do not take into account other origins like the creation of chimeric sequences due for example to the ligation of free adapters (Wright *et al*., 2016). Pooling libraries just prior to sequencing or adding a step to remove free adapters can also reduce these misassignment issues. The sequences of sets of indexes included in each run should be recorded for trace back purposes and to plan properly the succession of sequencing runs.

Pooling also requires that the amount of nucleic acid of each library in the pool be normalised. This minimises the pooling bias that causes the generation of uneven numbers of sequences between samples (Hébrant *et al*., 2018). The laboratory should be aware of the risk associated with pooling (see section 6.3 – Risk analysis) and demonstrate that the pooling strategy used, does not affect the test performance (e.g. lower analytical sensitivity, contamination). The pooling method depends on the desired read depth of the targets to be sequenced and should be optimized to ensure that the HTS test meets the criteria of its intended use (Hébrant *et al*., 2018).

6.2.1.6 Sequencing platforms and methods

Recommendation: **The sequencing platform and method should be appropriate for the intended use of the HTS test.**

The laboratory should consider the type of sequencing platform and the method of sequencing best suited for the intended use of the HTS test based on the following points:

- expected number of samples received per batch and reads number per sample: these two parameters will determine the number of reads requested on average per batch of HTS testing, knowing that several batches of samples can be sequenced on a single run

- required test turn-around time (e.g. urgent testing for imported perishable material),

- total number of generated reads per sequencing run: it should be compared to the requested reads per batch in order to determine if a batch requires a complete or partial sequencing run, which as an impact on the turn-around time,

- multiplexing capacity of the platform: is it compatible with the expected number of samples per batch?

- read length and type (e.g. single, paired, mate-pair): these two parameters will depend on the HTS test used: short single reads are appropriate for sRNA sequencing whereas amplicon sequencing might need the longest reads, provided the error rate is acceptable,

- error rate and type of error: the error rate varies between the sequencing platforms and between runs. It can be critical for some HTS technologies, such as for amplicon sequencing where a small number of errors in the sequence can modify its annotation, or for shotgun sequencing when SNPs are important,

- availability of bioinformatic support, laboratory resources and technical expertise and manufacturer level of technical support in order to solve quickly (re-)occurring problems.

Another key point to consider during the selection of the sequencing platform is the impact on the downstream bioinformatic analyses (depending on the number of sequences, their length, their quality and accuracy) (Budowle *et al*., 2014; Jennings *et al*., 2017).

A cost study may be carried out taking into account the previous criteria and also (i) the three main expenses involved in the operation of a sequencing machine: purchase, running (reagents and consumables, e.g. the cost per sequence) and maintenance; (ii) the personnel time and expertise needed to run and maintain the machine (Rehm *et al*., 2013). These considerations can be important for a decision to invest in a desktop or stand-alone sequencer or to outsource the sequencing step.

Sequencing platforms are regularly updated and the laboratory should closely monitor these updates and evaluate their potential impact on the test results (see section 5.5 – Management of modifications and section 6.6.3 – Impact of changes made to a validated HTS test).

6.2.1.7 Contamination

Recommendation: **The laboratory should prevent contaminations as they can critically impact the results of HTS tests.**

The issue of contamination is particularly important for HTS tests as they are as, or even more prone to contamination than PCR-based tests. The higher chance of contamination within HTS tests comes from the multiple handling steps and use of more reagents in the sample preparation process. There is also a higher chance to detect a contaminant in HTS tests because of their broad range of detection. Contamination can occur at different steps of the laboratory component (e.g. nucleic acid extraction, library preparation, sequencing). Sources of contamination may include sample handling, laboratory surface and equipment/tools contamination, reagents and carry-over (Asplund *et al*., 2019; Champlot *et al*., 2010; Dickins *et al*., 2014; Gaafar *et al*., 2020; Rosseel *et al*., 2014).

Contamination between successive uses of a sequencing machine, called carry-over contamination has often been observed (Quail *et al*., 2014). In addition, contamination can occur when multiplexing several samples in a single sequencing experiment, i.e. the cross-contamination between prepared nucleic acids due to traces of other samples or index-hopping between samples (see section 6.2.1.5 – Pooling of libraries) (Buschmann *et al*., 2014). It has also been demonstrated that contamination of laboratory reagents used for HTS, such as DNA extraction kits or molecular grade water, can impact the results obtained using shotgun or amplicon sequencing tests (Asplund *et al*., 2019; Galan *et al*., 2016; Salter *et al*., 2014).

The recommendations on the organisation of facilities to prevent contamination are described in section 5.1. In addition, the best practices for molecular laboratories should be applied (e.g. use of clean reagents, consumables, tools, equipment; frequent changes of disposable tools and frequent cleaning of benches, equipment and tools, use of different dedicated rooms, or physical separation of activities). The EPPO standard PM 7/98 (2019) provides guidance on how to avoid contamination with specific requirements for molecular laboratories and specific guidelines for monitoring contamination. In addition, the physical separation of samples suspected to contain a high load of organisms from samples suspected to contain a low load, is highly recommended. The laboratory should regularly decontaminate benches, equipment and tools with appropriate products (e.g. 0.1M sodium hydroxide followed by 70% (v/v) ethanol, Virkon).

Despite every precaution taken, some contaminations can still occur such as some cross contamination because of index hopping with pooled samples or when using standard library preparation for sequencing (i.e. not double indexes or unique molecular barcodes). Therefore, the level of contamination should be monitored during the bioinformatic analyses (see section 6.2.2.3 – Analysis of controls included in the HTS test) based on the relevant first line controls (see section 6.7.2.1 – First line controls) and should be taken in consideration during interpretation of the results. Noteworthy, the identification of contaminations in the sequencing datasets is not yet standardized and many scientific, technical and bioinformatic developments are expected in the near future to improve it.

6.2.2 Bioinformatic component

The bioinformatic pipeline consisting of a combination of software used to analyse the raw data, is a key element of the HTS test as it can generate false positive and/or false negative results. The results generated by the pipeline depend on each (version of) software used, the parameters and the thresholds applied, as well as the accuracy and completeness of sequence database(s) used for sequence comparison (see section 5.6.3 – Reference sequence database). The impact of the bioinformatic pipeline on the correct identification of target(s) has been shown by Massart *et al*. (2019) through a test performance study including 21 plant virology laboratories analysing 10 datasets.

Whatever the bioinformatic strategy, many pipelines have been developed that can operate either on a Linux system, statistic programmes, web interface, as well as commercial packages or OS-friendly open-source software. A current general trend is to simplify the use and the parameterization of these tools, making them usable without extended bioinformatic knowledge or, sometimes, as a "one-click" solution. For such simplified pipelines, it is paramount that the personnel implementing and using them understands what is being done and that the pipelines are appropriately used (according to the data and the goal of the analysis) by competent personnel (see section 5.2 – Personnel).

The bioinformatic component of the HTS process can be divided in three steps with several sub-steps which are described below. It should be noted that the result of each sub-step of the bioinformatic analyses depends on the selected parameters and metrics of the previous sub-step(s). For example, the selected minimal quality score of trimming from the first step of the bioinformatic analyses (see section 6.2.2.1 – Analysis of raw reads) can impact the quality of the reads assembly from the second step of the bioinformatic analyses (see section 6.2.2.2 – Identification of target(s)).

6.2.2.1 Analysis of raw reads

Recommendation: **The laboratory should eliminate low quality sequences and assign unambiguously the sequences to each sample. Optionally, redundant or background sequences can be eliminated.**

The analysis of raw reads consists of different sub-steps (Figure 2) which may be performed in a different order depending on the bioinformatic pipeline. These sub-steps may not all be relevant, depending on the HTS protocol used (e.g. single species vs metagenomics, short vs long reads, single-end vs paired–end).

The sub-steps needed and their order should be defined during test development/adaptation (see section 6.5 – Test development and optimisation) with their parameters and corresponding quality metrics and thresholds (Weiss *et al*., 2013; Hébrant *et al*., 2018). If these thresholds are not met, the decision on repeating or proceeding with the HTS run should be documented.

The first sub-step of the bioinformatic analyses is to check the overall quality of the sequencing dataset by looking at the metadata produced during the sequencing run (e.g. cluster densities, quality profiles, number of and size of reads) and the specification metrics. These metrics are platform-dependent and the most relevant metrics should be determined during validation with the setting of (a) minimal threshold(s) (Hébrant *et al*., 2018). Alternatively, the analysis of the metrics of this sub-step can be carried out after the trimming of primers, adapters, and indexes.

The raw reads can also be controlled for their quality. Quality control is a process that removes either nucleotides or the full sequences of reads whose quality does not meet an established threshold. The quality of reads is checked using the base quality scores (for example, Phred quality score) which can vary depending on the sequencing platform. A minimal threshold of the base quality scores should be defined during validation (see section 6.6.1 – Validation of HTS test). It should be noted that the choice of an optimal threshold for read trimming is always a trade-off between sequence loss and dataset quality (Del Fabbro *et al*., 2013). This score is logarithmically related to the base calling error probability which is used to measure the quality of the identification of each nucleotide by the sequencing platform (Lambert *et al*., 2018). The objective of quality filtering is to retain sequences of appropriate quality for the next steps of the bioinformatic analyses (Budowle *et al*., 2014; Hébrant *et al*., 2018; Weiss *et al.,* 2013).

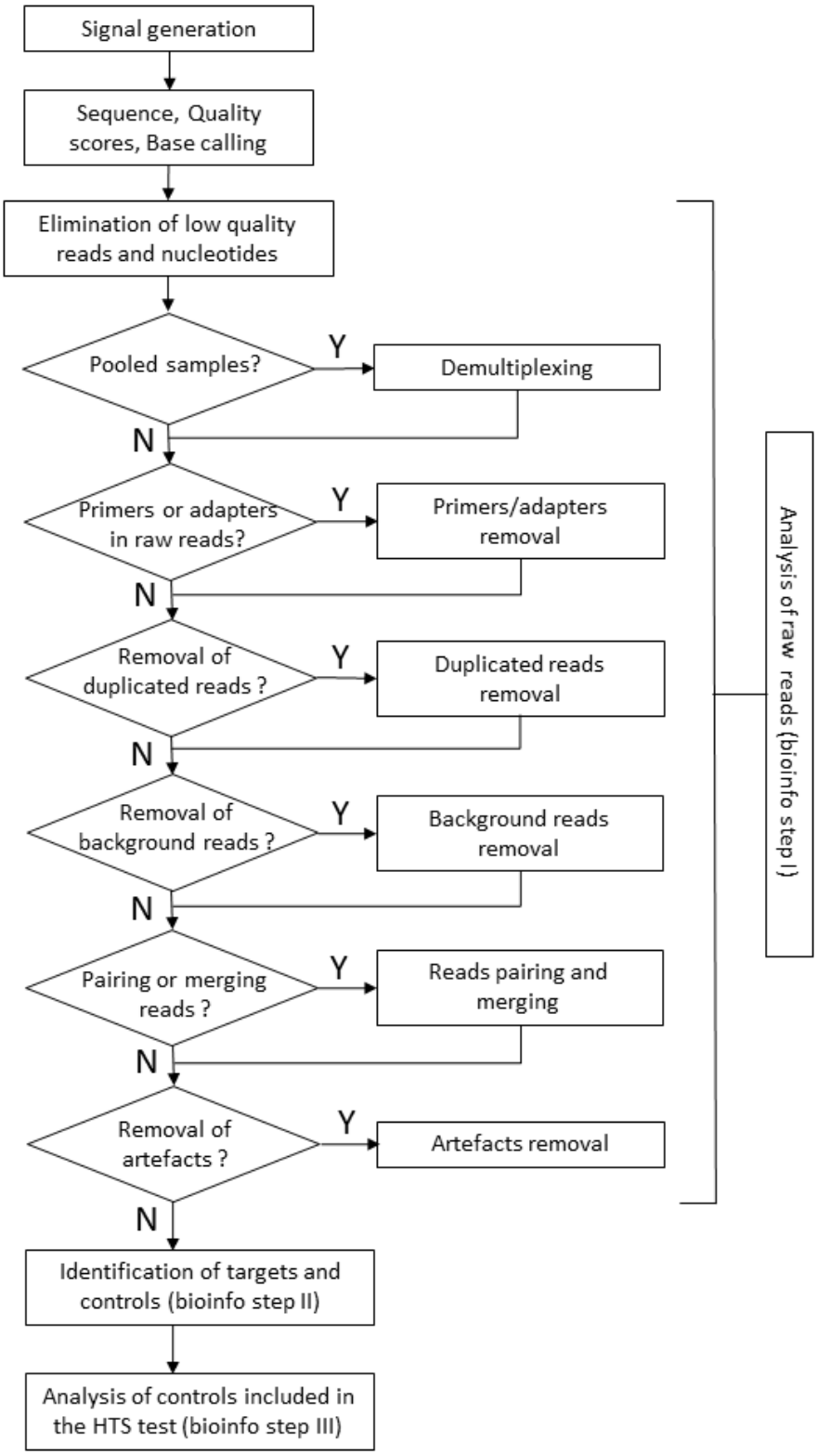

Figure 2. Example of the organisation of the first step of the bioinformatic analyses, i.e. analysis of raw reads. The order of sub-steps can be modified, depending on the bioinformatic workflow used (for example the elimination of low-quality reads and nucleotides can be carried out at any time during the process).

The other sub-steps (when relevant) are:

-*Demultiplexing*: If several libraries were pooled for sequencing, the reads are assigned *in silico* to their respective samples of origin by cross-checking the index sequences associated with each read (Budowle *et al*., 2014; Hébrant *et al*., 2018). For this, it is recommended to use an appropriate stringency, so that the tolerance of index errors cannot cause misassignation of the reads. It is also possible to search for index sequences that have not been used in the sequencing run to estimate and filter possible cross-contamination that may have occurred during the indexing or sequencing steps (i.e. inter-run contamination) (Galan *et al*., 2016; Kircher *et al*. 2011; van der Valk *et al*., 2018). Misassignment can occur during this step due to errors in index sequencing and inappropriate bioinformatic parameters (for example mismatch tolerance). The risk of misassignment also depends on the type of indexes (e.g. long and very different indexes, single vs dual indexes, see section 6.2.1.5 – Pooling of libraries).

- *Primer, adapter and indexes removal* (also known as clipping, trimming): primers, adapters and indexes (if used) included in the generated reads should be removed before continuing the bioinformatic analyses (Davis *et al*., 2013; Hébrant *et al*., 2018). The removal is usually done during the demultiplexing step (see above).

-*Background reads removal*: Some sequences not related to the target(s), called background reads (e.g. host sequences, ribosomal sequences, phage sequences, environmental contaminant sequences), can be removed to facilitate the search of target(s) sequences and to reduce the risk of reporting incorrect results (Lambert *et al*., 2018). These reads, are mainly associated with shotgun sequencing strategies, and their presence also depends on the nucleic acid extraction procedure used (e.g. total nucleic acid extraction vs. target enrichment or selection). They can be removed by reference subtraction (i.e. host genome reads or host rRNA reads removal). The host control and/or no template control can be useful in finding the background reads (e.g. host reads, environmental contaminant reads; see Table 5). The removal of background reads can be particularly important when the target(s) are present in low concentration (Baizan-Edge *et al*., 2019). Caution should also be taken when dealing with organisms that are capable of being completely or partially integrated in their host genome because they may be removed during this process (e.g. pararetroviruses in plants, bacteriophages in bacteria; Hohn *et al*., 2008, Sharma *et al*., 2017). A large-scale performance evaluation of sequence analysis strategies revealed that some laboratories were unable to identify the viral sequences integrated in the host genome present in the sequence dataset (Massart *et al.,* 2019). It should be noted that there may be a risk of removing the target reads during this process when high sequence identities exist between the host and target or if the reference genomes used for the removal of background reads contain themselves contaminants target reads. The awareness of the quality of the host reference genome used for background reads removal is hence very important. Some (typically lower quality) reference genomes in databases can contain contaminant sequences (from endophytes), or are incomplete and their annotations are still in progress. Very different results can be obtained with or without the removal of background reads and therefore this should be considered during test development/optimisation with parameters and thresholds settings based on the HTS test intended use (e.g. detection to species or genus level) (see section 6.5 – Test development and optimisation).

-*Duplicate reads removal*: Duplicate reads originate from the same amplified fragments. Their characteristics are common coordinates (e.g. the same start and end coordinates after mapping), same sequencing direction (or mapped strand) and identical sequences. The presence of duplicate reads depends on the Initial sequence complexity of extracted nucleic acids, the library preparation procedure and the sequencing technology. They can be generated during a fragmentation or tagmentation step or by an amplification-based technology (Hébrant *et al*., 2018; Maliogka *et al*.,

2018). A dataset containing lots of duplicated reads might also be the result of a failed library preparation, where too little input material was available. The high abundance of duplicated reads can limit the sensitivity of the HTS test as they can compete with low abundance targets, despite having a large total read number. It is therefore recommended to evaluate the proportion of duplicate reads during the quality control stage of data analyses. Excess duplicate reads can be removed by using read normalisation tools in order to facilitate the downstream analysis of sequences. The elimination of duplicated reads depends on the protocol and is not required in protocols that use the number of reads (sometimes identical) to estimate the relative abundance of a target like amplicon sequencing for metabarcoding.

-*Merging paired-end reads*: In paired-end sequencing, the DNA fragment is sequenced from both ends (sense and antisense sequencing). Depending on the intended use of the HTS test, it may be useful to merge both reads of a single DNA fragment, if they overlap. For some sequencing technology, like Illumina, the quality of the sequence tends to diminish towards the end of the reads (Kwon *et al*., 2014; Lambert *et al*., 2018). The pairing of reads can increase the overall quality and the length of the sequences. The parameters should be defined when the two sequences to generate the consensus sequence, are not identical.

- *Artefact removal*: Amplicon sequencing can generate chimeric sequences corresponding to a combination of different sequences from the original sample, leading to the formation of false sequences. The first part of the sequence is coming from a target organism while a second part is coming from another target organism as a result of an amplicon that accidentally acts as a primer during PCR. Similarly, whole genome amplification techniques such as multiple displacement amplification (MDA) or emerging single cell sequencing techniques are commonly used within low-input library preparation protocols for shotgun sequencing can produce chimeric sequences (Lasken and Stockwell, 2007; Quince *et al*., 2011). It is important to monitor and remove these sequences using appropriate tools before the target identification (Anslan *et al*., 2018; Lu *et al*., 2019; Quince *et al*., 2011).

- *Denoising / clustering (specific to metabarcoding)*: PCR and sequencing errors inherent to amplicon sequencing introduce noise through the generation of high numbers of unique amplicons differing from the original sequences by one or more nucleotides. As a consequence, spurious results can be generated and data analysis can become more complex. Within metabarcoding analyses sequencing reads are commonly clustered in representative bins called Operational Taxonomic Units (OTUs) using a nucleotide similarity threshold that ideally broadly approximates species boundaries (Mahé *et al*., 2015). Nevertheless, the optimal selection of threshold can vary across taxa and can result in over-clustering (putting different species together in one cluster) or under-clustering (splitting one species over different clusters) (Anslan *et al*., 2018; Quince *et al*., 2011).

Alternatively, denoising algorithms have been developed. They do not cluster the sequences based on their similarity but resolve erroneous sequences by assuming that erroneous sequences will be closely related and will show a similar ocurrence pattern than an authentic 'parent' haplotype while showing lower abundances and/or lower quality scores (Laehnemann *et al*., 2016; Yang *et al*., 2012). After read correction, this denoising process produces amplicon sequence variants (ASVs) or exact sequence variants (ESVs) that are taxonomically identified.

6.2.2.2 Identification of target(s)

Recommendation: **The laboratory should identify target(s) from the cleaned reads of the analysed samples and controls.**

The second step of the bioinformatic analyses aims to identify the target(s) in the datasets derived from samples and controls.

An accurate identification of the target(s) bioinformatically is important to avoid false positive (wrong taxonomic position, gene annotation or variant detection) or false negative (absence of identification) results.

The identification of target(s) in samples and controls consists of different sub-steps which may not all be relevant (Figure 3), depending on the HTS protocol and the sequencing platform used. These sub-steps may be performed by the laboratory or the sequencing providers in a different order depending on the workflow. The sub-steps that need to be performed and their order of processing should be defined during test development/adaptation (see section 6.5 - Test development and optimisation) along with their parameters and their corresponding quality metrics and thresholds (Budowle *et al*., 2014; Hébrant *et al*., 2018). If thresholds are not met, the decision of repeating or proceeding with the HTS run should be documented. In addition, if updating the algorithms or databases, the sub-steps and their order can be adapted in routine test once the appropriate verification has been carried out.

The optional sub-steps of the second step of the bioinformatic analyses are:

- *Direct annotation of individual reads*: The quality checked reads can be annotated at taxonomic or functional levels without any assembly, clustering or mapping. The specificity of the annotation process will depend on the length of the sequences and on the database(s) used (see taxonomic position and functional assignment sub-steps).

- *De novo assembly* (also called contiguous assembly, reads assembly): The quality checked reads from a shotgun sequencing library can be assembled *de novo* to create longer sequences, called contiguous sequences (or contigs) (Brinkmann *et al*., 2019). The reads are assembled when they present similar sequences on a portion or on the totality of their length. The reads assembly can be complex when they are short (like for small RNA sequencing) (Massart *et al*., 2019). canThe parameters for reads assembly depend on the type of algorithm used and should be defined during test development/optimisation, such as, the percentage of identity between reads, the minimum overlap, the minimal length of contigs, the k-mer length or bubble size. The quality of assembly in contigs can be evaluated, for example using N_{50} or U_{50} values (Castro and Ng, 2017) or for *de novo* assembly of complete genome, using CheckM or BUSCO (Parks *et al*., 2014; Seppey *et al*., 2019). Once the reads have been assembled into contigs, these can be annotated taxonomically and/or functionally (see below). In case there remains unused high-quality reads at the end of the *de novo* assembly process, these can be further analysed and some guidance is provided below.

-*Reference mapping* (also called reference assembly) for selected target(s): If (a) reference sequence(s) are(is) known for organism (e.g. host, pest) suspected to be present in the sample, the quality checked reads can be directly mapped against the targets reference sequence(s), which can be partial or full genome(s) (Budowle *et al*., 2014; Hébrant *et al*., 2018; Roy *et al*., 2018). Several reference sequences can be used for each target in order to take into account genetic variability (Massart *et al*., 2019) and improve the number of mapped reads and the annotation quality.

