DOI: 10.1111/epp.12863

ORIGINAL ARTICLE

Facilitating the adoption of high-throughput sequencing technologies as a plant pest diagnostic test in laboratories: A step-by-step description

Benedicte Lebas[1] | **Ian Adams**[2] | **Maher Al Rwahnih**[3] | **Steve Baeyen**[4] | **Guillaume J. Bilodeau**[5] | **Arnaud G. Blouin**[1] | **Neil Boonham**[6] | **Thierry Candresse**[7] | **Anne Chandelier**[8] | **Kris De Jonghe**[4] | **Adrian Fox**[2] | **Yahya Z. A. Gaafar**[9] | **Pascal Gentit**[10] | **Annelies Haegeman**[4] | **Wellcome Ho**[11] | **Oscar Hurtado-Gonzales**[12] | **Wilfried Jonkers**[13] | **Jan Kreuze**[14] | **Denis Kutjnak**[15] | **Blanca Landa**[16] | **Mingxin Liu**[17] | **François Maclot**[1] | **Martha Malapi-Wight**[18] | **Hano J. Maree**[19,20] | **Francesco Martoni**[21] | **Natasha Mehle**[15] | **Angelantonio Minafra**[22] | **Dimitre Mollov**[23] | **Adriana Moreira**[24] | **Mark Nakhla**[25] | **Françoise Petter**[26] | **Alexander M. Piper**[21] | **Julien Ponchart**[1] | **Robbie Rae**[27] | **Benoit Remenant**[10] | **Yazmin Rivera**[25] | **Brendan Rodoni**[21] | **Johanna W. Roenhorst**[28] | **Johan Rollin**[1] | **Pasquale Saldarelli**[22] | **Johanna Santala**[29] | **Rose Souza-Richards**[30] | **Davide Spadaro**[31] | **David J. Studholme**[32] | **Stefanie Sultmanis**[5] | **René van der Vlugt**[33] | **Lucie Tamisier**[1] | **Charlotte Trontin**[26] | **Ines Vazquez-Iglesias**[2] | **Claudia S. L. Vicente**[34] | **Bart T. L. H. Vossenberg**[28] | **Thierry Wetzel**[35] | **Heiko Ziebell**[9] | **Sebastien Massart**[1]

[1]Plant Pathology Laboratory, Terra-Gembloux Agro-Bio Tech, University of Liège, Liège, Belgium
[2]FERA Science Ltd, York Biotechnology Campus, York, UK
[3]Department of Plant Pathology, University of California-Davis, Davis, California, USA
[4]Plant Sciences Unit, Flanders Research Institute for Agriculture, Fisheries and Food (ILVO), Merelbeke, Belgium
[5]Canadian Food Inspection Agency, Ottawa, Ontario, Canada
[6]Institute for Agrifood Research Innovations, Newcastle University, Newcastle upon Tyne, UK
[7]UMR Biologie du Fruit et Pathologie, Université de Bordeaux, INRAE, Villenave d'Ornon, France
[8]Walloon Agricultural Research Centre, CraW, Gembloux, Belgium
[9]Institute for Epidemiology and Pathogen Diagnostics, Julius Kühn Institute – Federal Research Centre for Cultivated Plants, Braunschweig, Germany
[10]Unité de Bactériologie, Plant Health Laboratory, Virologie et détection des OGM, ANSES, Angers, France
[11]Plant Health Diagnostic Laboratory, Ministry for Primary Industries, Auckland, New Zealand
[12]Plant Germplasm and Quarantine Program, USDA-APHIS, PPQ, Beltsville, USA
[13]Bejo Zaden BV, Warmenhuizen, the Netherlands
[14]CGIAR, Consultative Group for International Agricultural Research, Lima, Peru
[15]Department of Biotechnology and Systems Biology, National Institute of Biology, Ljubljana, Slovenia
[16]Institute for Sustainable Agriculture, CSIC, Córdoba, Spain
[17]School of Natural Sciences, University of Tasmania, Burnie, Australia
[18]Biotechnology Risk Analysis Programs, USDA-APHIS, BRS, Washington, USA
[19]Department of Genetics, Stellenbosch University, Matieland, South Africa
[20]Citrus Research International, Matieland, South Africa
[21]Agriculture Victoria, AgriBio Centre for AgriBiosciences, Bundoora, Victoria, Australia

[22]IPSP-CNR, Institute for Sustainable Plant Protection, Consiglio Nazionale delle Ricerche, Bari, Italy
[23]Horticultural Crops Research Unit, USDA-ARS, Corvallis, USA
[24]FAO, International Plant Protection Convention, Rome, Italy
[25]Science and Technology, USDA-APHIS, PPQ, Hanover, USA
[26]European and Mediterranean Plant Protection Organization, Paris, France
[27]School of Biological and Environmental Sciences, Liverpool John Moores University, Liverpool, UK
[28]Netherlands Food and Consumer Product Safety Authority, National Plant Protection Organization, Wageningen, the Netherlands
[29]Plant Analytics, Finnish Food Authority, Helsinki, Finland
[30]International Seed Federation, Nyon, Switzerland
[31]Department of Agricultural, Forest and Food Sciences and AGROINNOVA – Centre of Competence for the Innovation in the Agro-environmental Sector, University of Torino, Turin, Italy
[32]Biosciences, University of Exeter, Exeter, UK
[33]Wageningen University and Research, Wageningen, the Netherlands
[34]Instituto Nacional de Investigação Agrária e Veterinária (INIAV I.P.), Quinta do Marquês, Oeiras, Portugal
[35]DLR Rheinpfalz, Institute of Plant Protection, Neustadt an der Weinstrasse, Germany

Correspondence
S. Massart, Plant Pathology Laboratory, Terra-Gembloux Agro-Bio Tech, University of Liège, Liège, Belgium.
Email: sebastien.massart@uliege.be

Funding information
This article is based upon work from the work package 2 of the project VALITEST (https://www.valitest.eu/), supported by the European Union's Horizon 2020 research and innovation programme under grant agreement no. 773139.