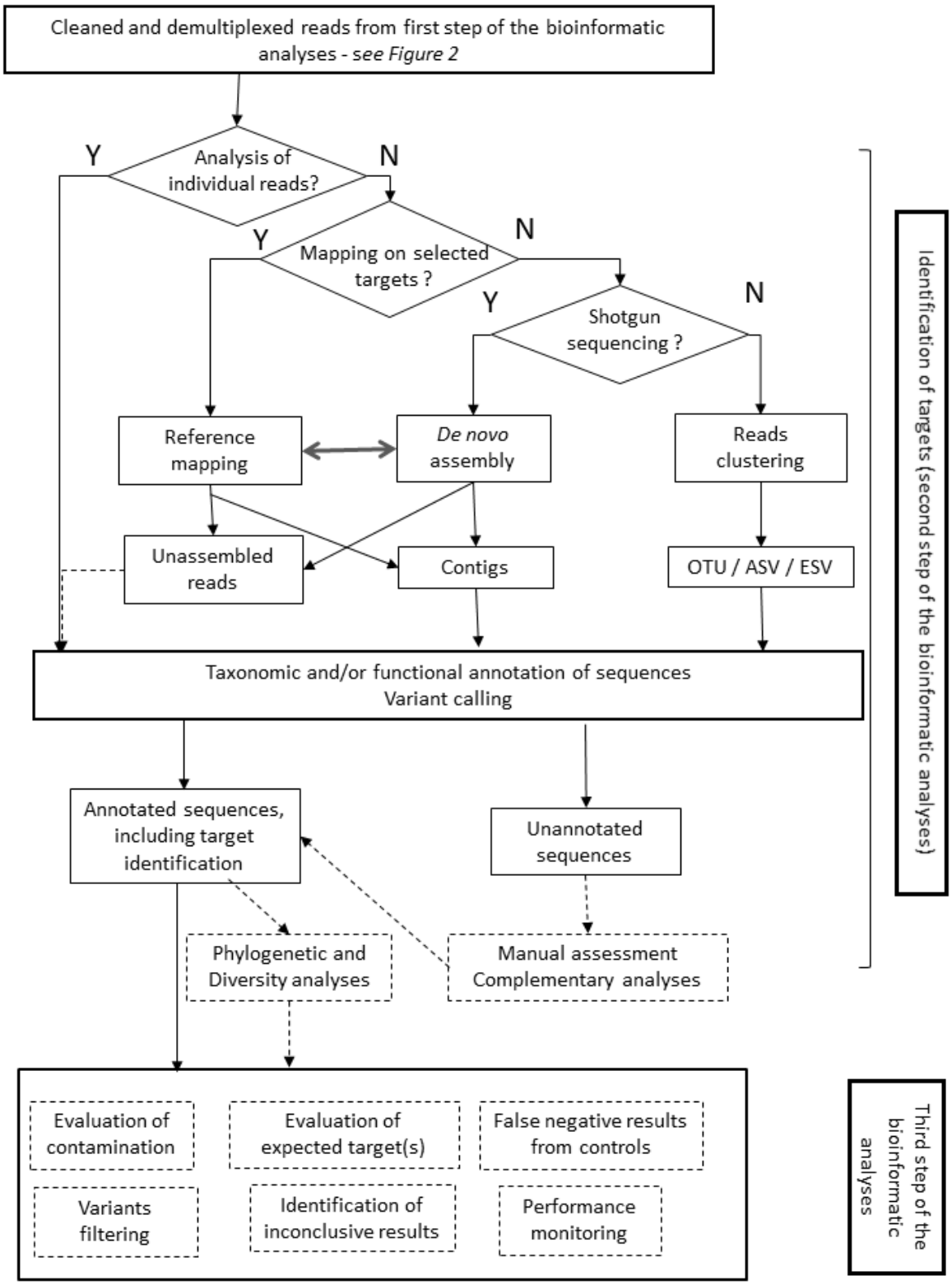

Figure 3. Overview of the second and third steps of the bioinformatic analyses, identification of target(s) and analysis of controls included in the HTS test, respectively. The selection of the sub-steps depends on the bioinformatic workflow used. Discontinued arrows are alternative steps. Meaning of acronyms: ASV: amplicon sequence variants, ESV: exact sequence variants, OTU: Operational taxonomic units.

The mapping parameters such as, number of mismatches or gaps allowed or minimal percentage of identity, are critical to avoid incorrect results. If the mapping parameters are set too low, non-specific mapping to another species can happen, while too stringent mapping parameters can result in the poor mapping of reads from a distant isolate (Roy *et al*., 2018; Weiss *et al*., 2013). Important mapping results metrics include genome completeness, average read depth, distribution of reads on the reference sequence and percentage of identity with reference sequence(s). Their individual relevance depends on the technology used (e.g. PCR amplified targets will result in greater read depth) (Asplund *et al*., 2019; Weiss *et al*., 2013).

A combination of reference mapping and *de novo* assembly can be required to increase the likelihood of identifying target(s) present in low concentration (Maliogka *et al*., 2018). The ordering of contigs along a genome (i.e. scaffolding) can improve downstream analyses like the taxonomic and functional annotation (Sahlin *et al*., 2016) or *de novo* assembled (meta)genome contiguity.

- *Taxonomic position for pest identification*: when using reference mapping, the taxonomic position can be obtained from the annotation of the reference sequences but there can be a risk of misassignment (reads belonging to another species are mapped on the reference used) and the contigs generated from reads assembly might need to be further annotated independently. For individual reads, clustered reads and *de novo* contigs, the taxonomic position should be determined using the latest taxonomy information, including up to date sequence-based demarcation criteria (see box below for a list of current databases) and appropriate sequence databases and software (see sections 5.6.3 – Reference sequence database and 5.5 - Managing modifications). Similarity searches performed from assembled contigs or reads using dedicated tools (e.g. ANI, AODP, BLAST, DIAMOND, EDNA, Mash, Kraken, KAIJU) provide indications on the taxonomic position and the closest organisms, most often with a confidence threshold (Lambert *et al.,* 2018; Maliogka *et al*., 2018; Massart *et al*., 2019). These similarity searches use algorithms analysing alignment, K-mer, signature short motifs, and are continuously evolving (Budowle *et al*., 2014; Lefebvre *et al*., 2019; Rott *et al*., 2017; Ye *et al*., 2019). Simply taking the top hit in a BLAST search can lead to wrong conclusions. In addition to sequence similarity searches, some taxonomic classifiers, like RDP classifier, QIIME or SYNTAX, also take into account other similar sequences in the reference database and provide a confidence score using approaches such as bootstrapping. The level of certainty of the similarity searches should always be retained and mentioned (e.g. e-value) together with the tool and database (version) used. Expert judgement may be needed to evaluate the result of a taxonomic position (Massart *et al*., 2017; Matthijs *et al*., 2016). This is particularly challenging when dealing with uncharacterized organisms or with a sequence identity close to the threshold of species demarcation. When it is possible to retrieve the whole genome of a target, through shotgun sequencing, genome completeness and read depth can support the result of a taxonomic annotation (i.e. the more complete the genome is, the more reliable the taxonomic position). Additional analyses such as phylogenetic analysis may also be required. For amplicon sequencing, the resolution of the taxonomic assignment of the OTUs depends on different factors with the chosen barcode, the completeness of the reference database and the taxonomic position algorithm as the main ones. Currently used barcodes are relatively short (a few hundred nucleotides), and hence can provide only a limited taxonomic resolution. Classification methods such as naive Bayesian classifiers, lowest common ancestor-based methods, or phylogenetic placement methods are more reliable, but often also more conservative, hence not always leading to a satisfactory species-level classification. These limitations are inherent to amplicon sequencing or to the annotation of individual reads from shotgun sequencing and should be considered and explored *in silico* during the test selection and development, to verify whether the barcode is suited to detect the target organism(s) at a satisfactory taxonomic level.

Currently (in 2020), the IPPC diagnostic protocols usually consider the following species lists for the latest taxonomy information:

International Committee on Taxonomy of Viruses (ICTV), https://talk.ictvonline.org/

International Committee on Systematics of Prokaryotes (ISCP), http://www.the-icsp.org/

International Commission on the Taxonomy of Fungi (ICTF), https://www.fungaltaxonomy.org/

Committee on Taxonomy of Plant Pathogenic Bacteria -International Society for Plant Pathology, https://isppweb.org/about_tppb.asp

Remark: some recently discovered species might not be listed in the official taxonomy list while being described in the literature and published in genome databases.

-*Functional assignment*: The determination of the (potential) function of genes, the (prediction of) genomic features related to pathogenicity, resistance to antibiotics or to pesticides, proof of irradiation of live insects (provoking nucleotide mutations) intercepted at a border or any other sequence feature that may be of importance to plant health (Davis *et al*., 2016; Leifert *et al*., 2013; Zheng *et al*., 2015) may be useful/required depending on the intended use of the HTS test.

- *Recovering the whole genome of pests*: Obtaining the (near) complete genome sequence may be required for example, to validate the taxa identified, to gain information on the gene content and population diversity, or to resolve properly the epidemiology and origin of an outbreak. Obtaining the (near) complete genome sequence for viruses is relatively easy because of their small genome sizes. The ability to recover a (near) complete genome becomes more complex with bacteria, phytoplasmas, and eukaryotic pests. When a (near) complete genome is needed, an iterative combination of reference mapping and *de novo* assembly with varying parameters can be carried out. Alternatively, a combination of sequencing strategies such as short and long read sequencing can assist in obtaining the (near) complete genome.

-*Variant calling*: Variants can consist in single nucleotide polymorphism (SNP), insertion and deletion of nucleotides (indels) or by the integration/deletion of entire genes compared to a reference sequence or compared to the consensus contigs generated (for example, the quasispecies complex of haplotypes for a virus isolate). The number of variants identified on a pest genome compared to a reference sequence can be used to evaluate if a new species (for viruses) or divergent isolate (for bacteria) has been identified and if the used reference is appropriate. To identify those variants accurately, longer reads can be used and if possible, retrieved from several samples and the associated metadata such as mapping quality, base-calling quality and strand bias should be checked (Gargis *et al*., 2015; Roy *et al*., 2018; Weiss *et al*., 2013). Replicates from the same sample can be processed in parallel to verify that the variant is identified in all datasets.

-*Unused high-quality reads:* A number of reads that have passed all the quality checks may still not be assembled, mapped or annotated after the bioinformatic analyses. These reads, also called unused reads or unmapped reads, can be gathered as a separate output during analysis and their number or proportion calculated. Depending on the purpose of the test and the algorithms used, these reads can be discarded or re-analysed using other algorithms in order to validate the absence of target sequences or of unforeseen organisms among them. Some individual sequences or some contigs, may still not be

annotated during the second step of bioinformatic analyses. These unannotated sequences are sometimes referred to as “dark matter”. Periodic re-analysis can be carried out to see if progress in strategies, algorithms or in knowledge of organisms allows a progress in annotation of such “dark matter”.

6.2.2.3 Analysis of controls included in the HTS test

Recommendation: **The laboratory should verify that all the controls used in the HTS test performed as expected.**

The third and last step of the bioinformatic analyses is to verify that all the controls included in the HTS run performed as expected. This step is important to identify potential false positive and/or negative results.

False positive results may come from contaminations which can occur at different sub-steps of the laboratory phase (e.g. nucleic acid extraction, library preparation, sequencing; see section 6.2.1.7 – Contamination). They may be due to the handling of samples during the process, inappropriate quality of nucleic acid extracts, libraries and/or pooling of libraries (see section 6.2.1.3 – Nucleic acid extraction, section 6.2.1.4 – Library preparation, section 6.2.1.4 – Pooling of libraries).

False negative results can come from several origins, including for example the inhibition of enzymatic reactions, the sample degradation or the generation or too few sequences.

To address false positive and/or negative results, different controls can be included at different stages of the HTS test. The type of controls that can be included is provided in Table 5 (see section 6.7.2.1 – First line controls). All controls should be checked and should meet their respective acceptance criteria (see Table 1). The origin of false positive and/or negative results should be investigated and addressed and the decision on whether to repeat (parts of)the HTS test should be documented.

The analysis of controls may consist of different sub-steps that may not all be relevant, depending on the HTS protocol and the controls used (see section 6.7 – Ensuring the validity of results). The sub-steps that need to be performed and their order should be defined during test development/optimisation (see section 6.5 - Test development and optimisation) along with their corresponding quality metrics and thresholds, when relevant (Budowle *et al*., 2014; Hébrant *et al*., 2018). If thresholds are not met, the decision on repeating (parts of) the HTS test should be documented and the reason for the failure of the control(s) should be investigated. These sub-steps are:

- *Evaluation of contamination*: Although the contamination rate has been decreasing with the improvement of laboratory protocols and the sequencing platforms (see section 6.2.1.7 contamination), there is still a need to monitor it qualitatively and quantitatively. To check for contamination that may have occurred during HTS test, positive, negative and alien controls (see section 6.7.2.1 – First line controls) can be used at different stages of the test (Table 5).

-*Evaluation of expected target(s)*: The evaluation of expected target(s) can be carried out using positive and alien controls (see section 6.7.2.1 – First line controls). These targets should all be detected according to the specified metrics (for example: genome completeness, number of generated sequences/reads, read depth and percentage of identity with their reference sequences).

-*False negative results from controls*: False negative results can be expected when one of the targets from the positive control(s) (see section 6.7.2.1 – First line controls) is not detected in the sequence

data. The result metrics for reference mapping (see section 6.2.2.1 - Analysis of raw reads) such as genome completeness, read depth and percentage of identity with reference sequences are important for filtering false negative results (Asplund *et al*., 2019; Weiss *et al*., 2013).

- *Variant filtering*: If of interest, variants generated due to sequencing errors during the HTS test should be flagged or filtered from the original sequence file (e.g. mapping quality, base-calling quality, strand bias) (Hébrant *et al*., 2018; Roy *et al*., 2018), empirical error rate definition or sequencing of parallel technical replicates. However, variant calling should always take into account that sequencing errors, polymerase errors or reverse transcriptase errors can also generate variant artefacts.

- *Inconclusive results*: If there are some issues with the controls of a sequencing run, for example when a quality metric is just above or below the defined threshold (i.e. inconclusive result or grey zone), the origin of the issue should be investigated and addressed (e.g. a reference sequence data set can be used to check that the bioinformatic pipeline performs as expected). The HTS test may need to be repeated or confirmatory tests other than HTS may be required to ascertain the HTS results. Whatever the decision, it should be documented as part of quality assurance (EPPO PM 7/77, 2019).

-*Performance monitoring*: The performance of HTS tests may be checked routinely by including appropriate controls (see Table 5). For example, for HTS tests used for the detection of quarantine pests, a positive control (see Table 5) close to the limit of detection should be included in each sequencing run and the results monitored over time (see Table 5).

6.3 Risk analysis

Recommendation: **The laboratory should conduct a risk analysis for each HTS protocol and update it regularly.**

The risks associated with running HTS tests should be identified before their use as diagnostics tests. Some risks linked to the management of the laboratory (e.g. equipment, personnel, consumables, environment, organisation of the laboratory) or the management of the documentation (e.g. quality management system with procedures, records, traceability of measures) are identical to the risks associated with other molecular tests. However, some additional risks and their associated metrics are specific to HTS tests.

As part of the risk analysis, the laboratory should define clearly the intended use of an HTS test as potential applications can be broad. Indeed, an HTS test can be used as a standalone test or as part of a series of tests for the detection and identification of specific pests (e.g. detection of quarantine pests at import or export, or as part of a phytosanitary certification programme) or for the broad detection of groups of organisms (e.g. some viruses, bacteria) with the likely detection and identification of uncharacterized organisms (e.g. surveillance studies). A range of questions and factors to consider when defining the intended use of an HTS test are provided in section 5.4.3 of EPPO standard PM 7/98 (2019). It is also recommended to define the desired taxonomic level for the HTS test: strains/isolates, pathovar, race, *formae speciales*, species, group of species, genus, family or higher levels of taxa (see section 6.6.1.2 – Special consideration for analytical specificity).

The risks can be analysed through the methodologies described in EPPO standard PM 7/98 (2019). An example, an Ishikawa diagram, adapted from Mehle *et al*. (2014), is proposed in Figure 4. Alternatively, an operational risk assessment framework with tools for the assessment and management of risks for a plant health laboratory has been proposed by Murugan and Kumarasinghe (2018). The risk analysis should be conducted by competent personnel (see section 5.2 – Personnel).

The risk analysis should be regularly updated (e.g. change in the level of and/or type of risks) depending on the results of the quality checks obtained during the development, validation, or verification phases and during the routine use of the HTS test. The risk analysis should be documented as it will be useful in the decision making for verification or validation after modification of an HTS test (see section 6.6.3 - Impact of changes made to a validated HTS test) as well as when troubleshooting errors that may arise (Hébrant *et al*., 2018; Jennings *et al*., 2017).

A non-exhaustive list of risks associated with HTS and their corresponding metrics/controls to monitor/evaluate is presented in Table 1. The risks listed in the table are not relevant to all possible HTS tests. This indicative table aims to guide laboratories in their risk analysis and should be adapted and complemented by the laboratories to fit the intended use of their HTS test. Specific considerations for the risk analysis for the test selection/test development, for the validation/verification of the test and for ensuring the validity of results during routine diagnostics are discussed in the sections below (sections 6.3.1, 6.3.2 and 6.3.3).

6.3.1 Risk analysis for HTS test selection/test development

For the selection of an HTS test (see section 6.4 – Test selection), the risk analysis enables the identification of the type of protocol required for the intended use of the test, taking into account the constraints of the laboratory (e.g. IT infrastructure, equipment, personnel competence/availability), the time constraints (e.g. expected timeline from sampling to result and acceptable delay of the results) and the consulted literature. Specific points related to the sequencing platforms and methods listed in section 6.2.1.6 should be taken into account (e.g. read length, number of generated reads, platform flexibility and scalability, platform error rate and the type of errors produced). Some points listed as input for the risk analysis for the validation/verification of a test in the EPPO standard PM 7/98 (2019) can also be useful.

When a development step is required (see section 6.5 – Test development and optimisation), the risk analysis performed during the selection of an HTS test can be iteratively updated depending on the sequencing data generated and analysed during this phase. These adaptations improve the reliability of the risk analysis for the other phases, e.g. validation or verification of the HTS test (see section 6.3.2 – Risk analysis for validation/verification of HTS test).

6.3.2 Risk analysis for validation/verification of HTS test

When a standardised protocol is available, either written after the development step or obtained from a publicly available standard, a risk analysis should be carried out before validation and/or verification of the test, to identify which performance criteria need to be evaluated and the extent of the evaluation needed, as stated in the EPPO standard PM 7/98 (2019). This standard lists the points that should be considered as input for the risk analysis such as the intended use of the test, constraints of the laboratory, review of validation data available, and review of altered conditions. Figure 4 and Table 1 provide a non-exhaustive guide for risk analysis of HTS tests.

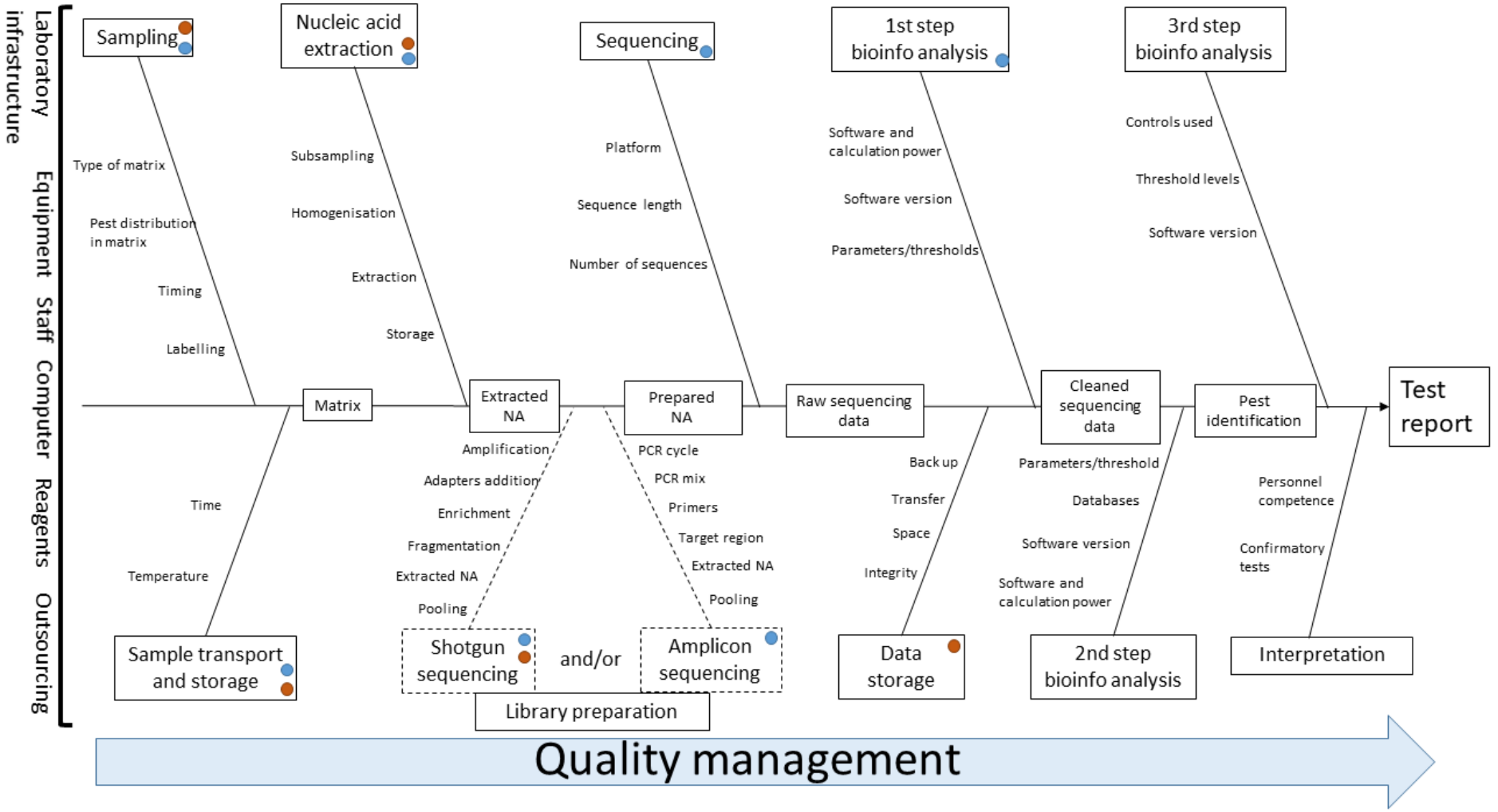

Figure 4. Ishikawa diagram representing the cause and effect of each component of HTS tests. Acronyms: NA: nucleic acids, bioinfo: bioinformatics, PCR: polymerase chain reaction. ● Risk of degradation, ● risk of contamination.

Table 1. Potential sources of errors[1] that may negatively affect the performance of HTS tests with their causes, consequences and mitigation. For details on controls, see section 6.7.2.1 – First line controls.