Abstract

High-throughput sequencing (HTS) is a powerful tool that enables the simultaneous detection and potential identification of any organisms present in a sample. The growing interest in the application of HTS technologies for routine diagnostics in plant health laboratories is triggering the development of guidelines on how to prepare laboratories for performing HTS testing. This paper describes general and technical recommendations to guide laboratories through the complex process of preparing a laboratory for HTS tests within existing quality assurance systems. From nucleic acid extractions to data analysis and interpretation, all of the steps are covered to ensure reliable and reproducible results. These guidelines are relevant for the detection and identification of any plant pest (e.g. arthropods, bacteria, fungi, nematodes, invasive plants or weeds, protozoa, viroids, viruses), and from any type of matrix (e.g. pure microbial culture, plant tissue, soil, water), regardless of the HTS technology (e.g. amplicon sequencing, shotgun sequencing) and of the application (e.g. surveillance programme, phytosanitary certification, quarantine, import control). These guidelines are written in general terms to facilitate the adoption of HTS technologies in plant pest routine diagnostics and enable broader application in all plant health fields, including research. A glossary of relevant terms is provided among the Supplementary Material.

Faciliter l'adoption des technologies de séquençage à haut débit pour les tests de diagnostic effectués dans les laboratoires phytosanitaires : une description étape par étape

Le séquençage haut débit (HTS) est un outil puissant qui permet, simultanément, la détection et l'identification potentielle de tout organisme présent dans un échantillon. L'application des technologies HTS suscite un intérêt croissant dans les laboratoires phytosanitaires pour les activités de diagnostic de routine et cet intérêt a conduit à l'élaboration de directives sur la manière de préparer les laboratoires à effectuer des tests HTS. Cet article décrit les recommandations générales et techniques, élaborées afin de guider les laboratoires dans le processus complexe de se préparer aux tests HTS, dans le cadre des systèmes d'assurance qualité existants. De l'extraction des acides nucléiques à l'analyse et interprétation des données, toutes les étapes sont décrites afin de garantir des résultats fiables et reproductibles. Ces

directives sont applicables pour la détection et l'identification de tout organisme nuisible aux végétaux (p. ex. arthropodes, bactéries, champignons, nématodes, plantes ou adventices envahissantes, protozoaires, viroïdes, virus), et à partir de tout type de matrice (culture microbienne pure, tissu végétal, sol, eau), quelle que soit la technologie HTS (p. ex. séquençage d'amplicons, séquençage shotgun) et l'application (p. ex. programme de surveillance, certification phytosanitaire, quarantaine, contrôle des importations). Ces directives sont rédigées avec des termes génériques afin de faciliter l'adoption des technologies HTS dans les activités de diagnostic phytosanitaire de routine et de permettre une application plus large dans tous les domaines de la santé des végétaux, y compris le domaine de la recherche. Un glossaire des termes utiles est fourni dans les documents complémentaires.

Содействие внедрению технологий высокопроизводительного секвенирования в качестве диагностического теста на присутствие вредителей растений в лабораториях: пошаговое описание

Высокопроизводительное секвенирование (HTS) - это мощный инструмент, позволяющий одновременно обнаруживать и потенциально идентифицировать любые организмы, присутствующие в образце. Растущий интерес к применению технологий HTS для рутинной диагностики в лабораториях, занимающихся здоровьем растений, требует разработку рекомендаций по подготовке лабораторий к проведению HTS-тестирования. В данном документе описаны общие и технические рекомендации, которые помогут лабораториям пройти сложный процесс подготовки лаборатории к проведению HTS-тестов в рамках существующих систем обеспечения качества. Рассматриваются все этапы для обеспечения надежных и воспроизводимых результатов, начиная с выделения нуклеиновых кислот и заканчивая анализом и интерпретацией данных. Настоящее руководство актуально для обнаружения и идентификации любого вредного организма растений (например, членистоногих, бактерий, грибов, нематод, инвазивных растений или сорняков, простейших, вироидов, вирусов) и из любого типа матрицы (например, чистая культура микроорганизмов, ткани растений, почва, вода), независимо от технологии ВПС (например, ампликонное секвенирование, дробовое секвенирование) и области применения (например, программа надзора, фитосанитарная сертификация, карантин, контроль импорта). Настоящее руководство составлено в общих терминах, чтобы облегчить внедрение технологий HTS в рутинную диагностику вредителей растений и обеспечить её более широкое применение во всех областях защиты растений, включая научные исследования. В дополнительных материалах приводится глоссарий соответствующих терминов.

1 | INTRODUCTION

High-throughput sequencing (HTS), also known as next-generation sequencing or deep sequencing, is the most significant new method in plant health diagnostics since polymerase chain reaction (PCR)-based detection was introduced in the late 1980s. High-throughput sequencing technologies have the potential to detect the nucleic acids of any organism present in a sample, including distant variants and uncharacterized organisms (Hadidi et al., 2016; Massart et al., 2014). This very large inclusivity is achieved without needing any a priori information on the content of the sample.

One of the most frequent uses of HTS technologies in plant pest diagnostics is the identification of pests causing novel diseases or diseases of unknown aetiology. This has led to the discovery of hundreds of previously uncharacterized organisms or strains of organisms associated

with symptomatic and asymptomatic plants (Aritua et al., 2015; Barba et al., 2014; Malapi-Wight et al., 2016; Maliogka et al., 2018). For example, HTS technologies allowed a rapid increase in the number of complete genomes of bacteria (e.g. Xu & Wang, 2019), fungi (e.g. https://mycocosm.jgi.doe.gov/mycocosm/home/1000-fungal-genomes), phytoplasmas (e.g. Palmano et al., 2012) and viruses (e.g. Rivarez et al., 2021) sequenced in the past decade. Sequencing the complete genome of a pest can provide key insights into the pathogenicity mechanisms, as shown for the bacterium *Xylella fastidiosa* (Simpson et al., 2000). The *X. fastidiosa* sequences were also used to identify subspecies and to allow tracing the origin of the pathogen in an incursion (Cella et al., 2018). In addition, sequencing the genomes of many isolates/specimens/strains of a pest provides a broader overview of its genetic diversity and consequently improves the inclusivity of diagnostic primers and targeted diagnostic protocols (Adams et al., 2018; An et al., 2015; Bonants et al., 2015; Catara et al., 2021; Katsiani et al., 2018; Kikuchi et al., 2011; Owati et al., 2019; Pritchard et al., 2016).

In phytosanitary certification schemes, HTS technologies can be used to certify nuclear stock, seeds and plant propagation material. They can also be used for (post-entry) quarantine testing to prevent the establishment of pests in a country or area (Candresse et al., 2014; Fox et al., 2019; Malapi-Wight et al., 2021), and to monitor imported commodities from different countries to avoid potential risks for plant health (Abdelfattah et al., 2019). High-throughput sequencing technologies have been evaluated as a generic method for virus and/or viroid detection in grapevine and fruit trees (Al Rwahnih et al., 2015; Rott et al., 2017; Soltani et al., 2021; Villamor et al., 2021). They have also been used for the identification of bacteria in the re-emerging disease 'acute oak decline' caused by a polymicrobial complex (Denman et al., 2018). In addition, HTS technologies have been evaluated for the detection of viable plant propagules at an international point of entry (Whitehurst et al., 2020) and used to detect plant viruses and variants in wastewater (Bačnik et al., 2020).

HTS technologies have also been used in surveillance programmes, monitoring and source tracking utilizing a PCR-based approach called metabarcoding or amplicon sequencing (Hamelin & Roe, 2019). Metabarcoding uses generic primers to target short genomic regions conserved across multiple organisms to provide an identification up to a defined taxonomic level. Amplicon sequencing has been recently applied in various ecosystems and areas for surveillance studies of, for example, airborne fungi and oomycetes, including plant pathogens (Abdelfattah et al., 2019; Aguayo et al., 2018; Chandelier et al., 2021; Franco Ortega et al., 2020; Mbareche et al., 2020; Nicolaisen et al., 2017; Nilsson et al., 2019; Núñez et al., 2017; Ovaskainen et al., 2020; Tremblay et al., 2018, 2019), insects (Braukmann et al., 2019; Elbrecht et al., 2019; Piper et al., 2019) and plants (Bruni et al., 2015; Núñez et al., 2017; Tremblay et al., 2019). Since this technique targets short genomic regions, it can be extremely versatile and can be used on a wide variety of samples.

With the decrease in sequencing cost, the availability of effective sequencing machines and the improved accessibility of bioinformatic tools for analysing HTS sequencing data, there is a rapidly increasing interest in implementing HTS technologies for routine diagnostics, including regulatory plant health (Catara et al., 2021). However, caution should be applied when interpreting the results of HTS technologies, in particular when the results are used to implement phytosanitary measures (IPPC Secretariat, 2019; Olmos et al., 2018).

One of the main challenges for a routine use of HTS technologies in plant health laboratories is the current lack of internationally recognized, harmonized guidelines covering both laboratory and bioinformatics steps (Adams et al., 2018; Olmos et al., 2018). These should address the specific challenges of HTS technologies such as personnel training and competence assessment, infrastructure, equipment and quality assurance that complies with national and international (e.g. ISO) standards. In addition, a range of factors, such as establishing the quality metric thresholds and their acceptable range of values, also need to be considered in a routine setting.

The aim of this paper is to propose general recommendations (e.g. laboratory and computing infrastructure, quality management system) and technical requirements for a laboratory to prepare for implementing HTS technologies for plant pest diagnostics. They cover all of the steps of the HTS process: from nucleic acid extraction to data analysis and interpretation. Importantly, they have been developed irrespective of the molecular reactions, sequencing platform and software, and can be applied to any plant pest in any matrix. An overview of the HTS process in plant health diagnostics is also provided as well as a glossary of relevant terms (See Appendix S1).

By following these recommendations, a laboratory should be ready to apply HTS technologies and start their development, validation or verification (Soltani et al., 2021). For this purpose, a complementary publication (Massart et al., 2022) provides technical and management guidelines covering the process of implementation of HTS technologies in a research or diagnostics laboratory (selection, development, verification and validation). The complementary publication also has a strong focus on risk analysis, the use of controls and result interpretation.

This publication is an output of work package 2 of the European project, VALITEST (https://www.valitest.eu/index), which aimed to improve the validation approaches for diagnostic technologies to maximize their usefulness for users (diagnosticians) and decision-makers (at Regional, National and European levels) and their use in routine diagnostics.

2 | OVERVIEW OF THE HTS PROCESS IN PLANT HEALTH DIAGNOSTICS

Two main HTS approaches are currently widely adopted in research to detect plant pests: firstly, the sequencing of amplicons generated by PCR or rolling circle amplification and related protocols (also called targeted sequencing or metabarcoding); and secondly, shotgun sequencing of nucleic acids (also known as metagenomics or random sequencing).

Irrespective of the approach, the HTS process can be divided into eight steps (Figure 1). After sampling (step 1), HTS tests can be divided into six distinct steps encompassing laboratory and bioinformatics components followed by the confirmation and interpretation of the results. For each step, a range of protocols are available and regular updates of the laboratory or bioinformatics protocols are expected in the near future as the technology advances.

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

Step 2: nucleic acid extraction. The source of nucleic acids can be a matrix containing multiple organisms or isolated organisms. The nucleic acids can be genomic DNA or RNA, total DNA or RNA, small interfering RNAs or double-stranded RNAs (Gaafar & Ziebell, 2020; Maliogka et al., 2018; Pecman et al., 2017; Visser et al., 2016).