HTS step	Risk	Cause	Consequence	Mitigation[2]
Sampling	Sampled material not fit for purpose	Unclear or absence of procedure for sampling.	False negative	Protocol describing the sampling procedure with sample acceptance criteria (e.g. type of matrix, material received in sufficient quantity, quality of material suitable for the HTS test).
	Cross-contamination between samples	Inappropriate hygiene procedure.	False positive	Protocol describing the measures to minimise cross-contamination. Use of appropriate high concentration alien controls and/or low concentration positive controls. Use appropriate threshold values to discard contamination sequences.
Sample handling	Inappropriate transport and storage conditions leading to sample degradation	Absent or unclear procedure for transport and storage.	False negative	Protocol describing transport and storage conditions, including the acceptance criteria.
	Cross-contamination between samples	See cross-contamination between samples during sampling.		
Nucleic acid extraction	Inappropriate integrity and purity of nucleic acids	Inhibitors in the matrix, reagent lot number or inappropriate handling and/or storage conditions of samples or reagents will affect the nucleic acid extraction.	False negative	Protocol describing the control of the integrity and purity of nucleic acids with appropriate thresholds. Use of appropriate internal positive controls.
	Low quantity of nucleic acids	Low quantity of input biological material or inappropriate nucleic acid extraction procedure	False negative	Protocol describing the minimal quantity of biological material required and the control of the quantity of

HTS step	Risk	Cause	Consequence	Mitigation[2]
				nucleic acids with appropriate thresholds. Use of appropriate internal positive controls.
	Bias in presence and relative abundance of targets	Protocol favouring specific targets (e.g. VANA or dsRNA protocol under representing some viral taxa, spore forming bacteria or fungal spores difficult to extract with certain extraction protocols and that may be missed or under represented).	False negative	Proper definition of the intended use of the HTS test and selection of the most appropriate extraction protocol. Use of appropriate positive control with relevant targets at known concentration.
	Cross-contamination between samples	See cross-contamination between samples during sampling.		
Library preparation for shotgun sequencing	Content bias (fragment size and concentration of nucleic acids) related to the protocol used	Library protocol favouring some targets (e.g. siRNA/VANA/dsRNA sequencing favouring different viruses, PCR amplification favouring amplicons of smaller size, primers not inclusive enough).	False negative	Proper definition of the intended use of the HTS test and selection of the most appropriate protocol with appropriate threshold.
PCR for amplicon sequencing and some shotgun sequencing	Amplification error	Fidelity of the polymerase can create errors in the generated strand, resulting in false mutation that can bias the sequence annotation.	False positive	Use high fidelity polymerase. Check target sequences from positive controls.
	Bias during amplification	Modification of proportion between targets. Creation of large proportion of duplicated reads.	False negative Bias in relative proportion for metabarcoding	Appropriate primer selection. Set a maximum number of PCR cycles. Use of appropriate positive controls with target(s) at low concentration. Monitor the reads duplication rate (under threshold set)

HTS step	Risk	Cause	Consequence	Mitigation[2]
	Cross-contamination between samples	See cross-contamination during sampling.		
Sample indexing and pooling	Unequal representation of samples in the pool with samples sequenced at very low depth.	Unreliable quantification of prepared nucleic acids between samples or favoured sequencing of target(s) from some samples.	False negative	Check the quantity of prepared nucleic acids using appropriate protocol. Monitor sequencing depth per sample after demultiplexing and set a minimal sequencing depth.
	Index-hopping (e.g. a small proportion of sequence is not tagged with the appropriate index)	Presence of free index in the reaction getting tagged randomly to sample.	False positive	Use highly distinct and unique index between two sequencing runs. Dual labelling index should be preferred for Illumina sequencing technology. Monitor index switching and cross-contamination between library preparation and between sequencing runs. Check the number or % of sequences unexpected in controls.
	Unequal size of the prepared nucleic acids between samples	The size of the prepared nucleic acids may influence the sequencing yield of the sample or the amplicon. For example, low sequencing yields are obtained with long fragments (>500-600 nucleotides) with Illumina technology.	False negative	Define acceptable range of sequence length in prepared libraries. Pool PCR products of similar length or shear samples at appropriate size.
Sequencing	Not enough sequences	Presence of inhibitors in one of the samples Sequencing reagents Under- or overloading the sequencing device with DNA.	False negative	Set minimal sequencing depths per sample.

HTS step	Risk	Cause	Consequence	Mitigation[2]
	Low quality of nucleotide bases	Sequencing reagents, accuracy of signal measurement, background signal. Low quality and incorrect sequences will be generated and can be eliminated during reads analysis or introduce errors during bioinformatic analyses.	False negative False positive	Set minimal sequencing depth per sample. Monitor the number of reads eliminated or remaining per sample after reads analysis.
	Sequencing errors	Type of sequencing platform and sequencing method (e.g. sequencing reagents).	False negative False positive	Use of appropriate controls to evaluate the sequencing fidelity (e.g. positive control, alien control, PhiX control with Illumina technology). Use appropriate threshold. Sequencing the DNA library in both direction to have overlap between reads.
	Inter-run contamination	Traces of libraries from a previous run are sequenced.	False positive	Proper washing of the sequencing machine and alternate the indexes used between runs. Detecting indexes used during previous run (when alternating indexes between runs)
Data storage	Not enough data storage capacity Corrupted data Loss of data Disclosure of data	IT computational system not adapted for handling and storing HTS data.	Corrupted and/or loss of data preventing their (re)analysis	Procedure detailing the required storage capacity including back up, the safe storage of HTS data with disclosure clause and the retention period.
First step of the bioinformatic analyses	GC bias	Target sequences with high GC content are currently not well sequenced (e.g. some bacterial taxa, operon). The sequence coverage of these regions can be very low or	False negative	Determine the extent of GC bias for HTS test targeting organisms with high GC content. Follow GC content statistics of the sequencing.

HTS step	Risk	Cause	Consequence	Mitigation[2]
		absent compared to the other regions.		
	Inappropriate demultiplexing	Inappropriate demultiplexing software and parameters might fail to assign a small proportion of the sequences to the proper sample. inappropriate combination of indexed samples within a run.	False positive	Verify the compatibility of demultiplexing parameters and software with the used indexes. Use highly distinct and unique index between two sequencing runs.
Second step of the bioinformatic analyses	Low quality of reads assembly	The use of too stringent parameters during the assembly of reads can hamper reads assembly while a higher tolerance to mismatches can create assembly artefacts.	False negative False positive	Set appropriate thresholds for the parameters of reads assembly quality. Monitor the parameters of reads assembly quality in each run using appropriate controls.
	Low mapping quality	The reference database used (and its completeness and curation level) and the parameter of mapping can create bias: incorrect mapping of reads on a reference from another species or absence of mapping due to the sequencing of a distant isolate of the pest.	False negative False positive	Set appropriate thresholds for the parameters of mapping quality, such as percentage of identity with reference sequence, quality of assembled genomes. Monitor the parameters of mapping quality in each run using appropriate controls Check the suitability of the reference sequence and/or sequence database.
Second step of the bioinformatic analyses (ctd)	Non-uniform coverage of the pest genome	Variation in read depth or absence of coverage can be due to low abundance of the pest, the use of distant reference genome, the difficulty for sequencing some regions of the genome.	False negative	Set the minimum criteria for the level of coverage across the sequenced regions (which may be partial or full genome). Monitor the uniformity of coverage in each run using appropriate controls.
	Wrong taxonomic and functional annotation	Database completeness and accuracy are essential for proper annotation of the sequences. Software	False negative False positive	Set appropriate threshold parameters to ensure the reliable identification of

HTS step	Risk	Cause	Consequence	Mitigation[2]
		methodology and threshold parameters could also influence the annotation. The particular case is outside the knowledge/know how of the diagnostician/ bioinformatician.		targets according to species demarcation criteria. Monitor the taxonomic and/or functional annotation using appropriate controls. Check the suitability of the reference sequence and/or sequence database. Use latest official disciplinary taxonomy rules for each species/group of species. Personnel competent in the bioinformatic analyses and the relevant plant health discipline.
	Variant not identified or false variant identified	The identification of variants can be done according to different methodologies, each depending on several parameters. It also depends on the quality control of the reads (e.g. SNPs can originate from low quality reads or sequencing errors).	False negative False positive	Set appropriate threshold parameters (e.g. minimal SNP frequency, coverage, base-calling quality, strand bias) to ensure the reliable identification of variants. Controlled artificial datasets could be used. Sequencing replicates of a single sample.
Third step of the bioinformatic analyses	Targets of control samples not retrieved	One or several threshold(s) of parameters from the previous bioinformatic steps of the HTS test have not been appropriately set.	False negative False positive	Set appropriate threshold parameters (e.g. ????) to ensure the reliable identification of controls. Controlled artificial datasets could be used.

[1]The list of potential sources of errors is not exhaustive.

[2]Other mitigation options may be applied.

The variety of organisms that may be detected and identified by HTS tests from a wide range of matrices makes it impossible to fully validate HTS for each target (which may be a species or a group of organisms). Thus, it is recommended to use a risk-based approach in which the validation data are complemented with the use of other tests performed during the overall diagnostic process, as well as with the use of controls at each step of the HTS process (see section 6.7.2.1 – First line controls). These steps are assessed for their impact and how they can be managed during the risk analysis (see section 6.3.3; Piper *et al*., 2019; Roenhorst *et al*., 2018).

6.3.3 Risk analysis for ensuring the validity of results

Given the complexity of an HTS test, it is recommended to assess holistically the factors influencing the results. The severity of their impact should be estimated, and appropriate measures should be implemented to reduce, minimise or when possible, eliminate the risk (Hébrant *et al*., 2018; Jennings *et al*., 2017).

The risk analysis will be the basis for establishing the critical parameters and quality checks of the HTS test for routine use. The thresholds, acceptable range and proper interpretation, should be defined in the procedure used during routine analysis and be used for continuous monitoring of the performance through time.

6.4 Test selection

Recommendation: **The laboratory should select an HTS test that is suitable to its intended use.**

The HTS tests should be selected according to their intended use, the laboratory constraints and the availability of validated tests.

The laboratory should follow the recommendations of EPPO standard PM 7/98 (2019) for the selection of an HTS test. Specifically, "*tests described in the legislation (e.g. European Union or national legislation)* ***<u>are mandatory</u>*** *for the countries concerned. If no test is mandatory, tests published as international, regional or national standards should, preferably, be used. Whenever such tests are not available or whenever performance could be improved, laboratory-developed or adapted tests can be considered (ISO 17025, 2017, points 7.2.1.1 and 7.2.1.4).*" In the latter case, the laboratory should select and define the most appropriate HTS test(s), including laboratory and bioinformatic components, for further development and optimisation (see sections 6.5 - Test development and optimisation and 6.6 – Validation/verification of HTS test).

6.5 Test development and optimisation

Recommendation: **In the absence of an official HTS standard protocol, the laboratory should carry out an iterative test development and optimisation which aims at writing a detailed protocol of the HTS test, that after validation can be used as a routine diagnostic test.**

The laboratory can be required to develop its own HTS test or adapt a previously published HTS tests to fit the intended use. The HTS test will need to be optimized to ensure it provides the appropriate level of confidence in its results (Hébrant *et al.*, 2018; Maree *et al.*, 2018). The development phase includes the following goals:

1) gaining the necessary experience with the test by identifying the critical steps, parameters and quality metrics that may affect the test performance;
2) defining the most appropriate controls and their continuous monitoring strategy (see section 6.7.2 – Internal and external quality checks);
3) establishing the quality metrics thresholds and acceptable ranges;
4) preparing a detailed protocol describing the optimized test conditions and analysis settings for validation and subsequent routine use.

The development phase should cover both the laboratory and the bioinformatics components of the HTS test. Iterative cycles of protocol development or updates should be performed until all HTS test conditions and analysis protocols meet the minimal predefined performance requirements. The results obtained with the optimized HTS protocol, including its quality metrics, and thresholds should be documented (Aziz *et al.*, 2015; Hébrant *et al.*, 2018; Budowle *et al.*, 2014; Jennings *et al.*, 2017; Roy *et al.*, 2018).

During this phase, the laboratory should also determine the number of samples that can be pooled per sequencing run to achieve the desired minimal number of sequencing reads (see section 6.6.1 – validation of HTS test) and establish baseline cost and turnaround time projections (Rehm *et al.*, 2013).

Designing the panel of samples to be used during the development phase is a key step. The type of matrix, the number of samples, the concentration ranges of targets, the type of genome (DNA vs RNA genomes for viruses), and the type of targets [e.g. closely related organisms that may cross-react with the expected target(s)] (Hébrant *et al.*, 2018; Jennings *et al.*, 2017), need to be defined depending on the intended use of the HTS test. The panel should contain reference material(s) that will be used later as control(s) for the continuous monitoring of the test performance (see Table 5 for the list of controls and section 6.7 – Ensuring the validity of results).

For an HTS test targeting multiple organisms of the same group or from different groups, the panel of samples should be composed at a minimum with the different types of samples expected to be tested routinely with a range of known targets representing the diversity of organisms that may be detected (e.g. bacteria, fungi, viroids, viruses). The concentrations of each target should mimic those in real samples and, for some, be close to the limit of detection. For example, for an HTS test targeting viruses, known viruses representing the different types of viral genomes (e.g. ssRNA, dsRNA, DNA) should be selected (Claverie *et al.*, 2018; Gaafar and Ziebell, 2020). Similarly, for metabarcoding of macroorganisms like insects or plants, individuals should be selected that represent a range of species of the expected main groups (Piper *et al.*, 2019). Depending on the test purpose, the inclusion of commercial microbial community standards as reference material (e.g. Zymo: https://www.zymoresearch.com/collections/zymobiomics-microbial-community-standards) can also be a relevant option for validation of the workflow.

6.6 Validation and verification of HTS test

The ISO 9000 (2015) standard provides a definition of validation: confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled. The ISO 17025 (2017) standard also recommends that "*the validation shall be as extensive as is necessary to meet the needs of the given application or field of application*". The EPPO standard PM 7/98 (2019) recommends the validation of the following performance criteria for a diagnostic test: analytical sensitivity, analytical specificity which includes inclusivity and exclusivity, selectivity, repeatability and reproducibility. As described in EPPO standard PM 7/98 (2019), when a validated test is available, the verification should provide objective evidence that the laboratory personnel is competent to perform this test and to meet the established performance characteristics.

It is advised to evaluate the performance criteria by covering all the steps of an HTS test, starting preferably from biological material. The bioinformatic pipeline can also be validated using sequence datasets obtained from biological reference material or artificial reference datasets (see section 6.7.1 – Reference material and Table 3 - First line controls) containing known target(s) (Budowle *et al*., 2014; Brinkmann *et al*., 2019; Massart *et al*., 2019; Trimme *et al*., 2015).

The reference samples used during validation or verification, should preferably include several targets, most of which at a concentration close to the limit of detection to minimise the risk of false positive due to cross-contamination. The proper combination of targets in such composite samples can allow the simultaneous evaluation of several criteria with fewer samples.

Before performing the validation or verification of an HTS test, a risk analysis should be carried out to identify which performance criteria need to be evaluated and to what extent (see section 6.3.2 – Risk analysis for validation/verification of HTS test). For tests that have been developed and optimized internally, this analysis can be based on the results of the risk analysis performed during the test development and optimisation phase. Figure 5 provides guidance on the decision-making process for validation or verification of an HTS test.

6.6.1 Validation of HTS test

Recommendation: **The laboratory should validate the HTS test.**

The performance criteria provided in EPPO standard PM 7/98 (2019) for the validation of a test (i.e. analytical sensitivity, analytical specificity which includes inclusivity and exclusivity, selectivity, repeatability and reproducibility) are mostly applicable to HTS tests, with some specific points (see sub-sections of 6.6.1).

For HTS tests targeting a broad range of organisms, including uncharacterized organisms (e.g. detection of viruses from *Solanaceae*), it is not possible to develop and validate protocols for analysis of all possible combinations of organism, host or matrix. The validation of the HTS test should focus on the use of key representatives of the targets/pests and mimicking the concentration and composition of real samples expected to be tested routinely (see sections 6.1 - Scope of HTS test and 6.3 - Risk analysis). This approach is similar to what currently proposed for the validation of HTS tests in oncology, where only a representative set of mutations are included in the validation (Hébrant *et al*., 2018; Jennings *et al*., 2017; Roy *et al*., 2018).

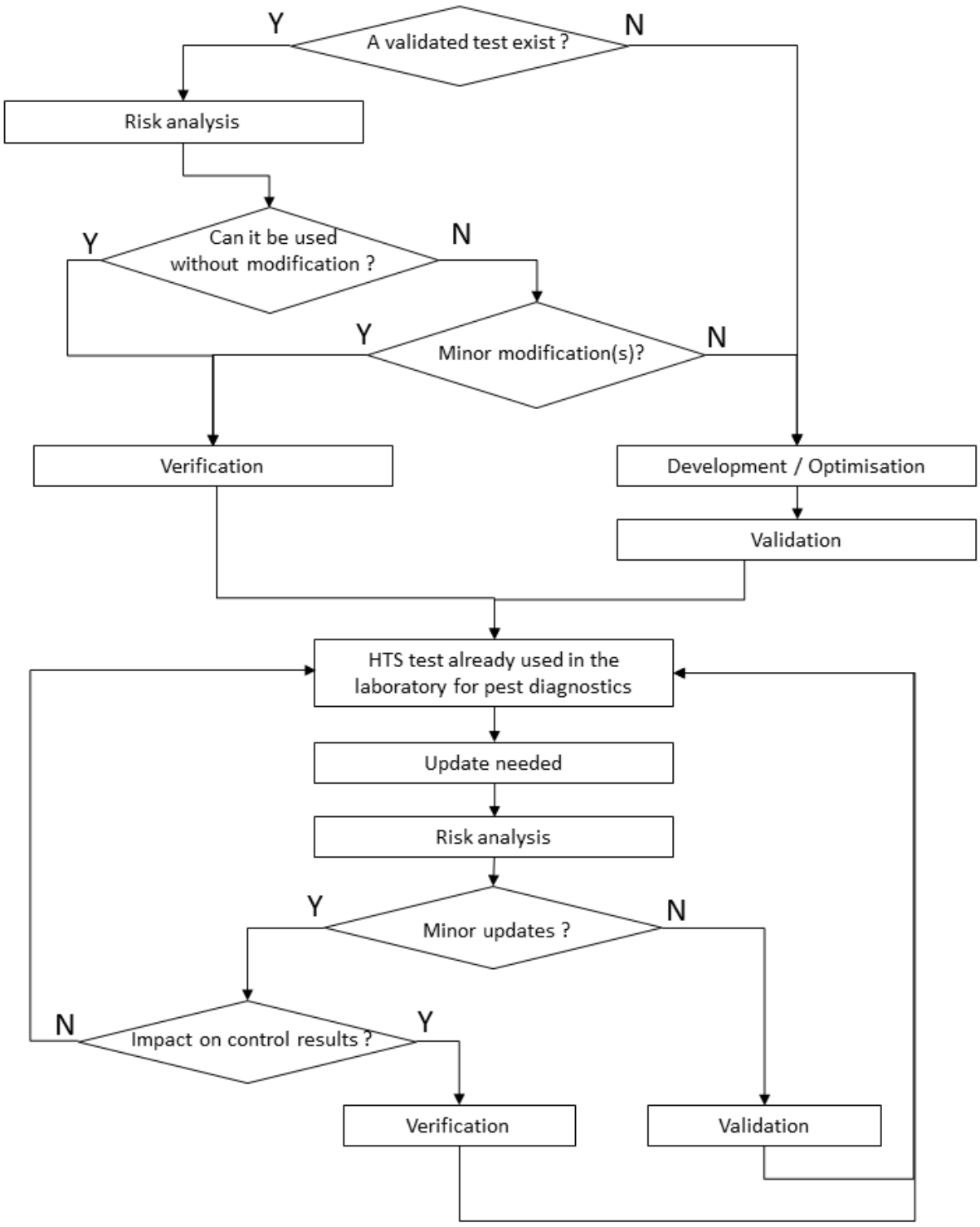

Figure 5. Decision tree on the validation and verification of an HTS test.

6.6.1.1 Specific considerations for analytical sensitivity

Determining the analytical sensitivity is particularly challenging for an HTS test. This is because the analytical sensitivity of an HTS test depends on the number of reads generated per sample. The total number of reads generated during the sequencing run, also called sequencing yield or sequencing depth, will vary between runs. In addition, when several samples are pooled together, the number of sequences per sample will vary and this variability can increase with the level of pooling (see section 6.2.1.5 – Pooling of libraries). This can have a significant impact on the limit of detection. Indeed, if only 1,000 sequences are generated for the sample and a target is present in the nucleic acid extract at a very low proportion, for example 1:10,000 (target nucleic acids : total nucleic acids), it will most probably not be detected and a false negative would be generated. The probability of detecting the target will rise with the increase of the number of generated reads. Indeed, if one million reads are generated, around 100 reads from the target could be expected to be detected, potentially leading to a true positive result. This has been demonstrated for virus detection (Pecman *et al*., 2017; Massart *et al*., 2019; Visser *et al*., 2016). However, increasing the number of reads per sample also increases the probability of detecting contaminating reads (Dr. Heiko Ziebell, Julius Kühn-Institut, pers. comms, Nov 2020). Therefore, an optimal number of generated reads per sample should be defined, as well as the minimal sequencing depth per sample, providing an analytical sensitivity that fits the intended use of the test. This level has been established in the literature using reference samples (Pecman *et al*., 2017), by comparing the results of dilution series of samples containing the relevant range of targets with those of PCR-based tests (Santala and Valkonen, 2018). One should however, also take into account other metrics, such as sequence duplication levels in case of shotgun sequencing. A sample can have many reads, but the diversity of the sequenced molecules could be low due to a poor library preparation.

During validation, the minimal sequencing depth can be evaluated by a 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.

Another observed phenomenon is the impact of co-infecting organisms on the ability of the HTS test to detect a target. This is the case when an organism infecting a sample can be missed because another organism is present at a very high concentration and “masks” that organism (called dilution factor; Maclot *et al*., 2020). This situation has been observed for shotgun sequencing in samples co-infected with viruses or fungi (Rolland *et al*., 2017) and has been shown in amplicon sequencing for fungi using dedicated controls (Chandelier *et al*., 2020). For amplicon sequencing, the composition of the community determines whether or not certain species will be detected, for example due to competition for the primers in the PCR reaction, or differences in copy number of the used barcode. Furthermore, the DNA extraction efficiency also plays an important role. This has been shown with artificial mixes of nematodes where some species were hard to detect by amplicon sequencing, even if they were the only species present in the mix. Because of poor cell lysis, contaminants could be more easily picked up and amplified than the target (Waeyenberge *et al*., 2019). Nevertheless, this phenomenon cannot be currently anticipated for all the combinations of targets tested. To mitigate this risk, the validation could include reference samples with different proportions/quantities of the targets, some very abundant while others at very low level. Such series of controls have been recently used for amplicon sequencing to survey the presence of fungal species in spore traps (Chandelier *et al*., 2020).

To conclude, the analytical sensitivity of an HTS test is theoretically very low as a single read from a target can be potentially identified by an appropriate bioinformatics pipeline. Nevertheless, the analytical sensitivity is limited by the contamination level between samples that can vary between sample batches and runs and, within a batch or a run, between target organisms. The analytical sensitivity will depend on the contamination threshold fixed for the run or the batch. In addition, it will be influenced by the presence of other samples containing the same target within the run or the batch. For example, 10 reads of a target have been detected in a sample. If there are at least one other sample in the batch with a very high abundance of this target (for example 500,000 reads), there is a risk of cross-contamination from this sample. If this target is not detected in any other sample from the batch, the 10 reads are more likely to represent a true infection at very low level. Whatever the situation, and depending on the intended use of the HTS test, a confirmation might or must be carried out (see section 7.1 – Confirmation of the identity of the pest(s)). Further scientific developments are expected to improve the determination of the analytical sensitivity of HTS tests.