Step 3: library preparation. The aim of the library preparation step is to isolate and, often also, to amplify a sufficient quantity of nucleic acids of appropriate size that are flanked with the adapters and indexes (oligonucleotide sequences) required for sequencing. The adapters are short nucleotide sequences specific to the sequencing platform that enable the nucleic acid fragments to anchor to the sequencer to start the sequencing process. The indexes are used for multiplexed runs (see below) to link each nucleic acid fragment to the sample it originated from. The nucleic acids are prepared according to the selected sequencing approach, whether amplicon sequencing or shotgun sequencing.

- For *amplicon sequencing*, specific genomic regions are amplified, mainly by PCR, and sequenced. Primers are usually composed of a sequence complementary to the target sequence, allowing the amplification(at the 3′-end) of the target region and, at the 5′-end, of an adapter sequence (optionally including an index). Another option is to first perform a PCR with the target primers followed by a second PCR with overlapping primers containing an adapter or, alternatively, to use an adapter ligation step (to avoid a second PCR reaction). Long amplified products can also be fragmented before being further sequenced as described below (shotgun sequencing).
- For *shotgun sequencing*, a reverse-transcription step is applied when starting from RNA, although direct RNA sequencing protocols are possible on some platforms (e.g. Oxford Nanopore Technologies, Pacific BioSciences single-molecule real-time; Zhao et al., 2019). Protocols for library preparation include shearing (sonication) or digestion (restriction enzymes or chemical lysis) of nucleic acids followed by end-repair, and ligation of adapter sequences. Alternatively, random hybridization and amplification using degenerated oligonucleotides or the use of transposases can be applied to fragment nucleic acids (van Opijnen & Camilli, 2013; Wilcox et al., 2018). Adapters are then ligated to one or both ends of the sample's fragmented nucleic acids. These steps can be complemented by an additional PCR amplification.
- *Enrichment or target selection* is an optional step of nucleic acid extraction or library preparation step. For shotgun sequencing, the target enrichment or selection can be performed by removing untargeted sequences [for example, the removal of plant ribosomal RNA (rRNA), called ribodepletion, or enriching dsRNA by cellulose capture, concentrated salt precipitation and/or nuclease digestion] or by using oligonucleotides specific to the targets, sometimes also referred to as probe capture (Adams & Fox, 2016; Gaafar & Ziebell, 2020; Maliogka et al., 2018).
- *Pooling of samples*, also called multiplexing, can be an option in many library preparation protocols. The sequencing of pooled samples in a single run allows a reduction of the cost per sample. A unique identifier, called an index (also known as barcode, tag, molecular identifier or MID) is a short oligonucleotide sequence flanked (or not) by an adapter that is used to tag each sample. This index is sequenced along with the target molecules or in a separate sequencing reaction, allowing each determined sequence to be linked to the appropriate sample after appropriate demultiplexing (see step 4) (Budowle et al., 2014; Piper et al., 2019). Indexes can be added to one end or to both ends (dual indexing) of the nucleic acid fragment by ligation or fusion primers (Tremblay et al., 2018).

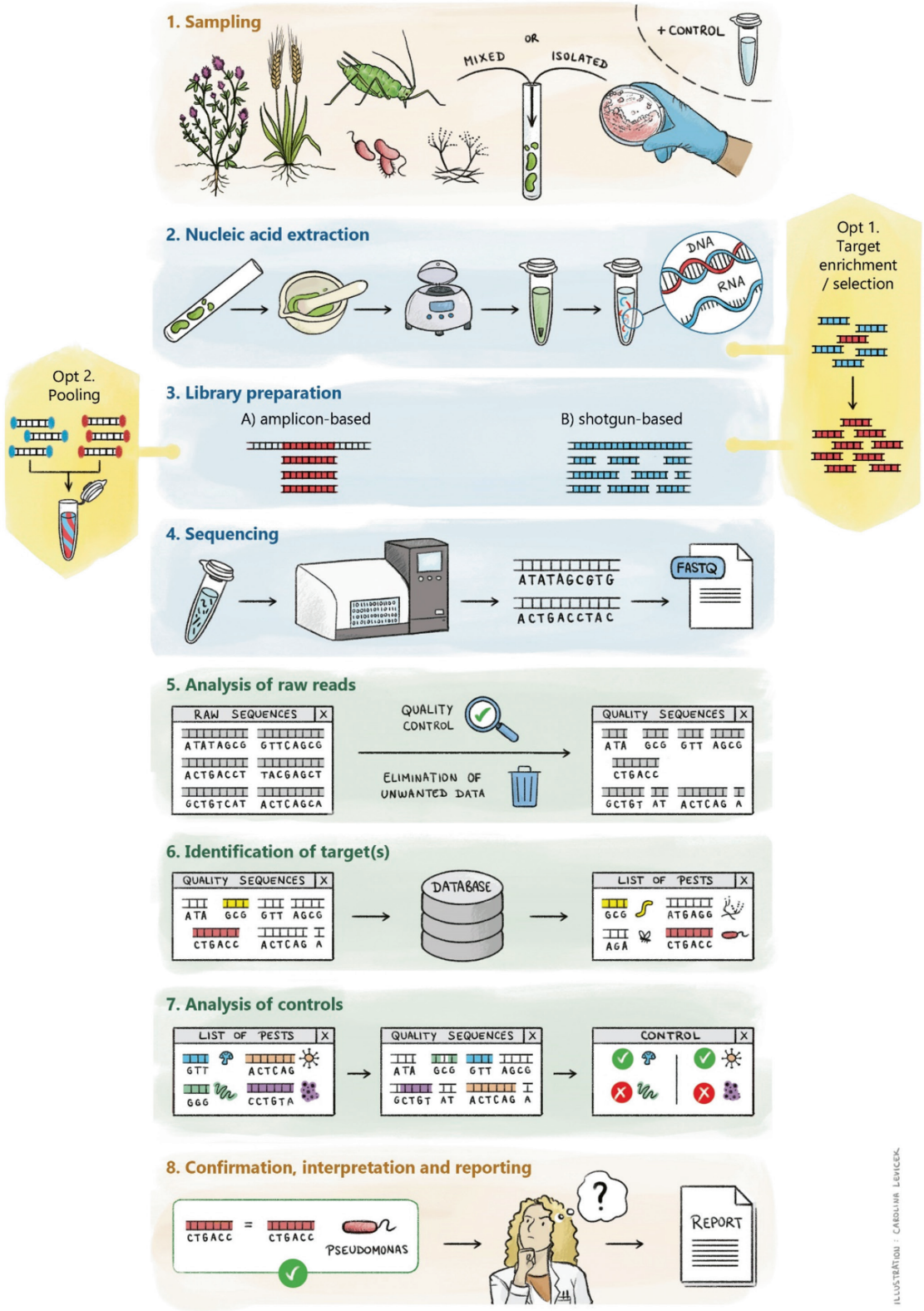

FIGURE 1 Schematic representation of the main steps of the high-throughput sequencing (HTS) process used in plant health diagnostics