6.6.1.2 Specific considerations for analytical specificity

The analytical specificity of an HTS test depends on the strategy used to generate the sequencing library, the genetic variability of the targeted agents, the software and parameters used for the bioinformatic analyses and the reference sequence database(s) (see section 5.6.3 – Reference sequence database). The desired taxonomic resolution (e.g. genes, isolates/strains, pathovars, *formae speciales*, species, genera or families relevant to plant health) should be determined when describing the scope and the intended use of the test (see section 6.1 - Scope of HTS test).

For amplicon sequencing, the analytical specificity of the target region can be at least partially evaluated theoretically by analysing all the targeted regions accessible in sequence databases, taking into account the intended use of the HTS test. The discrimination of closely related organisms based on certain genomic/sequence regions should be studied in depth through bioinformatic analyses. If sequence similarities exist between organisms that could potentially be present in the samples and might interfere with pest detection and identification, those organisms would need to be included in the development and/or validation phases. The taxonomic resolution based on a sequence region can also vary. For example, a genomic region may discriminate all the species in one genus but may be unable to discriminate the species of another genus because of the lack of divergence between these species in that genomic region. The analytical specificity could be evaluated by the use of artificial reads datasets with known pest composition or of positive controls containing a mix of targets whose presence has been confirmed by different methods. Ideally, the concentration of the target(s) should reflect as much as possible the concentration in real samples that will be tested.

When applying a shotgun sequencing protocol to a sample composed of multiple organisms, the analytical specificity might depend on the number of sequences generated from each organism, the percentage of the genome covered, the genomic regions that have been sequenced (conserved or specific) and their read depth. As for analytical sensitivity, the taxonomic resolution of a shotgun sequencing will depend on the sequencing depth and the appropriate target coverage to achieve the intended taxonomic resolution. Sufficient and reproducible sequence coverage and quality needs to be obtained and a minimal number of generated sequences per sample needs to be clearly stated during the development and/or validation phases.

For bacteria, determination of the analytical specificity can be complicated when applying HTS for specific detection of pathogenic bacteria from a complex sample (not a single colony). This is because

of their genome size and the presence of commensal bacterial species which may be related to the pathogenic ones present in the samples. For many bacteria, the appropriate discrimination between family, genus, species or strains may rely on a few specific genes, from which sequences need to be obtained. In cases where no specific genes are available for the identification to species level, the full genome should be obtained, which can be complicated for target organisms that are not isolated. Currently, the sequence databases are not yet representative of the diversity of bacterial species and the limited availability of genome sequences for bacteria will hamper their identification and may lead to the false positive detection of a related species whose genome is in the database(s). The use of a curated databases such as the Genome Taxonomy Database (https://gtdb.ecogenomic.org/) is encouraged.

For fungi, protists, nematodes, arthropod pests, invasive plants or weeds, the determination of the analytical specificity is even more difficult than for bacteria. This is because they have larger genomes than bacteria and of the limited availability of genomic sequences in current sequence databases (In 2020, it is estimated that only a very small proportion of fungal DNA is described in databases with about 1% of fungal species having DNA sequences annotated). In addition, low-quality reference genomes or sequences can be contaminated by microbial sequences (some fungi host bacterial cells, for example *Paenibacillus* spp. can live inside fungi) which could interfere with the calculation of the analytical specificity.

For viruses, determination of the analytical specificity can be achieved because of their small genomes that can be fully sequenced and of the sequence divergence that exists between and across species. However, the sequence variability of the envelopes or coat proteins of viral species is sometimes close to the species threshold. This can be an issue for establishing the limit between divergent isolates and closely related species (for example, four molecular discrimination criteria exist for the family Betaflexiviridae: nucleotides and amino acids percentage for the coat protein and the replication polymerase genes). Therefore, wherever possible, the full genome should be sequenced or at least, several genomic regions should be sequenced although some uncertainties can remain and the demarcation criteria could be met only partially.

6.6.1.3 Other criteria: selectivity, repeatability and reproducibility

According to EPPO PM 7/98 (2019), the selectivity of molecular methods aims to "*Determine whether variations of the matrix affect the test performance*". The variation of matrix can correspond to different cultivars of the same plant species, other plant species, different type of soil, different source of water, or to the community present (e.g. insects, spores, traps). The presence of inhibitors due to variation in the matrix can be monitored using an internal control (see section 6.7.2.1 – First line controls). The use of an internal control with the determination of a minimal sequencing depth (see 6.6.1.1 - Analytical sensitivity), can help in monitoring this criteria during routine diagnostics as the validation process cannot take into account any matrix variation.

A further confounding factor is that the genetic of plants, the composition of soil/water samples can influence the concentration of the target, e.g. different resistance/tolerance to organism(s), changes in the community mix. This can impact the detection of target(s).

For the evaluation of the repeatability of molecular methods, the EPPO standard PM 7/98 (2019) recommends to analyse three replicates of a sample with a low target concentration. For the reproducibility, the same approach is recommended as repeatability, but with different operator(s) if

possible, and on different days and with different equipment when relevant. Wherever possible, HTS tests should be evaluated in the same way.

6.6.2 Verification of an existing validated HTS test

Recommendation: **The laboratory should verify that it can perform an existing validated HTS test according to the established performance characteristics.**

The laboratory interested in using an existing validated HTS test as a routine diagnostic test, should demonstrate its ability to perform the test according to the relevant performance characteristics (section 5.4.2 of EPPO PM 7/98, 2019). The laboratory should prepare a verification plan based on the outcomes of the risk analysis (see section 6.3 - Risk analysis) as stated in EPPO standard PM 7/98 (2019).

The performance criteria provided in EPPO standard PM 7/98 (2019) for the verification of an existing validated test (i.e. analytical sensitivity, analytical specificity which includes inclusivity and exclusivity, selectivity, repeatability and reproducibility) are largely applicable with some additional points specific to HTS tests (e.g. selectivity).

6.6.3 Impact of changes made to a validated HTS test

Recommendation: **The laboratory should evaluate the impact of any changes to a previously validated HTS test and validate or verify its performance if needed.**

The rapid pace of development of protocols (e.g. library kits), sequencing platforms (e.g. length of reads, chemistry/technology) and bioinformatic tools and pipelines (e.g. novel algorithms) may require updates of HTS tests already validated and used in the laboratory. As recommended by EPPO in standard PM 7/98 (2019), the impact of the update should be evaluated by competent personnel (see section 5.2 – Personnel) who will decide whether a verification or a validation is required. This decision should be documented.

The laboratory should determine during the risk analysis (see section 6.3 - Risk analysis) under which conditions (e.g. changes of reference materials or positive control, starting a new batch of critical reagents, changes in a kit composition, changes to algorithms and parameters) and for which HTS steps (e.g. library preparation, bioinformatic analyses) verification or validation is required to ensure the continued performance of the HTS test after an update (Hébrant *et al*., 2018).

In some instances, such as a new version of the library preparation protocol or sequencing kit or an update of bioinformatic software, verification or validation may be not required. If a verification or validation is not required *a priori*, the results obtained with the controls using the modified laboratory protocol and/or bioinformatic pipeline during routine analysis should be used to check for the reliability of results. The laboratory should document the decision, including the reason(s) for which validation/verification is not required and how the modified HTS test will be monitored.

To check the bioinformatic pipeline, the same sequence datasets used during the validation of the HTS test (see section 6.6 – Validation/verification of HTS test) can be used. These sequence datasets can also be completed by datasets generated during routine use of the HTS test. In any case, the sequence

datasets must be representative for the analysed samples, e.g. with a realistic range of targets at different concentrations, including concentrations close to the limit of detection. In addition, the same sequence datasets should be run regularly through the bioinformatic pipeline to make sure that updates to packages or the operating system do not affect the results. This verification and the comparative analysis of the sequence datasets should be documented. After completion of the verification of the bioinformatic pipeline, the datasets from a previous HTS run can be reanalysed to check that they were analysed correctly.

6.7 Ensuring the validity of results

Recommendation: **The laboratory should perform quality checks at appropriate intervals to ensure the validity of test results and to monitor the performance of the test and of the laboratory.**

Recommendations on ensuring the validity of test results stated in EPPO standard PM 7/98 (2019) are valid for HTS tests. The validity of test results should be ensured at different levels, i.e. for each test and diagnostic process, as well as for global quality control of the laboratory (EPPO PM 7/98, 2019). They can be ensured by using reference material (see section 6.7.1) for internal and external quality checks (see section 6.7.2).

The quality checks should be performed at different levels. The first line of control monitors the actual performance of each HTS run (e.g. positive control), the second line of control checks the performance of a single operator within a laboratory (e.g. blind testing) and the third line of control evaluates the performance of the laboratory (e.g. inter-laboratory comparison) (EPPO PM 7/98, 2019).

The types and frequency of quality checks depend on the frequency of use and of the intended use of the HTS test (EPPO PM 7/76, 2018) and should be defined during the risk analysis (see section 6.3 - Risk analysis). If there is any issue revealed by the quality checks, the origin of the issue should be investigated and addressed to prevent its recurrence and thus prevent the reporting of incorrect results (EPPO PM 7/98, 2019; Hébrant *et al.*, 2018; Roenhorst *et al.*, 2018).

Quality metrics should be monitored for each sequencing run and routinely collected and compared to those of an optimal validated run. Any significant deviation should be investigated and may require repeating of the test (for example when one of the targets from the positive controls is not detected in the sequence data). Such data checks can also help investigating the source of the problem in an underperforming test (Hébrant *et al.*, 2018).

6.7.1 Reference material

Recommendation: **The laboratory should use controls, preferably reference materials, for validation and for monitoring the performance of an HTS test and of the laboratory.**

Controls are important in any diagnostic test, including HTS tests, as they "*provide essential traceability in testing*" that can be used, for example, to validate or verify tests or to monitor the performance of the test and/or of the laboratory (EPPO PM 7/98, 2019). These controls can correspond to commercially, publicly or internally available reference materials as defined by EPPO standard PM 7/76 (2018) or to samples characterized properly by the laboratory itself. The recommendations of EPPO on reference material are relevant to HTS tests and therefore should be followed (EPPO PM 7/98, 2019). The preparation of reference materials for HTS tests should follow the recommendations of the VALITEST deliverable 3.3 (2020) and the future EPPO standard on reference materials (under development).

Biological reference material can be fresh, frozen, dried or lyophilized samples containing known target(s) and their stability over time should be ensured. As stated in EPPO standard PM 7/98 (2019), the identity and purity of the biological reference material should be checked by the laboratory to maintain confidence in their status. Additional reference materials specific to HTS tests may also be used, for example, artificial reference materials such as synthesized DNA/RNA, or artificially generated sequencing datasets (generated by computer algorithms). Given the importance of bioinformatic analyses in HTS, the availability of reference sequence datasets is recommended. Such reference datasets should be stored and maintained properly and be preferably publicly available (see section 5.6 – Data management).

Biological reference material can be used to generate a range of working materials with associated data (e.g. annotation of sequences) which can be used as controls to monitor specific step(s) of the HTS process: these can be cultures, nucleic acid extracts, prepared libraries and reference sequence data. Sequence data generated from reference material can be re-analysed over-time to ensure the reliability of the HTS pipeline and be used when an update of the bioinformatic pipeline requires verification or validation. In addition, artificial reference datasets can be produced *in silico* and can supplement real reference datasets produced from reference material.

Depending on the scope of the HTS test, it might be impossible to test for every expected target nor for every single combination of target(s)/matrix. The reference materials should reflect the diversity of the targeted organisms (e.g. RNA viruses, DNA viruses, viroids, bacteria, fungi, phytoplasmas, nematodes and insects) They should be processed together with the samples. Table 2 provides examples of reference materials that can be used in HTS tests.

Table 2. Examples of material that can be used as reference material for HTS depending on the test and targeted pests.

Reference materials	HTS test	Pest range	Notes/remarks
(Mix of) pure culture(s) of strains belonging to representative species expected in the matrix	Amplicon sequencing	Bacteria or fungi	- Strains can be combined into a mock community - Relative proportion of the strains could vary, with some strains close to the limit of detection to evaluate the analytical sensitivity - Composition could also allow the continuous evaluation of the analytical specificity by combining taxonomically related strains
Individuals from taxonomically characterized species	Amplicon sequencing	Insects, nematodes	- a mix of individuals or of their extracted nucleic acids can be combined into a mock community - the relative proportion of the individuals could vary, with some species close to the limit of detection to evaluate sensitivity - composition could also allow the continuous evaluation of the analytical specificity by combining taxonomically related individuals
Individual plant or seed or pollen from taxonomically characterized species	Amplicon sequencing	Invasive plants Weeds	- A mix of individuals or of their extracted nucleic acid can be combined into a mock community - Relative proportion of the individuals could vary, with some species close to the limit of detection to evaluate sensitivity - Composition should also allow the continuous evaluation of the analytical specificity by combining taxonomically related individuals
Matrix infected with known pest(s).	Shotgun sequencing for pest detection	Viruses, viroids, bacteria, fungi, nematodes	- Pest concentration in plants can be variable in different plant organs, over space and time. Lyophilisation of the reference material can provide stable material - Matrix could be infected by a single or by multiple pests reflecting the diversity of expected targets - Several plants/matrices could be combined into a single composite reference sample - Different concentration of the pest: at least one high and at least one close to the limit of detection

Reference materials	HTS test	Pest range	Notes/remarks
			- Isolates/strains of a pest can be spiked in the sample matrix
A pure culture of a strain whose genome is fully sequenced and annotated	Shotgun sequencing for genome characterization and whole genome sequencing	Bacteria or fungi	- Need a representative set of major genetic variants at species, subspecies or pathovar level - Spike the pure culture with plant material as a mock infection
Reference synthetic nucleic acids	All HTS tests	All targets	-Artificial nucleic acids manufactured
Available reference datasets	All HTS tests	All targets	- Generated from the reference materials mentioned above - Artificial reference datasets, or real datasets spiked with reads from the pathogen(s) of interest

6.7.2 Internal and external quality checks

6.7.2.1 First line controls

Different types of first line controls are required for HTS tests. As for targeted molecular tests, HTS technologies require external negative and positive controls. As an alternative (or in addition) to the external positive controls, internal positive controls can be used (see below for details). Another type of control, called alien control for which a detailed description is given below, can be used in HTS tests as an alternative to the external negative and positive controls. Table 3 provides a description of first line controls that can be used in HTS tests and their application in relation to the main steps of the HTS process. The purpose of each category of control in an HTS context is detailed below. It should be emphasised that each step of the HTS test should be monitored during each run (Figure 1 - Main steps of HTS technologies).

Regardless of the type of control, the absence of targets (i.e. no targets in negative controls, absence of other targets in positive and alien controls) or the presence and abundance of target(s) (i.e. positive and alien controls) should be known unequivocally in each control and should be stable over time (see section 6.7.1 – Reference material). The known abundance of target(s) in controls is also important for the determination of a quantitative threshold for contamination (see section 6.2.1.7 – Contamination). This threshold can be an absolute number of reads and/or can be calculated as a relative proportion of reads from the alien targets in the samples and the positive controls. For example, 100 reads of the alien target have been detected as a contamination in a sample or another control. The relative level of contamination will be different if, within the run, the number of reads of this target in the alien control is 1,000 (meaning 10% contamination) or 10,000,000 (meaning 0.01% contamination).

Positive controls

The positive controls such as a positive isolation control (PIC) and a positive amplification control (PAC) are external controls used to monitor the correct detection of targets. Positive controls can be replaced by or used in addition to an alien control in HTS tests (see below section: Alien controls). Table 3 provides details on the application of the positive controls in relation to the main steps of the HTS process.

Positive controls are processed alongside the samples to be analysed. A positive control will usually contain a small but representative fraction of the possible targets because of the broad range of targets an HTS test could detect. It can be prepared as a mix of individual positive controls. It is recommended to use positive controls for which at least some target concentrations are close to the limit of detection. Low level targets are well suited to check the analytical sensitivity of the sequencing run and their low concentration limits the risk of contaminating other samples. Positive controls can also be used to monitor contamination. The detection of an unexpected target in the positive control (in addition to the expected target(s)) may be a signal of contamination from another sample that can be confirmed with the percentage of nucleotide identity of the potential source of contamination.

Negative controls

The negative controls of HTS tests are the same as for any other molecular test and include a negative isolation control (NIC) and a negative amplification control (NAC). They can be used to monitor for contaminants as the detection of target(s) in the negative controls indicates that contamination has occurred during the HTS test. Negative controls may be replaced by or used in addition to an alien control in HTS tests (see below section: Alien controls). Table 3 provides details on the application of the negative controls in relation to the main steps of the HTS process.

A very low amount of contamination by target sequences will often be present in the data generated from negative controls (see section 6.2.1.7 – Contamination). These traces of contamination might not be detectable by conventional or real-time PCR. Contamination can be more prevalent in amplicon sequencing because the amplification of traces of contaminant DNA will be very efficient in the absence of other DNA in the sample. This phenomenon will lead to the risk of overestimating the contamination as compared to a sample or control in which trace contaminant DNA is extremely low in sample DNA. For this reason, the use of positive and/or alien controls containing a DNA quantity similar to the analysed samples could allow a better estimation of contamination in this specific case.

Alien controls

In addition to or as an alternative to the negative and positive controls, a third group of controls, called alien control, can be used in HTS tests. The alien control is processed alongside the samples to monitor the detection of an alien target (role of positive control) and to check for cross contamination between samples (role of a negative control). Table 3 provides details on its application in relation to the main steps of the HTS process.

An alien control corresponds to a matrix containing a target (called alien target) which belongs to the same group as the target organism(s) but cannot be present in the samples to be tested. This alien target can be a pest or not. For example, an alien control can be a bacterial or fungal strain from a species or genus restricted to an ecological niche that is not related to the analysed matrix (e.g. extremophile species with plant samples or spore trapping). For insects or plants, a species restricted to temperate climates could be used as an alien control when analysing tropical crops or environments

(through traps) and *vice versa*. For viruses, a wheat sample infected by barley yellow dwarf virus (BYDV), a *Luteoviridae* infecting only Poaceae, can be used as alien control when analysing viruses infecting potato or banana samples. In this case, the detection of BYDV sequences in the analysed potato or banana samples would indicate that cross contamination has occurred.

The alien control should contain a high concentration of the alien target (for example a plant with a high virus concentration (the higher the better), purified viruses or a pure isolate of bacteria or fungi). A high concentration of the alien target allows a better detection and quantification of alien contamination in the analysed samples. The number and/or proportion of the alien target sequences in the samples can be analysed (e.g. maximum, average, standard deviation, distribution) and compared to the number and proportion of alien target sequences in the alien sample (relative quantification of contamination). If the alien control is also used as a positive control, at least another alien target should be present at a low concentration in addition to the alien target at high concentration to monitor the analytical sensitivity of the HTS test.

As the composition of the alien control is known, the presence of an unexpected target in the generated sequence data from the alien control would also indicate a potential contamination from a sample or another control (when) used in the HTS test (Galan *et al*., 2017).

Internal positive controls

As an alternative or in addition to positive and alien controls, internal positive controls (IPC) may also be used in an HTS test. Table 3 provides details on the application of the internal positive controls in relation to the main steps of the HTS process.

Internal positive controls can correspond to sequences that are expected to be always present in the nucleic acids extracted from the sample (endogenous nucleic acids), for example a plant gene (e.g. nad5 gene, 18S gene, COI) constitutively expressed when analysing RNA shotgun sequencing data from plants to identify pests. Ideally, the selected sequences should be present at a stable and low level in the analysed matrix but above the level of detection to ensure proper monitoring of the analytical sensitivity.

Alternatively each sample is spiked with synthetic nucleic acids or a known target not expected to be found in the samples to be analysed (this target could therefore be another alien control). An advantage of using synthetic nucleic acids is that they are more readily quantifiable than total nucleic acids. The spiked material should be easily and unambiguously detected by the HTS test. It should be spiked at a low concentration (ideally close to the detection level) to evaluate the analytical sensitivity of the test and to avoid masking the targets present in the sample. For example, black bean tissue containing an endornavirus has been used to spike grapevine samples to monitor the sensitivity of the assay and set a threshold for the presence or absence of the target (Kesanakurti *et al*., 2016).

In metabarcoding, synthetic 16S rRNA gene spike-in controls have been used to aid in sample tracking and to detect and quantify cross-contamination that may have occurred during the laboratory processes. A distinct spike-in or mixtures of spike-ins were added in low concentration(s) in each sample before starting the DNA extraction (Tourlousse *et al*., 2018). Similarly, synthetic ITS spike-in controls (mock communities) were used in metabarcoding of forestry fungi. These synthetic controls proved to be useful for monitoring index-hoping and parameterizing the bioinformatic pipelines (Palmer *et al*., 2018).

Table 3. Description of first line controls[1] that may be used in HTS tests and their application in relation to the main steps of the HTS process.

	Negative controls	**Positive controls**	**Alien controls**	**Internal controls**
Aim/Monitoring	Contamination	Contamination Monitor the analytical specificity of the test Monitor the analytical sensitivity of the test when used at low concentration	Contamination Monitor the analytical specificity of the test when use at high concentration Monitor the analytical sensitivity of the test when use at low concentration	Monitor the analytical sensitivity of the test
Description	Same matrix of the analysed samples but free of the target(s) or extraction buffer (NIC), or molecular grade water (NAC)	Same matrix and range of target(s) expected to be detected in the analysed samples and processed alongside the samples and preferably at low concentration (naturally infected or spiked)	A target not expected to be found in the analysed samples (i.e. alien target) and processed alongside the samples and not expected to be detected in the samples to be tested when used at high concentration and/or expected to be detected only in the alien control when used at low concentration	Non target nucleic acids not related to the sample targets naturally present (e.g. plant genes) or known target spiked at low concentration in the samples (e.g. synthetic nucleic acids, known target not expected to be found in the samples to be tested)
Analysis	Absence of target(s) Contamination target(s) below a set threshold (*) (**)	Presence of positive control targets Contamination targets below a set threshold (**)	Absence of the alien target in the analysed samples (when used at high concentration) (**) Contamination targets below a set threshold (**) Presence of expected alien target in the alien control (when used at low concentration)	Presence of internal control in each sample

HTS steps	Negative controls	Positive controls	Alien controls	Internal controls
Sampling and nucleic acids extraction	Negative isolation control (NIC): matrix without target(s), if not available, extraction buffer	Positive isolation control (PIC) Matrix containing target(s) from a single or pooled individual(s)	Matrix containing alien target processed alongside samples	Not applicable as included in the analysed samples
Library preparation	Nucleic acids previously extracted from a NIC during another HTS test (that can be used as a NAC) Molecular grade water to verify the absence of contamination (*)(***)	Nucleic acids previously extracted from a PIC during another HTS test (that can be used as a PAC)	Nucleic acids previously extracted from an alien control during another HTS test	Spiked nucleic acids to be analysed with non-target nucleic acids of natural, synthetic origin or known target not expected to be found in the samples to be analysed
Sequencing	Previously prepared libraries from the respective controls can be sequenced for specific monitoring of sequencing DNA sequence of the positive controls designed by the HTS technology manufacturer, present in the sequencing reagents.			
Bioinformatic analysis	Raw sequencing data generated during previous HTS tests from respective controls or artificially generated data can be used to specifically monitor the bioinformatic analysis			

[1]Abbreviations of first line controls: NAC: negative amplification control, NIC: negative isolation control, PAC: positive amplification control, PIC: positive isolation control.

(*) the absence of target sequences is practically nearly not possible in a negative control

(**) if an unexpected target is detected in any control or an alien target is detected in the samples, their presence should be quantified and compared with the controls and samples infected by the target

(***) for shotgun sequencing, the same matrix as the analysed samples but free of the target(s) is preferred over molecular grade water as negative control.

In shotgun sequencing, a synthetic community of artificial microbial genomes called sequins (standing for sequencing spike-ins) mimicking the microbial community of the real samples, can be added to environmental DNA samples prior to library preparation. This enables the measurement and mitigation of technical variation (e.g. library preparation protocols) that can influence sequencing. Sequins also provide a constant reference that can be used during the development and optimization of HTS tests (Hardwick *et al*., 2018). Synthetic RNA spike-ins sets have also been used on zebrafish total RNA extracts for monitoring size-selection of RNA and for sample-to-sample normalization of RNA in small RNA sequencing. This improves the technical reproducibility of the test (Locati *et al*., 2015) but such an approach has not yet been evaluated in plant pest diagnostics.

Internal sequencing controls designed by the HTS technology manufacturers, are available for some sequencing platforms. Manufacturer's instructions should be followed when using these controls. For example, for the Illumina technology, the PhiX phage is used to monitor the sequencing run and is included in sequencing reagents and is always spiked in any sequencing reaction. Its genome sequence is known and it is therefore used to automatically evaluate the accuracy of sequencing (e.g. the proportion of sequencing errors). Similarly, Oxford Nanopore Technologies have a control sequence that can be spiked in.

Commercialised spike-in controls are now becoming available. For example, a common set of external RNA controls called ERCC RNA spike-in mix, has been developed by ThermoFisher Scientific for RNA analysis, including gene expression profiling and whole transcriptome surveying. This control has been used routinely in some plant health diagnostic laboratories.

6.7.2.2 Replicates

Recommendation: **The laboratory could consider the use of biological and/or technical replicates appropriate to the intended use of the HTS test.**

Biological and/or technical replicates can be used to validate the results although the costs can be prohibitive for example for library preparation. Technical or biological replicates could be more affordable for amplicon sequencing due to the lower costs per sample.

Additive processing (i.e. pooling the replicates) can be useful for overcoming sampling stochasticity and controlling for false-negative results, while restrictive processing (i.e. only retaining sequences present in several replicates) effectively controls for cross-contamination. To balance the merits of both approaches, it may be best to include a minimum number of technical or biological replicates to allow a majority-rules approach (e.g. 2/3 replicates count as a detection; Piper *et al*., 2019). The processing of replicates could be systematic for only a few samples or the controls and would be limited by their costs.

6.7.2.3 Second and third line controls and performance monitoring

The second line of control checks the performance of a single operator within a laboratory (e.g. blind testing) and the third line of control evaluates the performance of the laboratory (e.g. inter-laboratory comparison). A list of second and third line controls can be found in EPPO standard PM7/98 (2019).

The types and frequency of quality checks depends on the frequency of use and of the intended use of the HTS test (EPPO PM 7/76, 2018) and should be defined during the risk analysis (see section 6.3). If there are any issue resulting from the quality checks, the source of the issue should be investigated and addressed to prevent its recurrence and thus preventing reporting of incorrect results (EPPO PM 7/98, 2019; Hébrant *et al*., 2018; Roenhorst *et al*., 2018).

7 Confirmation, biological interpretation and reporting

7.1 Confirmation of the identity of the pest(s)

Recommendation: **The laboratory should confirm the identity of the pest according to the risks it poses to plant health.**

The confirmation of the identity of pest(s) obtained by HTS tests is similar to any other diagnostic tests. The need to confirm the identity of a pest depends on the context of the analysis and on the type of organism identified (e.g. identification of a regulated pest [i.e. quarantine pest or regulated non-quarantine pest] should be confirmed). The results should be confirmed for the critical cases described in EPPO standard PM 7/76 (2018). These critical cases are "*the detection of a pest in an area where it is not known to occur, cases where a pest is identified by a laboratory for the first time",* and "*detection of a pest in a consignment originating from a country where the pest is declared to be absent*"(EPPO PM 7/76, 2018). The identity of any uncharacterized organism with potential risks to plant health should also be confirmed and should be documented. For example, an apparently virulent strain of *Xanthomonas sontii*, a species that is normally considered to be a harmless endophyte, was identified taxonomically by HTS. Further testing and Koch's postulates were required, given that it was a surprising candidate for causing disease (Mirghasempour *et al*., 2020).

For non-regulated organisms commonly found in a particular host, the requirement of confirmatory tests may not be necessary (for example viruses with a wide host range and geographical distribution such as cucumber mosaic virus). However, a confirmatory test of such organisms can be useful in some situations, for example for targeted pest management. Similarly, a confirmatory test can be useful when a non-regulated organism is found associated with unexpectedly severe or unusual symptoms or a new host. Some bacteria and fungi can cause very different severity of symptoms depending on the host plant species. Some examples are *Calonectria pseudonaviculata*, *Diplodia corticola* and *Xanthomonas* strains (Bérubé *et al*., 2018; Constantin *et al*., 2017; Malapi-Wight *et al*., 2016).

When confirmation is needed, it is recommended to use a test or a combination of tests based on different biological principles and/or re-extract nucleic acids from the same matrix and re-test by other molecular methods. When possible and relevant, the viability or pathogenicity of the pest should be confirmed (e.g. for the critical cases listed above). If available, validated tests should be preferred. General characteristics of methods for plant virology have been reviewed by Roenhorst *et al.* (2018), for plant-parasitic nematodes by Castagnone-Sereno *et al*. (2011) and for fungi, based on a case study of *Phytophthora* by Martin *et al.* (2000). Diagnostic protocols developed by EPPO, are available for a range of regulated pests (https://www.eppo.int/RESOURCES/eppo_standards/pm7_diagnostics).

If no other test is available to confirm the identity of the pest (i.e. an organism that has never been detected/identified before), primers should be designed based on the HTS sequence data and available sequence information in sequence databases. Whenever possible, the primers should be designed to

maximise the inclusivity of the test. Alternatively, generic primers allowing the detection of several pests, including the detected one(s), followed by Sanger sequencing of the amplicon could be used to confirm the identity of the pest.

Table 4. Examples of situations when a confirmatory test can be required.

Scenario	Confirmatory test required?	Example	Comment
Characterized organism found on known host	Confirmatory test is required for the critical cases described in EPPO standard PM 7/76 (2018) Confirmatory test is required for the confirmation of the identity of regulated pests [i.e. quarantine pests or regulated non-quarantine pests]	Any potato pest on EPPO A1 list, identified on potato in post entry quarantine testing EPPO standard PM 3/21 (2019) would need to be confirmed; *Xylella fastidiosa* (EPPO A2 list) can be identified by HTS (Bonants *et al*., 2019) but should be confirmed by molecular or serological tests (EPPO PM 7/24, 2019)	Known organisms with existing test
	Confirmatory test of non-regulated pests may be required in other situations such as export certification, seed and reproduction material certification, diagnostics from field samples or environmental samples like spore or insect traps	Southern tomato virus, a non-regulated virus was detected by HTS in tomato from Germany and its presence was confirmed by conventional RT-PCR (Gaafar *et al.*, 2019a); Screening of exotic forest pathogens using metagenomics (Tremblay *et al*., 2018)	
Characterized organism found on known host with unusually severe or novel symptoms	Confirmatory test on non-regulated pests may be required in some situations such as crop protection, pest management Confirmatory test is required for the confirmation of the identity of regulated pests [i.e. quarantine pests or regulated non-quarantine pests]	An isolate of pepino mosaic virus (CH2) responsible for different symptoms on tomato plants (Hanssen *et al.*, 2009) Resistance breaking isolates of barley yellow mosaic virus Y and barley mild mosaic virus in barley (Rollands *et al*., 2017)	Emerging pathogens with existing test

Scenario	Confirmatory test required?	Example	Comment
Characterized organism found on novel host	Confirmatory test on non-regulated pests may be required in some situations such as crop protection, pest management Confirmatory test is required for the confirmation of the identity of regulated pests [i.e. quarantine pests or regulated non-quarantine pests]	First report of natural infection of beetroot with beet soil-borne virus (Gaafar *et al.*, 2019b)	Known organisms with existing test
Uncharacterized organism or poorly characterized organism with potential risk to plant health	Confirmatory test may be required	Several uncharacterized viruses detected in *Ullucus tuberosus* by HTS and confirmed by real-time RT-PCR (Fox *et al*., 2019)	No existing tests
Uncharacterized organisms with unknown/unlikely risk to plant health	Confirmatory test optional	Mycoviruses, some endornaviruses and partitiviruses usually non-pathogenic to the plants (Lee Marzano and Domier, 2016)	Any sequence data related to uncharacterized organisms with unknown/unlikely risk to plant health should be kept for future analysis as sequence databases evolve and identification may become possible

Sometimes it may not be possible to confirm the presence and the identity of a pest in a sample. In such case(s), the laboratory should document the results and its decision for quality assurance purposes and in case further work should be conducted.

The laboratory should have a procedure describing when a confirmatory test is required (Aziz *et al*., 2015). Table 4 provides examples of situations when a confirmatory test can be required.

7.2 Interpretation of the biological relevance of the identified target(s)

Recommendation: **The laboratory should endeavour whenever needed to interpret the biological relevance of the target(s) identified by HTS**.

The interpretation of the biological relevance is important for evaluating the potential risk the detected organism(s) would pose to plant health. It applies mainly to poorly characterized and uncharacterized organisms and in some cases, to known organism unexpectedly found in a new host. Understanding the biological relevance of the target(s) identified by an HTS test, is important for determining the potential risk to plant health as it will be useful information to the national or international bodies conducting risk evaluation. However, the biological characterisation may take time or may not be possible for various reasons (e.g. lack of human and/or financial resources) or be carried out by another laboratory.

HTS data do not provide any information on the biological relevance of the sequences identified, whether they correspond to a pathogenic organism with associated risks or whether the detected nucleic acids come from living organisms. For example, detected viral sequences may correspond to a *bona fide* virus infecting other organisms associated with the sample, including bacteria, fungi or arthropods (Al Rwahnih *et al*., 2011; Marzano and Domier, 2016) or to viral sequences integrated into the plant genome (Baizan-Edge *et al*., 2019; Brinkmann *et al*., 2019; Massart *et al*., 2017 and 2019). Bacterial, fungal or viral sequences attributed to a pest species might be originating from closely related species that are not pathogenic but living as endophytes without causing any harm under the specific environmental conditions. Relevant scientific expertise (see section 5.2 - Personnel) is essential to biologically interpret HTS results and their implications, in particular in case of the identification of a target at a low concentration, a poorly characterized organism or an uncharacterized organism, and of viral sequences that might result from integration in the host genome (Brinkmann *et al*., 2019; Massart *et al*., 2019).

The extent of the biological characterisation depends on the potential risk the detected organism(s) would pose to plant health. The scaled and progressive scientific framework proposed by Massart *et al.* (2017) is a useful tool that can be used by plant health stakeholders to perform the biological characterisation and the risk assessment of an uncharacterized or poorly characterized plant virus detected by HTS. If carried out partially or totally by the diagnostic laboratory, it should always document the decisions related to the biological characterisation of the identified organism(s).

The interpretation of the biological significance should cover some or all of the following items, depending on the context of the analysis. The information should be documented.

-*Sample information*: The recommendations in EPPO standard PM 7/77 (2019) on the information of samples to be recorded should be followed. The following sample metadata can be used to support biological interpretation: information about the nature of the material (i.e. host identity to species and, whenever possible, to cultivar level and part of the plant sampled), the precise description of symptoms (if any) and time of appearance (if available), the time of sampling, the geographical origin of the sample and any other information relevant for the biological interpretation of the HTS results (e.g. estimation of the extent of infestation, hosts destined for import or export, size of the consignment) (EPPO PM 7/77, 2019; Massart *et al*., 2017).

-*Taxonomic information*: The (provisional) taxonomic position of a sequence can provide some information on its biochemical properties (e.g. bacteria belonging to a taxonomic group that have specific biochemical properties) and/or morphological characteristics (e.g. insects and nematodes) and

even its biology. For example, for plant viruses, the taxonomic position can give an indication on the putative host range and its potential pathogenicity to these hosts, the modes of horizontal and/or vertical transmission, including the identification of candidate vectors (Massart *et al*., 2017). However, these properties should be confirmed.

-*Genome information*: When relevant and wherever possible, identification of putative genes and the prediction of relevant gene products and functions (especially those associated with potential phytosanitary risks) should be determined (Budowle *et al.,* 2014). This is particularly important when an organism/agent new-to-science is detected, allowing to differentiate between known pathogenic and non-pathogenic organisms or strains (Zaluga *et al*., 2014). For example, virulence genes were found in three bacterial species consistently detected in the necrotic stem lesion of acute oak decline disease (Denman *et al*., 2018).

-*Confirmation of the results*: See section 7.1 - Confirmation of the identity of the pest(s).

-*Causation/aetiology:* Evidence of disease association is especially important when dealing with diseases potentially caused by several organisms (Lamichhane and Venturi, 2015). For example, a number of organisms (pathogenic and non-pathogenic) were identified in the disease of acute decline of oak (caused by bacteria) and in the disease of carrot internal necrosis (caused by viruses) (Adams *et al*., 2014; Denman *et al*., 2018). Some complex diseases may also be influenced by abiotic factors such as temperature, moisture, stage of host development (Denman *et al*., 2018). The understanding of organism interactions with the influence of abiotic factors and the evidence on the causative agent(s) will assist in minimising the potential risk to plant health by developing appropriate management strategies and by taking informed decisions in terms of phytosanitary action.

Fulfilling Koch's postulates, where one pathogen causes one disease, can be impractical in the HTS era and does not apply to diseases caused by several organisms and abiotic factors. Instead Denman *et al*. (2018) used a combination of sequencing and cultivation-based approaches to determine the biotic components of a complex decline-disease, acute decline of oak. Similarly, Adams *et al*. (2014) used a combination of molecular tests (conventional PCR and HTS) with a statistical approach to determine which viruses were associated with internal necrosis in carrots. Fox (2020) proposed a systematic integrated approach for plant virology, combining epidemiological observations supported by statistical analysis. The proposed approach may possibly be extended, with some modification to other plant health disciplines.

-*Viability of the organisms:* The determination of the viability of an organism can be required in some instances (depending on the intended use of HTS tests), e.g. bacteria, nematode cysts or insects after phytosanitary treatment. It can also be the case of organisms that could become viable, e.g. virus sequences integrated in plant genome leading to a replicative form; Massart *et al*., 2019). If the organism is a regulated pest, recommendations provided in EPPO standard PM 7/76 (2018) on the viability of organisms should be followed. Appropriate viability tests (when available) should be conducted (e.g. agar plating for bacteria and fungi, fluorescent viability stain for bacteria such as BacLight). This is particularly important when a pest is detected outside of its host(s) as it was shown for plant viruses in wastewater (Bačnik *et al*., 2020), for the bacterium, *Ralstonia solanacearum* in water courses (Caruso *et al*., 2005), *Phytophthoras* in soils (Riddell *et al*., 2019) and for the nematode, *Pratylenchus penetrans* (Orlando *et al*., 2020).

7.3 Reporting

7.3.1 General recommendations

Recommendation: **The laboratory should report diagnostic results of an HTS test according to the local, national and international legislation and should follow the recommendations of the EPPO standard PM 7/77 (2019) on the content of the report with additional information for the detection of uncharacterized organisms with potential risks to plant health.**

Regardless of the organism detected by an HTS test, either a known pest (expected or unexpected), a poorly characterized organism or an uncharacterized organism, the reporting of the diagnostic results should follow the recommendations of the EPPO standard PM 7/77 (2019), i.e. "*the result of a diagnosis should be reported accurately, clearly, unambiguously and objectively*." The diagnostic report should be adapted to the need of the client (e.g. different information may be required for a grower and the NPPO) (EPPO PM 7/77, 2019) and the confidentiality of the results to the client should be guaranteed (EPPO PM 7/98, 2019). The reporting of HTS test results should be accompanied with an expert judgement and with other confirmatory tests, when needed. This is particularly important for the reporting of uncharacterized organisms.

The laboratory should have a procedure to ensure that the findings of regulated pests or new pests are reported to the relevant NPPO, as recommended in EPPO standard PM 7/98 (2019). The laboratory should also have a procedure on reporting to the NPPO the finding of any uncharacterized organisms with a potential risk to plant health. Information to consider in the report to NPPO includes (if relevant):

-relationship with other organisms in the same taxon (e.g. closely related to an economically important pest)
-relationship with its host (e.g. mycovirus, insect virus)
-potential risk of causing damage to its host
-potential risk for other hosts (economically and/or ecologically important)
-potential risk of spreading
-location risk (e.g. horticultural area versus isolated area)
-viability of the organism (e.g. bacteria alive or dead, virus sequence integrated in plant genome leading to replicative form)
-possible influence of abiotic factors
-presence of other organisms in the same host (e.g. symbiotic or antagonistic effect)
-recommendation for re-sampling/re-testing or other extended analyses

7.3.2 Inconclusive results

Recommendation: **The laboratory should report inconclusive results of an HTS test following the recommendations of the EPPO standard PM 7/76 (2018).**

As for any other molecular diagnostic test, inconclusive results (e.g. unable to confirm the presence of an organism because of the HTS results falling in the "grey zone" and/or lack of confirmatory tests) may be obtained with an HTS test (Boukari *et al.*, 2020). Their reporting should follow the recommendations of EPPO standard PM 7/76 (2018), i.e. an explanation of the source of the

uncertainty should be provided. The sources of uncertainty in an HTS test can be that the level of the pest is close to the limit of detection, it is present only in a single technical replicate (e.g. one out of two or three replicates), the poor sample quality, the difficulty in distinguishing between episomal and integrated viruses, the discovery of a previously uncharacterized organism for which it is not clear if the tested plant is the host, the lack of completeness of the database, the lack of knowledge whether a bacterium/fungus can be endophytic for some plant species but harmful for others, the limitations of the barcode used. The repeat of the HTS test and/or resampling should be discussed with the client and the decision documented.

7.3.3 Detection of unexpected organisms

Recommendation: **The laboratory should have a policy on how to deal with the detection of unexpected organisms, and should follow local, national and international legislation.**

The detection of an unexpected organism is likely to happen during an HTS test targeting a group of organisms. The decision whether to report should be made by competent personnel (see section 5.2 – Personnel) (McGuire *et al*., 2013; Van El *et al*., 2013). The laboratory should have a policy on how to deal with the detection of unexpected organisms. The policy should be agreed with the NPPO and follow local, national or international guidelines, when available (Christenhusz *et al*., 2013).

Any unexpected organism detected by HTS that may pose a potential risk to plant health should be reported as per EPPO standard PM 7/77 (2019) after the recommended confirmatory testing (see section 7.1 – Confirmation of the identity of the pest(s)). For example, the detection of an unexpected organism in imported plants (e.g. a known organism of economic significance or a new to science organism that may pose a risk to plant health) could potentially have trade issues (Maree *et al*., 2018). A study on the screening for fungi by HTS (with confirmation by real-time PCR) in environmental samples originating from air and insect traps in Canada showed the importance of maintaining surveillance of the genus *Heterobasidion*, a genus of economic concern, although it is not part of the Canadian regulated pest list (Tremblay *et al*., 2018).

The reporting of unexpected organisms that may not pose a risk to plant health (e.g. endophytes, beneficial insects) would depend on the context of the analysis (e.g. metagenomics, metabarcoding). This is the case for example of cryptic viruses such as mycoviruses, endornaviruses, partitiviruses and some viruses infecting insects, where some interactions between the virus and its host can be beneficial (Kreuze *et al*., 2020; Roossinck, 2015). Similarly, insect or environmental DNA metabarcoding analyses could enable the record of unexpected beneficial insects (such as pollinators, parasitoids or biological control agents), previously unknown to be present in the area (Thomsen and Sigsgaard, 2019). A study on aerial spore samples by metabarcoding in Canada revealed the presence of *Diplodia corticola*, a fungal species considered an opportunistic plant endophyte capable of living asymptomatically for several years before changing to a pathogen when conducive conditions arise (Bérubé *et al*., 2019).

7.3.4 Additional remarks and disclaimers

Recommendation: **The laboratory should include in the report any additional remarks and disclaimers related to any limitation in the HTS test performance and analysis.**

The laboratory should include in the report additional remarks and disclaimers related to any limitation in the HTS test (for example, the impossibility to distinguish viable and dead pests) and in the performance analysis of the sample (Hébrant *et al*., 2018; Weiss *et al*., 2013). Indeed, the HTS test results depend on the algorithms and sequence databases used. If confirmatory tests have been carried out (like bioassay or viability assays), some limitations of the HTS test may not be relevant.

As for any other diagnostic test, the HTS test results may be affected by the quality of the sample received. In this case, the report may state that the results apply to the sample as received (ISO 17025, 2017).

8 Conclusion

This report provides technical recommendations for the selection, development, validation, verification and routine use of HTS with the assurance of the validity of results, the interpretation and reporting of HTS test results. It has been designed to be as practical as possible with enough flexibility to remain up to date as the technologies evolve. The proposed recommendations have been written based on real examples and when relevant, illustrated to explain complex processes. The guidelines apply to any HTS test used for the detection and identification of any plant pest from any type of matrices by plant health diagnostic laboratories.

The proposed guidelines will be presented to the Panel Diagnostics and Quality assurance for their consideration as to whether they could be the basis for a new EPPO standard on HTS tests.

REFERENCES

Abdelfattah A., Sanzani S.M., Wisniewski M., Berg G., Cacciola S.O., Schena L. (2019) Revealing cues for fungal interplay in the plant-air interface in vineyards. *Frontiers in Plant Science*, 10: 922, doi: 10.3389/fpls.2019.00922.

Ackelsberg J., Rakeman J., Hughes S., Petersen J., Mead P., Schriefer M., Kingry L., Hoffmaster A., Gee Jay E. (2015) Lack of evidence for plague or anthrax on the New York city subway. *Cell Systems*, 1: 4-5, http://dx.doi.org/10.1016/j.cels.2015.07.008.

Adams I.P., Skelton A., Macarthur R., Hodges T., Hinds H., Flint L., Nath P.D., Boonham N., Fox A. (2014) *Carrot yellow leaf virus* is associated with carrot internal necrosis. *PloS ONE* 9 (11): e109125, doi: 10.1371/journal.pone.0109125.

Adams I.P., Fox A. (2016) Diagnosis of plant viruses using next-generation sequencing and metagenomics analysis. *In*: Wang A., Zhou X. (eds.), Current research topics in plant virology, Springer International Publishing, Switzerland, pp. 323-335, doi: 10.1007/978-3-319-32919-2_14.