Step 4: sequencing. Currently, there are a limited number of sequencing platforms that have been widely commercialized. They have already been described in detail and reviewed elsewhere (Maljkovic Berry et al., 2020).

The next steps are related to the bioinformatic component of the HTS process which is broadly divided into three steps, each containing several sub-steps. It should be noted that the results of each sub-step

depend on the selected parameters and on metrics of the previous sub-step(s).

Step 5: analysis of raw reads. This bioinformatic step consists of several operations, including quality control of the generated sequences (depending on the sequencing technology and allowing the elimination of low-quality sequences and nucleotides) and the (optional) removal of adapter, index and primer sequences. In the case of pooled samples, demultiplexing enables the correct assignment of the generated sequences to each sample. Optionally, some additional analyses can be performed to reduce the amount of data and to improve the quality of the analysis, for example by merging forward and reverse reads, based on the overlapping region (if present), or removing duplicated (identical) reads.

Step 6: identification of target(s). This bioinformatic step, also called sequence annotation or assignment, aims to associate sequences with specific organisms. Depending on the use of the HTS test, sequence annotation can be a taxonomic classification (e.g. attributing reads to a species, genus or family) and/or a functional annotation (e.g. determining if a read belongs to a coding region, intron, promoter, micro RNA, long non-coding RNA, transposon or repeated sequence). Targeted identification currently always relies on 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) following de novo assembly of the reads into contigs; (iii) following mapping of the reads on reference sequences (reference assembly); or (iv) using a combination of these. In metabarcoding, reads are grouped into representative bins or clusters called operational taxonomic units (OTUs) or amplicon sequence variants (ASVs), and then compared against reference sequences database(s) to identify the most likely organism(s). Alternatively, reads with artefacts (also called noisy sequences) introduced during library preparation (e.g. nucleotide substitutions, length variation, chimeras) can be removed before OTU clustering (this is called denoising). Ultimately, once the reads have been assembled de novo, mapped to a reference sequence or grouped into OTUs, it is possible to identify variants of each sequence, corresponding to single nucleotide polymorphisms (SNPs), the insertion and deletion of nucleotides (indels) or the integration or deletion of larger parts of DNA/RNA (structural variants).

Step 7: analysis of controls. This analysis aims to verify that all the controls included in the HTS run produced the expected results to identify and eliminate potential false positive and/or false negative results.

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 the sample (in particular for uncharacterized organisms); and (iii) the reporting of the results of the HTS test.

3 | GENERAL RECOMMENDATIONS FOR IMPLEMENTING HTSTECHNOLOGIES

3.1 | Laboratory facilities and information technology infrastructure

As for any other molecular test, appropriate laboratory facilities help ensure reliable results. For example, contamination in HTS tests is particularly problematic because of the multiple handling steps and the use of many different reagents in the sample preparation process. The sources of contamination and how these can be monitored are discussed elsewhere (Asplund et al., 2019; Champlot et al., 2019; Dickins et al., 2014; Massart et al., 2019). General guidelines on PCR work such as those described in EPPO Standard PM 7/98 (2021) are applicable to HTS tests and should be followed. For example, the laboratory should have a forward workflow that follows the HTS process with dedicated areas for non-compatible steps such as nucleic acid extraction and amplification.

The implementation of HTS requires significant investment in information technology (IT) for both data storage and computing capacity. Large files (up to a few gigabytes per sample) are generated and need to be transferred, stored and properly backed up. Machines with high computational power or access to cloud-based services for advanced computing are required for running bioinformatic pipelines in a relatively short time frame (Olmos et al., 2018). A laboratory planning to implement HTS tests should always explore the most recent technological options available on the market, in order to acquire the IT infrastructure appropriate for the analysis that it is planning to perform. The IT infrastructure configuration for storage should take into consideration the expected number of samples, the volume of data per sample (including raw reads, intermediate data files and final results), the legal or commercial obligations related to data security and confidentiality, maintenance and data back-up. Further aspects to consider are the operating system environment (e.g. Windows, MacOS, Linux), which may impact the choice and version of bioinformatics algorithms available, as well as the computing power or server required to run the software(s) for timely delivery of results. The level of expertise needed when running and updating the infrastructure should also be

considered. A close collaboration with the IT department of the laboratory's organization is therefore recommended. To ensure proper traceability, registration of the log for all analyses and users in the bioinformatics pipeline is also preferable. Some steps of the HTS process, for example sequencing or bioinformatic analyses, may be outsourced to external sequencing facilities and bioinformatics service providers. Laboratories can also rent computational power and storage space on commercially available computer clusters. Requirements for outsourcing are discussed below.

3.2 | Personnel requirements

The use of HTS technologies requires trained personnel with expertise for each step of the process, including laboratory and bioinformatic components and in the biological interpretation of the results. Guidelines on the competence and expertise of personnel can be found in the EPPO Standard PM 7/98 (2021). The importance of scientific expertise has been highlighted during a proficiency test on the identification of food-borne pathogens in a simulated dataset (Brinkmann et al., 2019) and on the analysis of small RNAs datasets for the identification of plant viruses (Massart et al., 2019).