Adams I.P., Fox A., Boonham N., Massart S., De Jonghe K. (2018) The impact of high throughput sequencing on plant health diagnostics. *European Journal of Plant Pathology*, 152: 909-919, https://doi.org/10.1007/s10658-018-1570-0.

Aguayo J., Fourrier-Jeandel C., Husson C., Ioos R. (2018) Assessment of passive traps combined with high-throughput sequencing to study airborne fungal communities. *Applied and Environmental Microbiology*, 84: e02637-17, https://doi.org/10.1128/AEM.02637-17.

Ahmed M., Back M.A., Prior T., Karssen G., Lawson R., Adams I., Sapp M. (2019) Metabarcoding of soil nematodes: the importance of taxonomic coverage and availability of reference sequences in choosing suitable marker(s). *Metabarcoding and Metagenomics*, 3: 77-99, doi: 10.3897/mbmg.3.36408.

Al Rwahnih M., Daubert S., Úrbez-Torres J.R. (2011) Deep sequencing evidence from single grapevine plants reveals a virome dominated by mycoviruses. *Archives of Virology*, 156: 397-403, doi: 10.1007/s00705-010-0869-8.

Al Rwahnih M., Daubert S., Golino D., Islas C., Rowhani A. (2015) Comparison of next-generation sequencing versus biological indexing for the optimal detection of viral pathogens in grapevine. Phytopathology, 105: 758-763, https://doi.org/10.1007/s10658-018-1570-0.

An J.-H., Noh Y.-H., Kim Y.-E., Lee H.-I, Cha J.-S. (2015) Development of PCR and TaqMan PCR assays to detect *Pseudomonas coronafaciens*, a causal agent of Halo blight of oats. *The Plant Pathology Journal*, 31 (1): 25-32, http://dx.doi.org/10.5423/PPJ.OA.09.2014.0096.

Andersen M.T., Templeton M.D., Rees-George J., Vanneste J.L., Cornish D.A., Yu J., Cui W., Braggins T.J., Babu K., Mackay J.F., Rikkerink E.H.A. (2018) Highly specific assays to detect isolates of Pseudomonas syringae pv. *actinidiae* biovar 3 and *Pseudomonas syringae* pv. *actinidifoliorum* directly from plant material. *Plant Pathology*, *67* (5): 1220-1230, https://doi.org/10.1111/ppa.12817.

Aritua V., Musoni A., Kabeja A., Butare L., Mukamuhirwa F., Gahakwa D., Kato F., Abang M.M., Buruchara R., Sapp M., Harrison J., Studholme D.J., Smith J. (2015) The draft genome sequence of

Xanthomonas species strain Nyagatare, isolated from diseased bean in Rwanda. *FEMS Microbiology Letters*, 362 (4): 1-4, https://doi.org/10.1093/femsle/fnu055.

Anslan S., Nilsson R.H., Wurzbacher C., Baldrian P., Tedersoo L., Bahram M. (2018) Great differences in performance and outcome of high-throughput sequencing data analysis platforms for fungal metabarcoding. *MycoKeys*, 39: 29, https://doi.org/10.3897/mycokeys.39.28109.

Asplund M., Kjartansdóttir K.R., Mollerup S., Vinner L., Fridholm H., Herrera J.A.R., Friis-Nielsen J., Hansen T.A., Jensen R.H., Nielsen I.B., Richter S.R., Rey-Iglesia A., Matey-Hernandez M.L., Alquezar-Planas D.E., Olsen P.V.S., Sicheritz-Pontén T., Willerslev E., Lund O., Brunak S., Mourier T., Nielsen L.P., Izarzugaza J.M.G., Hansen A.J. (2019) Contaminating viral sequences in high-throughput sequencing viromics: a linkage study of 700 sequencing libraries. *Clinical Microbiology and Infection*, https://doi.org/10.1016/j.cmi.2019.04.028.

Aziz N., Zhao Q., Bry L., Driscoll D.K., Funke B., Gibson J.S., Grody W.W., Hegde M.R., Hoeltge G.A., Leonard D.G.B., Merker J.D., Nagarajan R., Palicji L.A., Robetorye R.S., Schrijver I., Weck K.E., Voelkerding K.V. (2015) College of American Pathologists' laboratory standards for next-generation sequencing clinical tests. *Archives of Pathology & Laboratory Medicine*, 139: 481-493.

Bačnik K., Kutnjak D., Pecman A., Mehle N., Tušek Žnidarič M., Gutiérrez Aguirre I., Ravnikar M. (2020) Viromics and infectivity analysis reveal the release of infective plant viruses from wastewater into the environment. *Water research*, 177: 115628, https://doi.org/10.1016/j.watres.2020.115628.

Baizan-Edge A., Cock P., MacFarlane S., McGavin W., Torrance L., Jones S. (2019) *Kodoja*: a workflow for virus detection in plants using *k-mer* analysis of RNA-sequencing data. *Journal of General Virology*, 100: 533-542, doi: 10.1099/jgv.0.001210.

Baldi P. La Porta N. (2017) *Xylella fastidiosa*: Host range and advance in molecular identification techniques. *Frontiers in Plant Science*, 08: 944, https://doi.org/10.3389/fpls.2017.00944.

Barba M., Czosnek H., Hadidi A. (2014) Historical perspective, development and applications of next-generation sequencing in plant virology. *Viruses*, 6: 106-136, doi:10.3390/v6010106.

Bérubé J.A., Gagne P.N., Ponchart J.P., Tremblay É.D. Bilodeau G.J. (2018) Detection of *Diplodia corticola* spores in Ontario and Québec based on high throughput sequencing (HTS) methods. *Canadian Journal of Plant Pathology*, 40 (3): 378-386, https://doi.org/10.1080/07060661.2018.1498394.

Bonants P.J.M., van Gent-Pelzer M.P.E., van Leeuwen G.C.M., van der Lee T.A.J. (2015) A real-time TaqMan PCR assay to discriminate between pathotype 1 (D1) and non-pathotype 1 (D1) isolates of *Synchytrium endobioticum*. *European Journal of Plant Pathology*, 143: 495-506, https://doi.org/10.1007/s10658-015-0702-z.

Bonants P., Griekspoor Y., Houwers I., Krijger M., van der Zouwen P., van der Lee T.A.J., van der Wolf J. (2019) Develoment and evaluation of a triplex TaqMan assay and next-generation sequence analysis for improved detection of *Xylella* in plant material. *Plant Disease*, 103: 645-655.

Boukari W., Mollov D.S., Wei C., Tang L., Grinstead S.C., Tahir N., Mulandesa E., Hincappie M., Beiriger R., Rott P. (2020) Screening for sugarcane yellow leaf virus in sorghum in Florida revealed its occurrence in mixed infections with sugarcane mosaic virus and a new marafivirus. *Crop Protection*, 139: 105373, https://doi.org/10.1016/j.cropro.2020.105373.

Braukmann T.W.A., Ivanova N.V., Prosser S.W.J., Elbrecht V., Steinke D., Ratnasingham S., de Waard J.R., Sones J.E., Zakhhharov E.V., Hebert P.D.N. (2019) Metabarcoding a diverse arthropod mock community. *Molecular Ecology Resources*, 19: 711-727, https://doi.org/10.1111/1755-0998.13008.

Brinkmann A., Andrusch A., Belka A., Wylezich C., Höper D., Pohlmann A., Nordahl Petersen T., Lucas P., Blanchard Y., Papa A., Melidou A., Oude Munnink B.B., Matthijnssens J., Deboutte W., Ellis R.J., Hansmann F., Baumgärtner W., van der Vries E., Osterhaus A., Camma C., Mangone I., Lorusso A., Maracci M., Nunes A., Pinto M., Borges V., Kroneman A., Schmitz D., Corman V.M., Drosten C., Jones T.C, Hendriksen R.S., Aarestrup F.M., Koopmans M., Beer M., Nitsche A. (2019) Proficiency testing of virus diagnostics based on bioinformatics analysis of simulated *in silico* high-throughput sequencing datasets. *Journal of Clinical Microbiology*, 57 (8): e00466-19, https://doi.org/10.1128/JCM.00466-19.

Bruni I., Galimberti A. ? Caridi L., Scaccabarozzi D., De Mattia F., Casiraghi M., Labra M. (2015) A DNA barcoding approach to identify plant species in multiflower honey. Food Chemistry, 170: 308-315, https://doi.org/10.1016/j.foodchem.2014.08.060.

Budowle B., Donnell N.D., Bielecka-Oder A., Colwell R.R., Corbett C.R., Fletcher J., Forsman M., Kadavy D.R., Markotic A., Morse S.A., Murch R.S., Sajantila A., Schmedes S.E., Ternus K.L., Turner S.D., Minot S. (2014) Validation of high throughput sequencing and microbial forensics applications. *Investigative Genetics*, 5: 9, http://www.investigativegenetics.com/content/5/1/9.

Buschmann T., Zhang R., Brash D.E., Bystrykh L.V. (2014). Enhancing the detection of barcoded reads in high throughput DNA sequencing data by controlling the false discovery rate. *BMC Bioinformatics*, 15: 264, doi:10.1186/1471-2105-15-264.

Bustin S.A., Benes V., Garson J.A., Hellemans J., Huggett J., Kubista M., Mueller R., Nolan T., Pfaffl M.W., Shipley G.L., Vandesompele J., Wittwer C.T. (2009) The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry*, 55 (4): 611-622, DOI: 10.1373/clinchem.2008.112797.

Candresse T., Filloux D., Muhire B., Julian C., Galzi S., Fort G., Bernardo P., Daugrois J.H., Fernandez E., Martin D.P., Varsani A. (2014) Appearances can be deceptive: revealing a hidden viral infection with deep sequencing in a plant quarantine context. *PLoS One*, 9 (7): e102945, doi: 10.1371/journal.pone.0102945.

Carew M.E., Coleman R.A., Hoffmann A.A. (2018) Can non-destructive DNA extraction of bulk invertebrate samples be used for metabarcoding? *PeerJ*, 6: e4980, https://doi.org/10.7717/peerj.4980.

Caruso P., Palomo J.L., Bertolini E., Alvarez B., López M.M., Biosca E.G. (2005) Seasonal variation of *Ralstonia solanacearum* biovar 2 populations in a Spanish river: recovery of stressed cells at low temperatures. *Applied and Environmental Microbiology*, 71 (1): 140-148, doi: 10.1128/AEM.71.1.140–148.2005.

Castagnone-Sereno P., Skantar A., Robertson L. (2011) Molecular tools for diagnosis. *In*: Jones J., Gheysen G., Fenoll C. (Editors), Genomics and Molecular Genetics of Plant-Nematode Interactions. *Springer, Dordrecht*, pp. 443-464.

Castro C.J. Ng T.F.F. (2017). U_{50}: A new metric for measuring assembly output based on non-overlapping, target-specific contigs. *Journal of Computational Biology*, 24 (11): 1071-1080, doi: 10.1089/cmb.2017.0013.

Champlot S., Berthelot C., Pruvost M., Bennett E.A., Grance T., Geigl E.-M. (2010) An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. *PLoS ONE*, 5 (9): e13042, doi: 10.1371/journal.pone.0013042.

Chandelier A., Hulin J., Martin S., Debode F., Massart S. (2020) Comparison of real-time PCR and metabarcoding methods as tools for the detection of airborne inoculum of forest fungal pathogens. *Phytopathology*, *In press*.

Christenhusz G.M., Devriendt K., Dierickx K. (2013) To tell or not to tell? A systematic review of ethical reflections on incidental findings arising in genetics contexts. *European Journal of Human Genetics*, 21: 248-255, doi:10.1038/ejhg.2012.130.

Claverie S., Bernardo P., Kraberger S., Hartnady P., Lefeuvre P., Lett J.-M., Galzi S., Filloux D., Harkins G.W., Varsani A., Martin D.P., Roumagnac P. (2018); From spatial metagenomics to molecular characterization of plant viruses: A geminivirus case study. *Advances in Virus Research*, 101: 55-83, https://doi.org/10.1016/bs.aivir.2018.02.003.

Cock P.J.A., Fields C.J., Goto N., Heuer M.L., Rice P.M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic acids Research*, 38 (6): 1767-1771, doi:10.1093/nar/gkp1137.

Constantin E.C., Haegeman A., Van Vaerenbergh J., Baeyen S., Van Malderghem C., Maes M. Cottyn B. (2017) Pathogenicity and virulence gene content of *Xanthomonas* strains infecting Araceae, formerly known as *Xanthomonas axonopodis* pv. *dieffenbachiae*. *Plant Pathology*, 66: 1539-1554, doi: 10.1111/ppa.12694.

Crous P.W., Hawksworth D.L., Wingfield M.J. (2015) Identifying and Naming Plant-Pathogenic Fungi: Past, Present, and Future. *Annual Review of Phytopathology*, 53: 247-267, https://doi.org/10.1146/annurev-phyto-080614-120245

Dal Molin A., Minio A., Griggio F., Delledonne M., INfantino A., Aragona M. (2018) The genome assembly of the fungal pathogen *Pyrenochaeta lycopersici* from single-molecule real-time sequencing sheds new light on its biological complexity. *PLoS ONE*, 13 (7): e0200217, https://doi.org/10.1371/journal.pone.0200217

Davis M.P.A., van Dongen S., Abreu-Goodger C., Bartonicek N., Enright A.J. (2013) Kraken: a set of tool for quality control and analysis of high-throughput sequence data. *Methods*, 63: 41-49, http://dx.doi.org/10.1016/j.ymeth.2013.06.027.

Davis J.J., Boisvert S., Brettin T., Kenyon R.W., Mao C., Olson R., Overbeek R., Santerre J., Shukla M., Wattam A.R., Will R., Xia F., Stevens R. (2016) Antimicrobial resistance prediction in PATRIC and RAST. *Scientific Reports*, 6: 27030, doi: 10.1038/srep27930.

Del Fabbro C., Scalabrin S., Morgante M., Giorgi F.M. (2013) An extensive of read trimming effects on Illumina NGS data analysis. *PLoS ONE*, 8 (12): e85024, https://doi.org/10.1371/journal.pone.0085024.

Denman S., Doonan J., Ransom-Jones E., Broberg M., Plummer S., Kirk S., Scarlett K., Griffiths A.R., Kaczmarek M., Forster J., Peace A., Golyshin P.N., Hassard F., Brown N., Kenny J.G., McDonald J.E.

(2018) Microbiome and infectivity studies reveal complex polyspecies tree disease in Acute oak decline. *The ISME Journal*, 12: 386-399, doi:10.1038/ismej.2017.170.

Dickins B., Rebolledo-Jaramillo B., Su M.S.-W., Paul I.M., Blankenberg D., Stoler N., Makova K.D., Nekrutenko A. (2014) Controlling for contamination in re-sequencing studies with a reproducible web-based phylogenetic approach. *Biotechniques*, 56 (3): 134-141, doi:10.2144/000114146.

Dormontt E.E., Van Dijk K.J., Bell K.L., Biffin E., Breed M.F., Byrne M., Caddy-Retalic S., Encinas-Viso F., Nevill P.G., Shapcott A., Young J.M. (2018) Advancing DNA barcoding and metabarcoding applications for plants requires systematic analysis of herbarium collections—an Australian perspective. *Frontiers in Ecology and Evolution*, 6: 134, doi: 10.3389/fevo.2018.00134.

Duan Y., Zhou L., Hall D.G., Li W., Doddapaneni H., Lin H., Liu L., Vahling C.M., Gabriel D.W., Willimas K.P., Dickerman A., Sun Y., Gottwald T. (2009) Complete genome sequence of Citrus Huanglongbing bacterium '*Candidatus* Liberibacter asiaticus' obtained through metagenomics. *Molecular Plant-Microbe Interactions*, 22 (8): 1011-1020, doi:10.1094/MPMI -22-8-1011.

Elbrecht V., Braukmann T.W.A., Ivanova N.V., Prosser S.W.J., Hajibabaei M., Wright M., Zakharov E.V., Hebert P.D.N., Steinke D. (2019) Validation of COI metabarcoding primers for terrestrial arthropods. *PeerJ*, 7: e7745, https://doi.org/10.7717/peerj.7745.

EPPO PM 3/21 (3) (2019) Post-entry quarantine for potato. *Bulletin OEPP/EPPO Bulletin*, 49 (3): 452-479, https://onlinelibrary.wiley.com/doi/epdf/10.1111/epp.12613.

EPPO PM 7/24 (4) (2019) *Xylella fastidiosa*. *Bulletin OEPP/EPPO Bulletin*, 49 (2): 175-227, https://onlinelibrary.wiley.com/doi/epdf/10.1111/epp.12575.

EPPO PM 7/76 (5) (2018) Use of EPPO diagnostic standards. *Bulletin OEPP/EPPO Bulletin*, 48 (3): 373-377, https://onlinelibrary.wiley.com/doi/epdf/10.1111/epp.12506.

EPPO PM 7/77 (3) (2019) Documentation and reporting on a diagnosis. *Bulletin OEPP/EPPO Bulletin*, 49 (3): 527-529, https://onlinelibrary.wiley.com/doi/epdf/10.1111/epp.12628.

EPPO PM 7/84 (2) (2018) Basic requirements for quality management in plant pest diagnostic laboratories. *Bulletin OEPP/EPPO Bulletin*, 48 (3): 378-386, https://onlinelibrary.wiley.com/doi/epdf/10.1111/epp.12507.

EPPO PM 7/98 (4) (2019) Specific requirements for laboratories preparing accreditation for a plant pest diagnostic activity. *Bulletin OEPP/EPPO Bulletin*, 49 (3): 530-563, https://onlinelibrary.wiley.com/doi/epdf/10.1111/epp.12629.

EPPO PM7/129 (1) (2016) DNA barcoding as an identification tool for a number of regulated pests. *Bulletin OEPP/EPPO Bulletin*, 46 (3): 501-537, https://onlinelibrary.wiley.com/doi/epdf/10.1111/epp.12344.

EPPO PM 7/130 (1) (2016) Guidelines on the authorization of laboratories to perform diagnostic activities for regulated pests. *Bulletin OEPP/EPPO Bulletin*, 46 (3): 538-540, https://onlinelibrary.wiley.com/doi/epdf/10.1111/epp.12343.

FAO (2019) Preparing to use high-throughput sequencing (HTS) technologies as a diagnostic tool for phytosanitary purposes. Commission on Phytosanitary Measures Recommendation, No 8, https://www.ippc.int/en/publications/87199/.

Fox A. (2020) Reconsidering causal association in plant virology. *Plant Pathology*, https://bsppjournals.onlinelibrary.wiley.com/doi/epdf/10.1111/ppa.13199.

Fox A., Fowkes A.R., Skelton A., Harju V., Buxton-Kirk A., Kelly M., Forde S.M.D., Pufal H., Conyers C., Ward R., Weekes R. (2019) Using high-throughput sequencing in support of a plant health outbreak reveals novel viruses in *Ullucus tuberosus* (Basellaceae). *Plant Pathology*, 68 (3): 576-587, https://doi.org/10.1111/ppa.12962.

Gaafar Y.Z.A., Ziebell H. (2020) Comparative study on three viral enrichment approaches based on RNA extraction for plant virus/viroid detection using high-throughput sequencing. *PLoS ONE*, 15 (8): e0237951, https://doi.org/10.1371/journal.pone.0237951.

Gaafar Y., Lüddecke P., Heidler C., Hartrick J., Sieg-Müller A., Hübert C., Wichura A., Ziebell H. (2019a) First report of Southern tomato virus in German tomatoes. *BSPP New Disease Reports*, 40: 1, https://www.ndrs.org.uk/article.php?id=040001.

Gaafar Y., Sieg-Müller A., Lüddecke P., Hartrick J., Seide Y., Müller J., Maaß C., Schuhmann S., Richert-Pöggeler K.R., Blouin A.G., Massart S., Ziebell H. (2019b) First report on natural infection of beetroot with Beet soil-borne virus. *BSPP New Disease Reports*, 40: 5, https://www.ndrs.org.uk/article.php?id=040005.

Galan M., Razzauti M., Bard E., Bernard M., Brouat C., Charbonnel N., Dehne-Garcia A., Loiseau A., Tatard C., Tamisier L., Vayssier-Taussat M., Vignes H. (2016) 16S rRNA amplicon sequencing for epidemiological surveys of bacteria in wildlife. *Clinical Science and Epidemiology*, 1 (4): e00032-16, doi:10.1128/mSystems.00032-16.

Gargis A.S., Kalman L., Bick D.P., da Silva C., Dimmock D.P., Funke B.H., Gowrisankar S., Hegde M.R., Kulkarni E., Mason C.E., Nagarajan R., Voelkerding K.V., Worthey D.A., Aziz N., Barnes J., Bennett S.F., Bisht H., Church F.M., Dimitrova Z., Gargis S.R., Hafez N., Hambuch T., Hyland F.C.L., Lunna R.A., MacCannell D., Mann T ., McCluskey M.R., McDaniel T.K., Ganova-Raeva L.M., Rehm H.L., Reid J ., Campo D.S., Resnick R.B., Ridge P.G., Salit M.L., Skums P., Wong L.-J.C., Zehnbauer B.A., Zook J.M., Lubin I.M. (2015) Good laboratory practice for clinical next-generation sequencing informatics pipelines. *National Biotechnologies*, 33 (7): 689-693, doi:10.1038/nbt.3237.

Giangacomo C., Mohseni M., Kovar L., Wallace J.G. (2020) Comparing DNA extraction and 16s amplification methods for plant-associated bacterial communities. *bioRxiv*, doi: https://doi.org/10.1101/2020.07.23.217901.

Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnirke A., Rhind N., di Palma F., Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N., Regev A. (2011) Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nature Biotechnology*, 29: 644-652, doi: 10.1038/nbt.1883.

Hadidi A., Flores R., Candresse T., Barba M. (2016) Next-generation sequencing and genome editing in plant virology. *Frontiers in Microbiology*, 7: 1325, doi: 10.3389/fmicb.2016.01325.

Hamelin R.C., Roe A.D. (2020) Genomic biosurveillance of forest invasive alien enemies: A story written in code. *Evolutionary Applications*, 13 (1): 95-115, doi: 10.1111/eva.12853.

Hanssen I.M., Paeleman A., Candewoestijne E., Van Bergen L., Bragard C., Lievens B., Vanachter A.C.R.C., Thomma B.P.H.J. (2009) *Pepino mosaic virus* isolates and differential symptomatology in tomato. *Plant Pathology*, 58: 450-460, doi: 10.1111/j.1365-3059.2008.02018.x.

Hardwick S.A., Chen W.Y., Wong T., Kanakamedala B.S., Deveson I.W., Ongley S.E., Santini N.S., Marcellin E., Smith M.A., Nielsen L.K., Lovelock C.E., Neilan B.A., Mercer T.R. (2018) Synthetic microbe communities provide internal reference standards for metagenome sequencing and analysis. *Nature Communications*, 9: 3096, doi: 10.1038/s41467-018-05555-0.

Hébrant A., Froyen G., Maes B., Salgado R., Le Mercier M., D'Haene N., De Keersmaecker S., Claes K., Van der Meulen J., Aftimos P., Van Houdt J., Cuppens K., Vanneste K., Dequeker E., Van Dooren S., Van Huysse J., Nollet F., van Laere S., Denys B., Ghislain V., Van Campenhout C., Van den Bulcke M. (2018) The Belgian next generation sequencing guidelines for haematological and solid tumours. *The Belgian Journal of Medical Oncology*, 11 (2): 56-67.

Herren G.L., Habraken J., Waeyenberge L., Haegeman A., Viaene N., Cougnon M., Reheul D., Steel H., Bert W. (2020) Effects of synthetic fertilizer and farm compost on soil nematode community in long-term crop rotation plots: a morphological and metabarcoding approach. *PLoS ONE*, 15 (3): e0230153, https://doi.org/10.1371/journal.pone.0230153

Hohn T., Richert-Pöggeler K.R., Staginnus C., Harper G., Schwarzacher T., Teo C.H., Teycheney P.Y., Iskra-Caruana M.L., Hull R. (2008) Evolution of integrated plant viruses. *In*: Plant Virus Evolution, Roossinck M.J. (Ed.), Springer-Verlag, Berlin, Germany, pp. 53-81.

ISO 9000 (2015) Quality management systems. International organization for standardization, Geneva, Switzerland.

ISO 17025 (2017) General requirements for the competence of testing and calibration laboratories. International organization for standardization, Geneva, Switzerland.

ISPM 5 (2020) Glossary of phytosanitary terms, IPPC / FAO https://www.ippc.int/en/publications/622/.

Jennings L.J., Arcila M.E., Corless C., Kamel-Reid S., Lubin I.M., Pfeifer J., Temple-Smolkin R.L., Voelerding K.V., Nikiforova M.N. (2017) Guidelines for validation of next-generation sequencing – based oncology panels, A joint consensus recommendation of the Association for Molecular Pathology and College of American Pathologists. *The Journal of Molecular Diagnostics*, 19 (3): 341-365, https://doi.org/10.1016/j.jmoldx.2017.01.011.

Johne R., Müller H., Rector A., van Ranst M., Stevens H. (2009) Rolling-circle amplification of viral DNA genomes using phi29 polymerase. *Trends in Microbiology*, 17 (5): 205-211, https://doi.org/10.1016/j.tim.2009.02.004.

Jung H., Winefield C., Bombarely A., Prentis P., Waterhouse P. (2019) Tools and strategies for long-read sequencing and *de novo* assembly of plant genomes. *Trends in Plant Sciences*, 24 (8): 700-724, https://www.cell.com/trends/plant-science/pdf/S1360-1385(19)30121-9.pdf.

Karczewski K., Riss W., Meyer E.I. (2017) Comparison of DNA-fingerprinting (T-RFLP) and high-throughput sequencing (HTS) to assess the diversity and composition of microbial communities in groundwater ecosystems. *Limnologica*, 67: 45-53, http://dx.doi.org/10.1016/j.limno.2017.10.001.

Katsiani A., Maliogka V.I., Katis N., Svanella-Dumas L., Olmos A., Ruiz-García A.B., Marais A., Faure C., Theil S., Lotos L., Candresse T. (2018) High-throughput sequencing reveals further diversity of *Little Cherry Virus 1* with Implications for Diagnostics. *Viruses*, 10: 385, doi: 10.3390/v10070385.

Kesanakurti P., Belton M., Saeed H., Rast H., Boyes I., Rott M. (2016) Screening for plant viruses by next generation sequencing using a modified double strand RNA extraction protocol with an

internal amplification control. *Journal of Virological Methods*, 236: 35-50, http://dx.doi.org/10.1016/j.jviromet.2016.07.001.

Kikuchi T., Cotton J.A., Dalzell J.J., Hasegawa K., Kanzaki N., McVeigh P., Takanashi T., Tsai I.J., Assefa S.A., Cock P.J.A., Otto T.D., Hunt M., Reid A.J., Sanchez-Flores A., Tsuchihara K., Yokoi T., Larsson M.C., Miwa J., Maule A.G., Sahashi N., Jones J.T., Berriman M. (2011) Genomic insights into the origin of parasitism on the emerging plant pathogen *Bursaphelenchus xylophilus*. *PLoS Pathogens*, 7 (9): e1002219, doi: 10.1371/journal.ppat.1002219.

Kircher M., Sawyer S., Meyer M. (2011) Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research*, 40 (1): e3, doi: 10.1093/nar/gkr771.

Kreuze J.F., Perez A., Gargurevich M.G., Cuellar W.J. (2020) Badnaviruses of sweet potato: symptomless coinhabitants on a global scale. *Frontiers in Plant Science*, 11: 313, doi: 10.3389/fpls.2020.00313.

Kwon S., Lee B., Yoon S. (2014) CASPER: context-aware scheme for paired-end reads from high-throughput amplicon sequencing. *BMC Bioinformatics*, 15 (suppl. 9): S10, http://www.biomedcentral.com/1471-2105/15/S9/S10.

Lamichhane J.R., Venturi V. (2015) Synergisms between microbial pathogens in plant disease complexes: a growing trend. *Frontiers in plant science*, 6: 385, doi: 10.3389/fpls.2015.00385.

Laehnemann D., Borkhardt A., McHardy A.C. (2016) Denoising DNA deep sequencing data—high-throughput sequencing errors and their correction. *Briefings in bioinformatics*, 17 (1): 154-179, https://doi.org/10.1093/bib/bbv029.

Lambert C., Braxton C., Charlebois R.L., Deyati A., Duncan P., La Neve F., Malicki H.D., Ribrioux S., Rozelle D.K., Michaels B., Sun W., Yang Z., Khan A.S. (2018) Considerations for optimisation of high-throughput sequencing bioinformatic pipelines for virus detection. *Viruses*, 10: 528, doi:10.3390/v10100528.

Lasken R.S., Stockwell T.B. (2007) Mechanism of chimera formation during the Multiple Displacement Amplification reaction. *BMC biotechnology*, 7: 19, doi:10.1186/1472-6750-7-19.

Lee Marzano S.-Y., Domier L.L. (2016) Novel mycoviruses discovered from metatranscriptomics survey of soybean phyllosphere phytobiomes. *Virus Research*, 213: 332-342, http://dx.doi.org/10.1016/j.virusres.2015.11.002.

Lefebvre M., Theil S., Ma Y., Candresse T. (2019) The VirAnnot pipeline: a resource for automated viral diversity estimation and operational taxonomy units (OUT) assignation for virome sequencing data. *Phytobiomes Journal*, https://doi.org/10.1094/PBIOMES-07-19-0037-A.

Leifert W.R., Glatz R.V., Siddiqui M.S., Collins S.R., Taylor P.W., Fenech M. (2013). Development of a test to detect and quantify irradiation damage in fruit flies. *Final Report for Horticulture Australia Ltd.*, Project VG09160, https://ausveg.com.au/app/data/technical-insights/docs/130056_VG09160.pdf.

Locati M.D., Terpstra I., de Leeuw W.C., Kuzak M., Rauwerda H., Ensink W.A., van Leeuwen S., Nehrdich U., Spaink H.P., Jonker M.J., Breit T.M., Dekker R.J. (2015) Improving small RNA-seq by using a synthetic spike-in set for size-range quality control together with a set for data normalization. *Nucleic Acids Research*, 43 (14): e89, doi: 10.1093/nar/gkv303.

Lu N., Li J., Bi C., Guo J., Tao Y., Luan K., Tu J., Lu Z. (2019) ChimeraMiner: An Improved Chimeric Read Detection Pipeline and Its Application in Single Cell Sequencing. *International journal of molecular sciences*, 20 (8): 1953, doi: 10.3390/ijms20081953.

Luscombe N.M., Greenbaum D., Gerstein M. (2001), What is bioinformatics? A proposed definition and overview of the field. *Methods of Information in Medicine*, 40: 346-358.

MacConaill L.E., Burns R.T., Nag A., Coleman H.A., Slevin M.K., Giorda K., Light M., Lai K., Jarosz M., McNeill M.S., Ducar M.D., Meyerson M., Thorner A.R. (2018) Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. *BMC Genomics*, 19: 30, doi: 10.1186/s12864-017-4428-5.

Maclot F., Candresse T., Filloux D., Malmstrom C.M., Roumagnac P., van der Vlugt R., Massart S. (2020) Illuminating an ecological blackbox: Using High Throughput Sequencing to characterize the plant virome across scales. *Frontiers in Microbiology*, 11: 578064, doi: 10.3389/fmicb.2020.578064.

Mahé F., Rognes T., Quince C., de Vargas C., Dunthorn M. (2015) Swarm v2: highly-scalable and high-resolution amplicon clustering. *PeerJ*, 3: e1420, doi: 10.7717/peerj.1420.

Malapi-Wight M., Salgado-Salazar C., Demers J., Clement D.L., Rane K., Crouch J.A. (2016). Sarcococca Blight: use of whole genome sequencing for fungal plant disease diagnosis. *Plant Disease*, 100 (6): 1093–1100.

Maliogka V.I., Minafra A., Saldarelli P., Ruiz-García A.B., Glasa M., Katis N., Olmos A. (2018) Recent advances on detection and characterization of fruit tree viruses using high-throughput sequencing technologies. *Viruses*, 10: 436, doi:10.3390/v10080436.

Maljkovic Berry I., Melendrez M.C., Bishop-Lilly K.A., Rutvisuttinunt W., Pollett S., Talundzic E., Morton L., Jarman R.G. (2020) Next generation sequencing and bioinformatic methodologies for infectious disease research and public health: approaches, applications, and considerations for development of laboratory capacity. *The Journal of Infectious Diseases*, 221 (S3): S292-307, doi: 10.1093/infdis/jiz286.

Maree H.J., Fox A., Al Rwahnih M., Boonham N., Candresse T. (2018) Application of HTS for routine plant virus diagnostics: state of the art and challenges. *Frontiers in Plant Science*, 9: 1082, doi: 10.3389/fpls.2018.01082.

Marquina D., Ronquist F., Łukasik P. (2020) The effect of ethanol concentration on the preservation of insects for biodiversity studies. *bioRxiv*, 2020, https://doi.org/10.1101/2020.03.05.978288.

Martin R.R., James D., Lévesque C.A. (2000) Impacts of molecular diagnostic technologies on plant disease management. *Annual Review of Phytopathology*, 38: 207-239.

Martin F.N., Abad Z.G., Balci Y., Ivors K. (2012) *Phytophthora*: reviewing our progress, identifying our needs. *Plant Disease*, 96 (8): 1080-1103, doi: 10.1094/PDIS-12-11-1036-FE.

Massart S., Olmos A., Jijaki H., Candresse T. (2014) Current impact and future directions of high throughput sequencing in plant virus diagnostic. *Virus Research*, 188 (8): 90-96, https://doi.org/10.1016/j.virusres.2014.03.029.

Massart S., Candresse T., Gil J., Lacomme C., Predajna L., Ravnikar M., Reynard J.-S., Rumbou A., Saldarelli P., Škorić D., Vainio E.J., Valkonen J.P.T., Vanderschuren H., Varveri C., Wetzel T. (2017) A framework for the evaluation of biosecurity, commercial, regulatory, and scientific impacts of

plant viruses and viroids identified by NGS technologies. *Frontiers in Microbiology*, 8: 45, doi: 10.3389/fmicb.2017.00045.

Massart S., Chiumenti M., De Jonghe K., Glover R., Haegeman A., Koloniuk I., Komínek P., Kreuze J., Kutnjak D., Lotos L., Maclot F., Maliogka V., Maree H.J., Olivier T., Olmos A., Pooggin M.M., Reynard J.-S., Ruiz-García A.B., Safarova D., Schneeberger P.H.H., Sela N., Turco S., Vainio E.J., Varallyay E., Verdin E., Westenberg M., Brostaux Y., Candresse T. (2019) Virus detection by high-throughput sequencing of small RNAs: large-scale performance testing of sequence analysis strategies. P*hytopathology*, 109: 488-497, https://apsjournals.apsnet.org/doi/10.1094/PHYTO-02-18-0067-R.

Matthijs G., Souche E., Alders M., Corveleyn A., Eck S., Feenstra I., Race V., Sistermans E., Sturm M., Weiss M., Yntema H., Bakker E., Scheffer H., Bauer P. (2016) Guidelines for diagnostic next-generation sequencing. *European Journal of Human Genetics*, 24: 2-5, doi:10.1038/ejhg.2015.226.

Mbareche H., Veillette M., Bilodeau G., Duchaine C. (2020) Comparison of the performance of ITS1 and ITS2 as barcodes in amplicon-based sequencing of bioaerosols, *PeerJ*, 8: e8523, https://doi.org/10.7717/peerj.8523.

McGuire A.L., McCullough L.B., Evans J.P. (2013) The indispensable role of professional judgment in genomic medicine. *Journal of the American Medical Association*, 309 (14): 1465-1466, doi:10.1001/jama.2013.1438.

McInerney P., Adams P., Hadi M.Z. (2014) Error rate comparison during polymerase chain reaction by DNA polymerase. *Molecular Biology International*, 2014: 287430, http://dx.doi.org/10.1155/2014/287430.

Mehle N., Dreo T., Jeffries C., Ravnikar M. (2014) Descriptive assessment of uncertainties of qualitative real-time PCR for detection of plant pathogens and quality performance monitoring. *Bulletin OEPP/EPPO Bulletin*, 44 (3): 502-509, doi: 10.1111/epp.12166.

Mehle N., Gutiérrez-Aguirre I., Kutnjak D., Ravnikar M. (2018) Water-mediated transmission of plant, animal, and human viruses. *Advances in Virus Research*, 101: 85-128, https://doi.org/10.1016/bs.aivir.2018.02.004.

Mirghasempour S.A., Huang S., Studholme D.J., Brady C.L. (2020) A grain rot of rice in Iran caused by a *Xanthomonas* strain closely related to *X. sacchari*. *Plant Disease,* 104 (6):1581-1583, doi: 10.1094/PDIS-01-20-0179-SC.

Mullis K., Faloona F., Scharf S., Saiki R., Horn G., Erlich H. (1986) Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology*, vol. LI., 0-87969-052-6/8.

Murugan S., Kumarasinghe L. (2018) An operational risk assessment framework for a plant health diagnostic laboratory. *Bulletin OEPP/EPPO Bulletin*, 48 (2): 225-233, doi: 10.1111/epp.12477.

Nicolaisen M., West J.S., Sapkota R., Canning G.G.M., Schoen C., Justesen A.F. (2017) Fungal communities including plant pathogens in near surface air are similar across Northwestern Europe. *Frontiers in Microbiology*, 8: 1729, https://doi.org/10.3389/fmicb.2017.01729.

Nielsen M., Gilbert M.T.P., Pape T., Bohmann K. (2019). A simplified DNA extraction protocol for unsorted bulk arthropod samples that maintains exoskeletal integrity. *Environmental DNA*, 1(2): 144-154, https://doi.org/10.1002/edn3.16.

Nilsson R.H., Anslan S., Bahram M., Wurzbacher C., Baldrian P., Tedersoo L. (2019) Mycobiome diversity: high-throughput sequencing and identification of fungi. *Microbiology*, 17: 95-109, https://doi.org/10.1038/ s41579-018-0116-y.

Núñez A., Amo de Paz G., Ferencova Z., Rastrojo A., Guantes R., García A.M., Alcami A., Montserrat Gutiérrez-Bustillo A., Moreno D.A. (2017) Validation of the high-type spore trap for simultaneous monitoring of prokaryotic and eukaryotic biodiversity in urban air samples by NGS. *Applied and Environmental Microbiology*, doi: 10.1128/AEM.00472-17.

OIE (2018) Normes pour le séquençage à haut debit, la bioinformatique et la génomique computationnelle. Volume 1, Chapitre 1.1.7. https://www.oie.int/fileadmin/Home/fr/Health_standards/tahm/F_summry.htm.

Olmos A., Boonham N., Candresse T., Gentit P., Giovani B., Kutnjak D., Liefting L., Maree H.J., Minafra A., Moreira A., Nakhla M.K., Petter F., Ravnikar M., Rodoni B., Roenhorst J.W., Rott M., Ruiz-García A.B., Santala J., Stancanelli G., van der Vlugt R., Varveri C., Westenberg M., Wetzel T., Ziebell H., Massart S. (2018) High-throughput sequencing technologies for plant pest diagnosis : challenges and opportunities. *Bulletin OEPP/EPPO Bulletin*, 48 (2): 219-224, doi: 10.1111/epp.12472.

Orlando V., Grove I.G., Edwards S.G., Prior T., Roberts D., Neilson R., Back M. (2020) Root-lesion nematodes of potato: Current status of diagnostics, pathogenicity and management. *Plant Pathology*, 69 (3): 405-417, https://doi.org/10.1111/ppa.13144.

Ovaskainen O., Abrego N., Somervuo P., Palorinne I., Hardwick B., Pitkänen J.-M., Andrew N.R., Niklaus P.A., Schmidt N.M., Seibold S., Vogt J., Zakharov E.V., Hebert P.D.N., Roslin T., Ivanova N.V. (2020) Monitoring fungal communities with the global spore sampling project. *Frontiers in Ecology and Evolution*,7: 511, https://doi.org/10.3389/fevo.2019.00511.

Owati A., Agindotan B., Burrows M. (2019) Development and application of real-time and conventional SSR-PCR assays for rapid and sensitive detection of *Didymella pisi* associated with Ascochyta blight of dry pea. *Plant Disease*, 103: 11, https://doi.org/10.1094/PDIS-02-19-0381-RE.

Palmano S., Saccardo F., Martini M., Ermacora P., Scortichini M., Abbà S., Marzachì C., Loi N., Firrao G. (2012) Insights into phytoplasma biology through next generation sequencing. *Journal of Plant Pathology*, 94 (4): S4.50.

Palmer J.M., Jusino M.A., Banik M.T., Lindner D.L. (2018) Non-biological synthetic spike-in controls and the AMPtk software pipeline improve mycobiome data. *PeerJ*, 6: e4925, doi: 10.7717/peerj.4925.

Parks D.H., Imelfort M., Skennerton C.T., Hugenholtz P., Tyson G.W. (2014) Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, 25: 1043-1055, https://ecogenomics.github.io/CheckM/.

Pecman A., Kutnjak D., Gutiérrez-Aguirre I., Adams I., Fox A., Boonham N., Ravnikar M. (2017) Detection and discovery of plant viruses and viroids: comparions of two approaches. *Frontiers in Microbiology*, 8: 1998, doi: 10.3389/fmicb.2017.01998.

Pervaiz T., Lotfi A., Haider M.S., Haifang J., Fang J. (2017) High throughput sequencing advances and future challenges. *Journal of Plant Biochemistry & Physiology*, 5 (2): 1000188, doi: 10.4172/2329-9029.1000188.

Piper A.M., Batovska J., Cogan N.O.I., Weiss J., Cunningham J.P., Rodoni B.C., Blacket M.J. (2019) Prospects and challenges if implementing DNA metabarcoding for high-throughput insect surveillance. *GigaScience*, 8: 1-22, doi: 10.1093/gigascience/giz092.

Prezelj N., Nikolić P., Gruden K., Ravnikar M., Dermastica M. (2013) Spatiotemporal distribution of flavescence dorée phytoplasma in grapevine. *Plant Pathology*, 62 (4): 760–766, doi: 10.1111/j.1365-3059.2012.02693.x.

Pritchard L., Glover R.H., Humphris S., Elphinstone J.G., Toth I.K. (2016) Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Analytical Methods*, 8: 12-24, doi: 10.1039/c5ay02550h.

Quail M.A., Smith M., Jackson D., Leonard S., Skelly T., Swerdlow H.P., Gu Y., Ellis P. (2014) SASI-Seq: sample assurance Spike-Ins, and highly differentiating 384 barcoding for Illumina sequencing. *BMC Genomics*, 15: 110, http://www.biomedcentral.com/1471-2164/15/110.

Quince C., Lanzen A., Davenport R.J., Turnbaugh P.J. (2011) Removing noise from pyrosequenced amplicons. *BMC bioinformatics*, 12 (1): 38, https://doi.org/10.1371/journal.pone.0139857.

Ratnasingham S., Hebert P.D.N. (2007) BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Molecular Ecology Notes*, 7: 355–364.

Rehm H.L., Bale S.J., Bayrak-Toydemir P., Berg J.S., Brown K.K., Deignan J.L., Friez M.J., Funke B.H., Hedge M.R., Lyon Y. (2013) ACMG clinical laboratory standards for next-generation sequencing. *Genetics in Medicine*, 15 (9): 733-747, doi:10.1038/gim.2013.92.

Reuter J.A., Spacek D., Snyder M.P. (2015) High-throughput sequencing technologies. *Molecular Cell*, 58 (54): 586-597, doi:10.1016/j.molcel.2015.05.004.

Riddell C.E., Frederickson-Matika D., Armstrong A.C., Elliot M., Forster J., Hedley P.E., Morris J., Thorpe P., Cooke D.E., Pritchard L., Sharp P.M. (2019) Metabarcoding reveals a high diversity of woody host-associated *Phytophthora* spp. in soils at public gardens and amenity woodlands in Britain. *PeerJ*, 7: e6931, doi: 10.7717/peerj.6931.

Ritter C.D., Häggqvist S., Karlsson D., Sääksjärvi I.E., Muasya A.M., Nilsson R.H., Antonelli A. (2019). Biodiversity assessments in the 21st century: the potential of insect traps to complement environmental samples for estimating eukaryotic and prokaryotic diversity using high-throughput DNA metabarcoding. *Genome*, 62 (3): 147-159, dx.doi.org/10.1139/gen-2018-0096.

Robinson C.V., Porter T.M., Wright M.T.G. Hajibabaei M. (2020) Propylene glycol-based antifreeze as an effective preservative for DNA metabarcoding of benthic arthropods. *bioRxiv*, 2020, https://doi.org/10.1101/2020.02.28.970475.

Roenhorst J.W., de Krom C., Fox A., Mehle N., Ravnikar M., Werkman A.W. (2018) Ensuring validation in diagnostic testing is fit for purpose: a view from the plant virology laboratory. *Bulletin OEPP/EPPO Bulletin*, 48 (1): 105-115, doi: 10.1111/epp.12445.

Rolland M., Villemot J., Marais A., Theil S., Faure C., Cadot V., Valade R., Vitry C., Rabenstein F., Candresse T. (2017) Classical and next generation sequencing approaches unravel Bymovirus diversity in Barley crops in France. *PLOS One*, 12 (11): e0188495, https://doi.org/10.1371/journal.pone.0188495.

Roossinck M.J. (2015) Plants, viruses and the environment: ecology and mutualism. *Virology*, 479-480: 271-277, http://dx.doi.org/10.1016/j.virol.2015.03.041.

Rosseel T., Pardon B., De Clercq K., Ozhelvaci O., Van Borm S. (2014) False-positive results in metagenomics virus discovery: a strong case for follow-up diagnosis. *Transboundary and Emerging Diseases*, 61: 293-299, doi:10.1111/tbed.12251.

Rott M., Xiang Y., Boyes I., Belton M., Saeed H., Kesanakurti P., Hayes S., Lawrence T., Birch C., Bhagwat B., Rast H. (2017) Application of next generation sequencing for diagnostic testing of tree fruit viruses and viroids. *Plant Disease*, 101: 1489-1499, http://dx.doi.org/10.1094/PDIS-03-17-0306-RE.