As with any other molecular diagnostic test, only qualified and trained personnel should process the samples. For sequence data analysis, specific IT infrastructure and expertise in bioinformatics is needed. The bioinformatics component requires trained personnel able to run bioinformatic pipelines correctly (installation, development, validation, routine use and regular update of the software and databases). In addition, relevant scientific expertise is needed for the choice of biologically specific settings and parameters (such as the choice of a similarity threshold to generate OTUs), as well as for the appropriate interpretation of the data (to avoid reporting false positive and/or false negative results) and evaluation of their biological relevance. Relevant scientific expertise may also be required for decision-making on possible follow-up actions (e.g. confirmatory testing). Some specific expertise can be outsourced such as the development and implementation of a bioinformatic pipeline under conditions described below.

3.3 | How to ensure consistent operation and traceability

The laboratory should have a quality management system in place (including a documentation system), which would enable any operation carried out to be traced back and to identify the origin of samples or contamination. The documentation system should describe all of the procedures required to perform an HTS test from sampling to results reporting, including the different steps in the laboratory, and the bioinformatic components (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 the bioinformatic pipeline(s) used. These recommendations 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 application of HTS tests, including their bioinformatic component. The laboratory should ensure that the procedures are kept up to date and that the current versions are used by the personnel (EPPO PM 7/98, 2021). As part of the quality control system, the laboratory should keep records of personnel training related to HTS testing, of test development (when relevant), of validation efforts (when appropriate, including those after changes have been made to an HTS test), of test runs (including the values of relevant quality metrics, version of software, pipeline, sequence databases), of diagnostic results (see EPPO PM 7/77, 2019), of critical equipment (e.g. maintenance and calibration certificates), of critical kits/reagents (e.g. lot number, expiration dates of reagents) and of sample and administrative information. Records should be kept for a period that meets customer and legal requirements. For example, EPPO recommends a minimum data retention period of 5 years unless the national requirements specify otherwise (EPPO PM 7/77, 2019).

3.4 | Checking the quality performance of the outsourced services

Laboratories can outsource parts of an HTS test (e.g. nucleic acid extraction, library preparation, sequencing services, bioinformatic analyses) and the requirements are stated in EPPO Standard PM 7/130 (2016).

It is recommended to select a provider that has at least the same level of quality assurance management as the diagnostic laboratory, ideally with an official accreditation or certification such as ISO 9001 or ISO 17025. 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 reported on anonymized samples. It should be noted that although some steps of the HTS process can be outsourced, the diagnostic laboratory remains responsible for the interpretation and reporting of results.

3.5 | Monitoring, implementing and documenting modifications

High-throughput sequencing technologies and protocols evolve quickly in both their laboratory and bioinformatic

components. This situation might often require updates in protocols, sequence databases and even bioinformatic pipelines. The laboratory should make efforts to keep track of any relevant changes by monitoring, implementing and documenting the modifications. For example, modification of the laboratory protocols, owing to new versions of kits or to newly available kits for library preparation or sequencing, should be documented. Regarding bioinformatic analyses, the laboratory should keep track of software versions and updates/upgrades with algorithms and parameter settings and keep records of changes to the underlying operating systems which might affect how pipelines and tools perform (e.g. integrate a log system to track all versions in the bioinformatic pipeline). For any modification, as recommended in EPPO Standard PM 7/98 (2021), an expert judgement should be made as to whether the update to a validated HTS test requires re-validation or verification. This evaluation should be documented.

3.6 | Appropriate IT infrastructure

Large data files are generated during each HTS run. These datasets need to be easily transferable within the IT infrastructure of the laboratory and, if relevant, from the sequencing provider to the laboratory. The network should be secured to ensure the integrity and confidentiality of the data. Numerous algorithms have been developed to check the integrity of the transferred file (e.g. md5sum – see https://www.ncdc.noaa.gov/nomads/documentation/user-guide/MD5-hash-files).

Data can be stored internally on the laboratory's own data storage system or externally on a cloud-based computing resource. When data is 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. Data should be backed up, ideally on a mirror server so as to prevent data loss in case of server failure.

The laboratory should have a documented procedure for data transfer, backup and storage. The procedure should describe how and where files generated during sequencing and bioinformatic analyses should be stored to ensure their integrity and confidentiality. The laboratory should also describe which files (i.e. input files, intermediate files and output files) should be kept and for how long. Any issue related to data transfer, backup and storage should be recorded if they can influence the test results (Aziz et al., 2015; Hébrant et al., 2018).

3.7 | Sequence databases for data analysis

Sequence databases are a critical part of bioinformatic analyses and are therefore a key focus point. They can be incomplete or contain errors while their content is constantly evolving because of scientific discoveries or changes in the taxonomy of pests. Thus, the selection of appropriate sequence databases is important for a correct taxonomic assignment and to avoid false negative or false positive results (Massart et al., 2019; Nilsson et al., 2019; Piombo et al., 2021; Piper et al., 2019). For example, it is highly recommended to include non-pest organisms closely related to the pests of interest. When the focus of the HTS test is on a limited range of known pests, a curated database can be created with verified sequences that are annotated and not redundant. However, when searching for uncharacterized or unexpected organisms, a more extensive and less curated database might prove more effective than a well-curated database with a limited number of entries (Lambert et al., 2018; Piper et al., 2019).

Sequence databases can be publicly available or can be developed (preferably from documented reference material) and maintained by the laboratory (i.e. in-house sequence databases). In either case, sequence databases should be evaluated to ensure the accuracy of the sequences in identifying at least the expected target(s). Also, it is important to use sequences that have been generated from accurately identified specimens (i.e. reference materials, EPPO Standard PM 7/98, 2021) for the compilation of curated sequence databases, avoiding incorrectly annotated sequences from morphological or phenotypic misidentification, and therefore erroneous pest reports (e.g. Taylor & Martoni, 2020). For example, a custom-made database has improved the taxonomic assignment of 16S rRNA sequences generated by amplicon sequencing from human intestinal microbiota (Ritari et al., 2015). Nevertheless, developing such curated databases is a time-consuming activity.

Sequence databases 'should be kept up to date and readily available' (EPPO PM 7/98, 2021) and information on these databases should be documented. Such information includes, but is not limited to, the version number, the date of download and the original source or location. The recording of the database version is important because sometimes the names of organisms change from one version to the next. Also, the laboratory needs to make sure that the target organisms are still part of the databases when upgrading to a novel version.

The laboratory should endeavour to upload sequence(s) – partial or (near) complete genome sequences, variants sequences – with biological information when available, to an online database such as the National Center of Biotechnology Information (https://www.ncbi.nlm.nih.gov/) and its European counterpart the European Nucleotide Archive (https://www.ebi.ac.uk/ena/browser/home), the Barcode of Life Data System (http://v4.boldsystems.org/; Ratnasingham & Hebert, 2007) or the EPPO Q-bank database (https://qbank.eppo.int/). Whenever possible, the sequence(s) should be permanently linked to a voucher specimen. This specimen should be kept by the laboratory and/or stored in a depository such as

the microorganisms collections of Belgian Coordinated Collections of Microorganisms (https://bccm.belspo.be/about-us/bccm-lmg), German Collection of Microorganisms and cell cultures (https://www.dsmz.de/) or International Collection of Microorganisms from Plants (https://www.landcareresearch.co.nz/tools-and-resources/collections/icmp-culture-collection/). Uploading sequences to public sequence databases will assist the scientific community to identify organisms.

4 | TECHNICAL RECOMMENDATIONS FOR IMPLEMENTING AN HTS TEST

4.1 | Scope of the HTS test

A clear and unequivocal definition of the intended use of the HTS test (i.e. detection or identification), the target organism(s) and the tested matrix is mandatory to ensure that the HTS protocol is fit-for-purpose. When using 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 from a range of matrices (e.g. plant, soil, water). 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 'use of amplicon sequencing for the surveillance of insects, bacteria or fungi collected from traps' (Aguayo et al., 2018; Núñez et al., 2017; Piper et al., 2019). In defining the scope, sample quality and quantity should be considered as re-sampling or re-testing may not be possible when processing certain diagnostic samples.

4.2 | Laboratory component

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

4.2.1 | The sampling protocol

The type of sample (e.g. different plant parts) and the season of sampling can affect the results of any diagnostic test, including HTS tests (e.g. organisms not detected, leading to misleading negative results; Malapi-Wight et al., 2021; Prezelj et al., 2013). Although the laboratory may not be involved in sampling, it may be necessary for the laboratory to recommend a sampling procedure. Such a procedure should describe the type of material (e.g. tissue for plants), the minimum amount of material needed, the number of samples to make up a batch and, when relevant, the season of sampling and the requirements for sampling symptomatic and/or asymptomatic material (EPPO Standard PM 7/98, 2021). 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 the preservation of DNA and RNA throughout the sampling phase should be taken into consideration. Examples of DNA preservatives include different concentrations of ethanol (Marquina et al., 2021), dimethyl sulfoxide (Moreau et al., 2013), propylene glycol (Martoni et al., 2021; Robinson et al., 2021) and RNA*later*® (Vink et al., 2005). Preservation of plant DNA and/or RNA may also be required at the time of sampling (e.g. RNA*later*® for RNA preservation) for shipment to the laboratory or to external services.

4.2.2 | Sample handling

As with any diagnostic test, the quality of samples can affect the results of HTS tests. The laboratory should have a procedure that includes measures to prevent cross-contamination between samples, subsampling, registration and traceability of samples, sample preservation between collection and laboratory reception (e.g. insects preserved in glycol/ethanol), transportation to the laboratory (e.g. cold-chain box containers, plastic bags to avoid dehydration), assessment of samples' condition on receipt, storage (e.g. cold room storage upon arrival), aliquoting, retention and disposal (EPPO Standard PM 7/98, 2021).

4.2.3 | Ensuring the quality and quantity of nucleic acids

The quality (in terms of purity and integrity) and quantity (i.e. ng/μL) of nucleic acids are important as they can affect the results of an HTS test. 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 have higher nucleic acid integrity and/or minimal concentration requirements. This is often the case with long-read HTS technologies. Ultimately, the quality of the extracted nucleic

acids should be checked and minimal thresholds should be pre-determined (e.g. minimum, average or maximum fragment length, minimal acceptable purity or yield). Based on their experience, laboratories may have a preference for certain extraction protocols based on their user preference/experience and reagents availability in their region.

The extraction allows for the removal of inhibitors that can negatively impact the test result. For instance, extraction methods producing a high yield and a minimum of PCR inhibitors are necessary for the detection of organisms found at a very low concentration in the sample. Some kits/methods are better for bacteria, others for fungi; the most appropriate should be used. Kits involving paramagnetic beads usually do not provide high yields but can allow higher throughput processing of samples. The selection of the extraction method depends on the type of the genome target(s) expected to be detected by the HTS test (e.g. DNA vs. RNA genomes) and the type of matrix from which the nucleic acids are extracted (e.g. plant parts – seed, leaf, fruit, stem, roots; purified cultures, soil, water, insects). The composition of the matrix can also affect the extraction of nucleic acids, as demonstrated by studies comparing DNA extraction methods of plant-associated bacterial communities from soil, xylem sap or different plant species prior to amplicon sequencing (Giangacomo et al., 2020; Haro et al., 2021). Both RNA extraction methods and sequencing platforms have resulted in significant differences in the detection of viruses and viroids from citrus samples (Bester et al., 2021).

Specific adaptations to a protocol may be needed for specific organisms/matrices. 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 lysis; Nielsen et al., 2019; Waeyenberge et al., 2019; Wesolowska-Andersen et al., 2014). Non-destructive DNA extraction may be preferred for macro-organisms (e.g. insects, nematodes) to preserve morphological voucher specimens, creating a permanent link with the DNA sequence for the confirmation of results or to inform future studies (Batovska et al., 2021; Carew et al., 2018; Nielsen et al., 2019; Piper et al., 2019).