Roy S., Coldren C., Karunamurthy A., Kip N.S., Klee E.W., Lincoln S.E., Leon A., Pullambhatla M., Temple-Smolkin R.L., Voelkerding K.V., Wang C., Carter A.B. (2018) Standards and guidelines for validating next-generation sequencing bioinformatic pipelines, a joint recommendation of the Association for Molecular Pathology and the College of American Pathologists. *The Journal of Molecular Diagnostics*, 20 (1): 4-27, https://doi.org/10.1016/j.jmoldx.2017.11.003.

Sahlin K., Chikhi R., Arvestad L. (2016) Assembly scaffolding with PE-contaminated mate-pair libraries. *Bioinformatics*, 32 (13): 1925-1932, doi: 10.1093/bioinformatics/btw064.

Salter S.J., Cox M.J., Turek E.M., Calus S.T., Cookson W.O., Moffatt M.F., Turner P., Parkhill J., Loman N.J., Walker A.W. (2014) Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology*, 12: 87, http://www.biomedcentral.com/1741-7007/12/87.

Santala J., Valkonen J.P.T. (2018) Sensitivity of small RNA-based detection of plant viruses. *Frontiers in Microbiology*, 9: 939, doi: 10.3389/fmicb.2018.00939.

Scibetta S., Schena L., Abdelfattah A., Pangallo S., Cacciola S. (2018) Selection and experimental evaluation of universal primers to study the fungal microbiome of higher plants. *Phytobiomes*, 2 (4): 225-236.

Sekse C., Holst-Jensen A., Dobrindt U., Johannessen G.S., Li W., Spilsberg B., Shi J. (2017) High throughput sequencing for detection of foodborne pathogens. *Frontiers in Microbiology*, 8: 2029, doi: 10.3389/fmicb.2017.02029.

Seppey M., Manni M., Zdobnov E.M. (2019) BUSCO: assessing genome assembly and annotation completeness. *In*: Kollmar (eds.) *Gene Prediction. Methods in Molecular Biology*, Humana, New York, vol. 1962, 227-245, https://doi.org/10.1007/978-1-4939-9173-0_14.

Sharma S. Chatterjee S., Datta S., Prasad R., Dubey D., Prasad R.K., Vairale M.G. (2017) Bacteriophages and its applications: an overview. *Folia Microbiology*, 62: 17-55, doi: 10.1007/s12223-016-0471-x.

Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M., Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S., Barros M.H., Bonaccorsi E.D., Bordin S., Bové J.M., Briones M.R.S., Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H., Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M., Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H., Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A., Fraga J.S., França S.C., Franco M.C., Frohme M., Furlan L.R., Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A., Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P., Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C., Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A., Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L., Marques M.V., Martins E.A.L.,

Martins E.M.F., Matsukuma A.Y., Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B., Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S., Nhani Jr A., Nobrega F.G., Nunes L.R., Oliveira M.A., de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A., Peixoto B.R., Pereira G.A.G., Pereira Jr H.A., Pesquero J.B., Quaggio R.B., Roberto P.G., Rodrigues V., de M. Rosa A.J., de Rosa Jr V.E., de Sá R.G., Santelli R.V., Sawasaki H.E., da Silva A.C.R., da Silva A.M., da Silva F.R., Silva Jr W.A., da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A., de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H., Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L., Zago M.A., Zatz M., Meidanis J., Setubal J. C. (2000) The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature*, 406: 151-159, https://doi.org/10.1038/35018003.Strom S.P. (2016) Current practices and guidelines for clinical next-generation sequencing oncology testing. *Cancer Biology & Medicine*, 13: 1, doi: 10.28092/j.issn.2095-3941.2016.0004.

Taylor G.S., Martoni F. (2019) Case of mistaken identity: resolving the taxonomy between *Trioza eugeniae* Froggatt and *T. adventicia* Tuthill (Psylloidea: Triozidae). *Bulletin of Entomological Research*, 110 (3): 340-351, https://doi.org/10.1017/S0007485319000695.

Thomsen P.F., Sigsgaard E.E. (2019) Environmental DNA metabarcoding of wild flowers reveals diverse communities of terrestrial arthropods. *Ecology and Evolution*, 9: 1665-1679, https://doi.org/10.1002/ece3.4809.

Tourlousse D.M., Ohashi A., Sekiguchi Y. (2018) Sample tracking in microbiome community profiling assays using synthetic 16S rRNA gene spike-in controls. *Scientific Reports*, 8: 9095, doi:10.1038/s41598-018-27314-3.

Tremblay E.D., Duceppe M.-O., Bérubé J.A., Kimoto T., Lemieux C., Bilodeau G.J. (2018) Screening for exotic forest pathogens to increase survey capacity using metagenomics. *Phytopathology*, 108 (12): 1509-1521, https://doi.org/10.1094/PHYTO-02-18-0028-R.

Tremblay E.D., Duceppe M.-O., Thurston G.B., Gagnon M.-C., Côté M.-J., Bilodeau G.J. (2019) High-resolution biomonitoring of plant pathogens and plant species using metabarcoding of pollen pellet contents collected from a honey bee hive. *Environmental DNA*, 1: 155-175, https://doi.org/10.1002/edn3.17.

Trimme R.E., Rand H., Shumway M., Trees E.K., Simmons M., Agarwala R., Davis S., Tillman G.E., Defibaugh-Chavez S., Carleton H.A., Klimke W.A., Katz L.S. (2015) Benchmark datasets for phylogenomic pipeline validation, applications for foodborne pathogen surveillance. *PeerJ*, 5: e3893, doi: 10.7717/peerj.3893.

VALITEST deliverable 3.3 (2020) Guidelines and standard operating procedure (SOP) finalised for the production of the reference materials. https://www.valitest.eu/media/files/deliverables/WP6/D3.3_Guidelines_and_Standard_Operating_Procedures_(SOP)_finalised_for_the_production_of_the_reference_materials.pdf.

van der Valk T., Vezzi F., Ormestad M., Dalén L., Guschanski K. (2018) Index hopping on the Illumina HiseqX platform and its consequences for ancient DNA studies. *Molecular Ecology Resources*, 00: 1-11, doi: 10.1111/1755-0998.13009.

Van El C.G., Cornel M.C., Borry P., Hastings R.J., Fellmann F., Hodgson S.V., Howard H.C., Cambon-Thomsen A., Knoppers B.M., Meijers-Heijboer H., Scheffer H., Tranebjaerg L., Dondorp W., de Wert G.M.W.R. (2013) Whole-genome sequencing in health care, Recommendations of the European Society of Human Genetics. *European Journal of Human Genetics*, 21: S1-S5, doi:10.1038/ejhg.2013.46.

van Opijnen T., Camilli A. (2013) Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nature Review, Microbiology*, 11 (7): 435–442. doi: 10.1038/nrmicro3033.

Visser M., Bester R., Burger J.T. Maree H.J. (2016) Next-generation sequencing for virus detection: covering all the bases. *Virology Journal*, 13: 85, doi: 10.1186/s12985-016-0539-x.

Waeyenberge L., de Sutter N., Viaene N., Haegeman A. (2019). New insights into nematode DNA-metabarcoding as revealed by the characterization of artificial and spiked nematode communities. *Diversity*, 11: 52, doi:10.3390/d11040052.

Weiss M.M., Van der Zwaag B., Jongbloed J.D.H., Vogel M.J., Bruggenwirth H.T., Lekanne Derez R.H., Mook O., Ruivenkamp C.A.L., van der Stoep N. (2013) Sequencing applications in genome diagnostics: A national collaborative study of Dutch genome diagnostic laboratories. *Human Mutation*, 34 (10): 1313-1321, doi:10.1002/humu.22368.

Wesolowska-Andersen A., Iain Bahl M., Kristiansen K., Sicheritz-Pontén T., Gupta R., Rask Licht T. (2014) Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomics analysis. *Microbiome*, 2: 19, http://www.microbiomejournal.com/content/2/1/19.

Whitehurst L.E., Cunard C.E., Reed J.N., Worthy S.J., Marsico T.D., Lucardi R.D., Burgess K.S. (2020) Preliminary application of DNA barcoding toward the detection of viable plant propagules at an initial, international point-of-entry in Georgia, USA. *Biological Invasions*, 1-22, https://doi.org/10.1007/s10530-020-02204-w.

Wilcox T.M., Zarn K.E., Piggott M.P., Young M.K., McKelvey K.S., Schwartz M.K. (2018) Capture enrichment of aquatic environmental DNA: A first proof of concept. *Molecular Ecology Resources*, 18 (6): 1392–1401, https://doi.org/10.1111/1755-0998.12928.

Wright E.S., Vetsigian K.H. (2016). Quality filtering of Illumina index reads mitigates sample cross-talk. *BMC genomics*, 17 : 876, doi : 10.1186/s12864-016-3217-x.

Yang X., Chockalingam S.P., Aluru S. (2012) A survey of error-correction methods for next-generation sequencing. *Briefings in bioinformatics*, 14 (1): 56-66, doi:10.1093/bib/bbs015.

Ye S.H., Siddle K.J., Park D.J., Sabeti P.C. (2019) Benchmarking metagenomics tools for taxonomic classification. *Cell*, 178: 779-794, https://doi.org/10.1016/j.cell.2019.07.010.

Zaluga J., Stragier P., Baeyen S., Haegman A., Van Vaerenbergh J., Maes M., De Vos P. (2014) Comparative genome analysis of pathogenic and non-pathogenic *Clavibacter* strains reveals adaptations to their lifestyle. *BMC Genomics*, 15: 392, http://www.biomedcentral.com/1471-2164/15/392.

Zheng Z., Hou Y., Cai Y., Zhang Y., Li Y., Zhou M. (2015) Whole-genome sequencing reveals that mutations in myosin-5 confer resistance to the fungicide phenamacril in *Fusarium gramniearum*. *Scientific reports*, 5: 8248, doi: 10.1038/srep08248.

APPENDIX 1 – Terms, abbreviations and definitions

Alien control: It corresponds to a matrix infected by a target (called alien target) which belongs to the same group as the target organism but cannot be present in the samples of interest. The alien control can be used as an alternative to the negative and positive controls.

Alignment: Put sequences (reads, contigs or genomes) into correct relative positions with each other based on their nucleotide or amino acid similarity.

Amplicon sequencing: HTS test based on PCR amplification, such as metabarcoding. The PCR primers are usually designed to broadly amplify a specific genome region for a range of target organisms (e.g. bacteria, fungi, plants, viruses, insects, nematodes) and should be able to generate sequences from as many species as possible within this range.

Annotation: Information describing properties and features of a sequence region; sequence annotation can be either taxonomic (e.g. giving a taxonomic rank) or functional (e.g. identifying functional element like coding region, intron, promoter, miRNA, lncRNA, transposon, repeated sequences) depending on the intended use of the HTS test

Artefacts: Errors introduced in a sequence during one or several technical step(s) of the HTS process.

ASV: amplicon sequence variants are erroneous sequences generated during PCR and sequencing.

Background reads removal: A sub-step of the bioinformatic component of the HTS process in which non-target sequences are completely or partially excluded from the dataset. Also called background depletion or subtraction, reference subtraction or negative selection.

Background reads: Sequences not related to the targets (defined below). These sequences may be for example (part of) host sequences, phage sequences, environmental contaminants sequences (e.g. bacteria commonly found in the air, on plant surfaces, in reagents).

Barcode: A molecular method of species identification using a short section of DNA from a specific gene or genes that are sequenced and compared to reference DNA sequence databases.

Base quality scores: Indicates the probability that a base is called incorrectly. Each base in a read is assigned a quality score by a phred-like algorithm. A quality score of 10 means there is a 1/10 chance that the base call is incorrect (0.9% accurate); a score of 20 means there is a 1/100 chance that the base call is incorrect (0.99% accurate) and a score of 30 means there is a 1/1000 chance that the base call is incorrect (0.999% accurate) (Illumina, https://www.illumina.com/science/technology/next-generation-sequencing/plan-experiments/quality-scores.html). Also called Phred quality score.

Bioinformatic: "*Conceptualising biology in terms of molecules (in the sense of physical chemistry) and applying "informatics techniques" (derived from disciplines such as applied maths, computer science and statistics) to understand and organise the information associated with these molecules, on a large scale."* (Luscombe *et al*., 2001). Can also be called dry bench, dry lab, *in silico* analysis.

Bioinformatic pipeline: a suite of several software that usually follow each other in order to conduct the complete bioinformatic analyses.

Blind testing: Processing of "*samples with known levels of pests between routine samples*" (EPPO PM 7/98, 2019) with the sample status unknown to the operator. Blind testing is considered as second line control in EPPO PM 7/98 (2019).

Clustering: A bioinformatic operation (used in metabarcording and metagenomics) in which reads with related sequences (e.g. similar genomic features, same or homologous gene or protein) are grouped together.

Cq: quantification cycle in a real-time PCR, previously known as threshold cycle (Bustin *et al*., 2009).

COI: cytochrome c oxidase subunit I is a gene commonly used as a marker in barcoding and metabarcoding to distinguish species of fungi, insects and nematodes. Also abbreviated as COX-1 or MT-CO1).

Contiguous sequences: Assembly of overlapping reads that together form a consensus region of DNA. Also called contigs.

***De novo* assembly**: A computational process in which the HTS generated reads are assembled into longer, continuous sequences, and sometimes full-length sequences without using a reference sequence (see definition).

Denoising: A bioinformatic operation (specific to metabarcording) in which reads with artefacts introduced during PCR amplification and sequencing (noisy sequences, e.g. nucleotide substitutions, length variation) are removed from the read data set in order to preserve the best reads.

DNA: Deoxyribonucleic acid.

dsRNA: Double-stranded RNA.

Duplicated reads: Identical reads generated during a sequencing run. Also called duplex reads.

ELISA: Enzyme-linked immunosorbent assay.

EPPO: European and Mediterranean Plant Protection Organization.

Genome completeness: Proportion of obtained sequence compared to the (near) complete genome. Also called completeness of the sequence, genome length coverage, horizontal coverage.

ICTV: International Committee on Taxonomy of Viruses, https://talk.ictvonline.org/taxonomy/

Indels: Insertions or deletions of nucleotides.

Index: A short sequence of oligonucleotides added during the library preparation when sequencing several samples in parallel. It is unique to each sample and allows the assignment of the generated sequences to the corresponding samples.

Index-hopping: a known phenomenon that has impacted HTS technologies from the time sample multiplexing was developed. It causes incorrect assignment of reads to libraries from the expected index to a different index (in the multiplexed pool)." . Can also be called barcode bleeding, barcode-hopping, barcode misassignment, cross-talk, index misassignment, index switching, miss-tag.

IPC: Internal positive control used to monitor the test performance in each individual sample. Positive internal controls can either be genes present in the matrix DNA or RNA (e.g. plant cytochrome oxidase gene or eukaryotic 18S rDNA for DNA and plant mRNA, nad5 gene for RNA) or added to the

DNA solutions (e.g. synthetic internal amplification controls) (instructions to authors of EPPO diagnostic protocols, https://www.eppo.int/ACTIVITIES/plant_quarantine/diagnostics).

IPPC: International Plant Protection Convention.

IT: Information technology.

High-throughput sequencing (HTS): Sequencing of thousands or millions of DNA molecules simultaneously. Also known as next generation sequencing (NGS), massive(ly) parallel sequencing (MPS) or deep sequencing.

LAMP: Loop-mediated isothermal amplification is a nucleic acid amplification technique performed in a single tube at a constant temperature.

Library preparation: Laboratory preparation of nucleic acids to make them compatible to the sequencing platform. There are two main ways to prepare the libraries to plant pest detection: the random sequencing of nucleic acids (also called shotgun sequencing) and the targeted sequencing of PCR products, also called amplicon sequencing (e.g. metabarcoding).

Mapping: See Reference mapping.

Metabarcoding: HTS method in which a specific DNA sequence (from a gene or genes of target species) can be amplified and sequenced at high throughput. It allows the simultaneous identification of many taxa within a single sample.

Metagenomics: Study of genetic material recovered directly from environmental samples, typically untargeted (i.e. by shotgun sequencing). In case RNA is sequenced instead of DNA, it is called metatranscriptomics.

Method: In plant health diagnostics, a procedure used for the detection and/or identification of a pest. "*Methods include: bioassay methods, fingerprint methods, isolation/extraction methods, molecular methods, morphological and morphometric methods, pathogenicity assessment and serological methods.*" (EPPO PM 7/76, 2018). Also called technique.

N_{50}: "*The length of the smallest contig such that 50% of the sum of all contigs is contained in contigs of size N_{50} or larger*" (Castro and Ng, 2017).

NAC: "*Negative amplification control [external control] to rule out false positives due to contamination during the preparation of the reaction mix: application of the amplification procedure to molecular grade water that was used to prepare the reaction mix.*" (instructions to authors of EPPO diagnostic protocols, https://www.eppo.int/ACTIVITIES/plant_quarantine/diagnostics).

NGS: Next generation sequencing. See high-throughput sequencing.

NIC: "*Negative isolation control [external control] to monitor contamination during nucleic acid extraction: nucleic acid extraction and subsequent amplification preferably of a sample of uninfected matrix or if not available clean extraction buffer*." (instructions to authors of EPPO diagnostic protocols, https://www.eppo.int/ACTIVITIES/plant_quarantine/diagnostics).

NPPO: National Plant Protection Organization.

Non-quarantine pest: "*Pest that is not a quarantine pest* (see below) *for an area [FAO, 1995]."* (ISPM 5, 2019).

OTU: Operational Taxonomic Unit (OTU) is a cluster of sequences based on their sequence similarities in order to assign it to relevant taxonomic levels.

PAC: **"***Positive amplification control [external control] to monitor the efficiency of the amplification: amplification of nucleic acid of the target organism. This can include nucleic acid extracted from the target organism, total nucleic acid extracted from infected host tissue, whole genome amplified DNA or a synthetic control (e.g. cloned PCR product). For PCRs not performed on isolated organisms, the PAC should preferably be near to the limit of detection.*" (instructions to authors of EPPO diagnostic protocols, https://www.eppo.int/ACTIVITIES/plant_quarantine/diagnostics).

PCR: Polymerase chain reaction, a method for targeted amplification of nucleic acid sequences.

Percentage of coverage: The proportion of the genome covered by at least one read, expressed as a percentage of the length of the genome. Also called coverage length.

Pest: "*Any species, strain or biotype of plant, animal or pathogenic agent injurious to plants or plant products."* (ISPM 5, 2019).

PIC: "*Positive isolation control [external control] to ensure that nucleic acid of sufficient quantity and quality is isolated: nucleic acid extraction and subsequent amplification of the target organism or a matrix sample that contains the target organism (e.g. naturally infected host tissue or host tissue extract spiked with the target organism).*" (instructions to authors of EPPO diagnostic protocols, https://www.eppo.int/ACTIVITIES/plant_quarantine/diagnostics).

Poorly characterized organism: A known organism for which there are no existing tests for diagnosis.

Quasispecies (viruses): A population of closely related viral genome from a single species subject to high mutation rates as a result of viral replication and selection, and which act as a unit of selection..

Quarantine pest: "*A pest of potential economic importance to the area endangered thereby and not yet present there, or present but not widely distributed and being officially controlled."* (ISPM 5, 2019).

Read: Inferred sequence of nucleotides corresponding to a complete or partial unique DNA fragment, resulting from a high-throughput sequencing experiment.

Read depth: Number of aligned reads covering a specific nucleotide position. Also called vertical coverage, depth of coverage, coverage read depth. It is calculated as follows: for a given genome, the mean read depth is the number of reads mapping that genome multiplied by the read lengths and divided by the genome length. Also called coverage depth or coverage fold.

Reference mapping: Sequences (reads or contigs) compared to an existing reference sequence (or backbone sequence), building a sequence that is similar but not necessarily identical to the reference sequence (see definition below). Can also be called reference assembly.

Reference material: "*Material appropriate to the test and diagnostic being performed such as live cultures, infested plant material, DNA/RNA preparations, images of a diagnostic quality or mounted specimens. The reference material used should be documented and appropriate to the test and diagnosis being performed. It should be ensured that the material used is producing the features for which it was selected, for example expressing a desired antigen for use in serological diagnosis, or display specific physical features (e.g. sporulation) if used for morphological diagnosis.*" (EPPO PM 7/76, 2018).

Reference sequence: sequences (partial or (nearly) complete genome, gene) used to map or annotate the reads or contigs.

Regulated non-quarantine pest: "*A non-quarantine pest* (see above) *whose presence in plants for planting affects the intended use of those plants with an economically unacceptable impact and which is therefore regulated within the territory of the importing contracting party."* (ISPM 5, 2019).

Regulated pest: "*A quarantine pest* (see above) *or a regulated non-quarantine pest* (see above)*."* (ISPM 5, 2019).

RNA: Ribonucleic acid.

rRNA: Ribosomal RNA.

RT: Reverse-transcription, the process of synthesizing complementary DNA from an RNA template by the enzyme reverse transcriptase.

Scaffold: Created by joining contigs together using additional information (introducing arbitrary N letters) about the relative position and orientation of the contigs in the genome (Jung *et al*., 2019).

Sequencing run: Single use of a sequencing machine to sequence one or several libraries.

Shotgun sequencing: Random sequencing of any DNA molecule present in a sample, whatever its origin: for example, pest, endophytic micro- and macroorganisms, host (e.g. plant). Also called random sequencing.

siRNA: Small interfering RNA.

SNP: Single nucleotide polymorphism.

ssRNA: Single-stranded RNA.

Strand bias: On a single genome position, a strand bias occurs when the proportion of reads from a forward sequence and from its corresponding pairs from an expectation of equal likelihood of sequencing the plus and minus strands.

Tagmentation: Illumina defined it as the "*step included in shotgun library preparation which involves the transposon cleaving and tagging of the double-stranded DNA with a universal overhang*".

Target: One or more variants, strains, species, genus, family or group of organisms (e.g. bacteria, fungi, viruses) that are being tested for. They can include pest and non-pest organisms. The type of target depends on the scope of the test.

Test: "*The application of a method to a specific pest and a specific matrix.*" (EPPO PM 7/76, 2018). In these HTS guidelines, the method is HTS.

Trimming: Removal of nucleotides at one or both extremities of reads. These nucleotides usually correspond to low quality nucleotides or to nucleotides added to the sample DNA (e.g. primers, adapters, indexes). The aim is to either to remove nucleotides not of interest or to keep reads and nucleotides of appropriate quality for further analysis.

U_{50}: "*The length of the smallest contig such that 50% of the sum of all unique, target-specific contigs is contained in contigs of size U_{50} or larger*" (Castro and Ng, 2017).

Uncharacterized organism: An organism that has never been detected/identified before (e.g. new species or new variant/strain, new to science organism) or whose biological properties (e.g. host range, transmission, symptomatology) are not known.

VANA: Virion-associated nucleic acids, a nucleic acid extraction method for plant viruses.

Variants: Single nucleotide polymorphism (SNP), insertion or deletion (indel), integration or deletion of genes (structural variants [SV]) or homologous recombination observed in a sequence compared to reference sequence target(s).

Variant read number: Number of independent reads supporting the presence of a variant (Strom, 2016). Often expressed as a proportion to the total number of reads observed on that position.