The concentration of targets in a sample can be very low in some types of matrices such as water samples (Mehle et al., 2018), which may result in a failure to detect them. A target-enrichment or -selection step can be included in, or precede, the nucleic acid extraction protocol to improve the analytical sensitivity of the HTS test and decrease cost by requiring fewer reads per sample. The selection of the enrichment protocol depends on the target genome [e.g. single-stranded RNA, double-stranded RNA (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) and the matrix (e.g. plants, soil, water). For plant samples, there are a number of protocols that can improve sensitivity. For example, viral particle enrichment by ultracentrifugation, depletion of ribosomal RNA (rRNA) from total RNA or the enrichment of dsRNA by cellulose affinity chromatography with or without additional nuclease treatment(s) (Adams & Fox, 2016; Pantaleo & Chiumenti, 2018). Rolling circle amplification is also frequently used as an enrichment procedure when targeting DNA viruses with circular genomes (Johne et al., 2009). Targeted enrichment for a particular pest at a low concentration can also be designed to improve the sensitivity of the HTS test (Cai et al., 2019).

4.2.4 | The library preparation protocol

Whatever the HTS approach, selection of the protocol for library preparation depends on the HTS technology used.

For shotgun sequencing, the protocols are often provided as kit(s) with all reagents included. Their selection depends on technical criteria (e.g. the minimum required quantity and the integrity of the extracted nucleic acids and expected proportion of target nucleic acids), 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 size selection or on the use of specific oligonucleotides either to eliminate non-target nucleic acids (such as 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 can result in a 10-fold enrichment of viral sequences (Adams & Fox, 2016).

For amplicon sequencing, which usually relies on a PCR step, special care should be taken in selecting primers to ensure that the target organisms can be amplified, 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 minimize amplification errors owing to the misincorporation of nucleotides (Budowle et al., 2014; McInerney et al., 2014). The number of PCR cycles should be selected to ensure that the PCR is still in the exponential phase. Metabarcoding PCR amplification targets a small region of the genome, the barcode, generally corresponding to the partial sequence of a gene. Since the first work proposing the use of DNA barcoding (Hebert et al., 2003), barcodes have been proposed and described in EPPO Standard PM 7/129 (EPPO, 2021) for a range of organisms affecting plants. Some of these barcodes have been successfully used in metabarcoding (Ahmed et al., 2019; Dormontt et al., 2018; Nilsson et al., 2019; Ritter et al., 2019; Tremblay et al., 2018). When choosing the barcode for metabarcoding analyses, the risk of

potentially amplifying host sequences should be considered and can be reduced by selecting appropriate primers (Hanshew et al., 2013) or by using blocking oligonucleotides (Lundberg et al., 2013).

4.2.5 | Pooling level of libraries

Several libraries, each tagged independently by a specific sequence of nucleic acids (also called MID or index), can be pooled together to reduce the sequencing costs while taking into account the minimal number of reads expected per sample. However, the process of pooling introduces a higher variability in the number of generated reads per sample and increases the risk of assignment of reads to an incorrect sample owing to cross-contamination of tagging that can occur during library preparation and sequencing (i.e. index-hopping or index switching) or between sequencing runs (i.e. inter-run contamination if 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 owing to sequencing errors in indexes. The risk increases when high sequencing depths are obtained with pooled libraries, since a very low level of error can be detected (Budowle et al., 2014; Massart et al., 2019). The laboratory should also be aware of the expected index misassignment rates as differences exist between sequencing platforms used.

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 that increase the bioinformatic power in identifying 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 hopping as they do not take into account other origins like the creation of chimeric sequences owing to, for example, the ligation of free adapters (Wright & Vetsigian, 2016). Pooling libraries just prior to sequencing or adding a step to remove free adapters can also reduce these (mis)assignment issues. The sequences of sets of indexes included in each run should be recorded for trace-back purposes and for planning successive sequencing runs.

Pooling also requires that the amount of nucleic acid of each library in the pool is normalized. This minimizes, but does not eliminate, 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 and demonstrate that the pooling strategy used does not affect test performance (e.g. lower level of detection, higher contamination). The pooling method depends on the desired number of reads from 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).

4.2.6 | The sequencing platform

Based on the points listed below and on relevant publications (e.g. comparison of two platforms for the detection of citrus viroids and viruses; Bester et al., 2021), the laboratory should consider the sequencing platform and sequencing output best suited for the intended use of the HTS test.

A non-exhaustive list of parameters influencing the selection of the sequencing platform is presented below:

- Expected number of samples received per batch and number of reads needed per sample. Generally, the higher the number of reads generated per sample, the higher the chances are that all targets present in the sample will be identified, but this will increase the cost and, potentially, the the risk of detecting low-level contamination.
- Total number of generated reads per sequencing run. Generally, a higher throughput of data (and a lower cost per sample) can be obtained from technologies that produce more output and therefore allow a higher number of samples per run. This can be illustrated by the sequencing machines MiSeq and NovaSeq from Illumina which, in 2021, generated a maximum of 50 million and 20 billion reads per run, respectively.
- Required test turn-around time (e.g. urgent testing for perishable materials). Some technologies require a longer running time for a single analysis. The operator might prefer to choose a more expensive albeit faster technology if the results are required urgently.
- Read length and type (e.g. single, paired, mate-pair). Short single reads are appropriate for sRNA sequencing whereas amplicon sequencing might need longer reads.
- Error rate and type of error, which vary between sequencing platforms and between runs. Technologies generating longer reads, but at a higher error rate, are usually considered more appropriate for genome sequencing of a single purified organism, where the same region is sequenced multiple times and a sequencing error can thus be corrected. On the other hand, metabarcoding analysis should generally prioritize a lower error rate, since it aims to sequence the same gene region from multiple individuals and a sequencing error could be misinterpreted as genetic variation.
- Impact on the downstream bioinformatic analyses (depending on the number of sequences, their length, their quality and accuracy).
- Availability of bioinformatic support, platforms (when outsourced), laboratory resources and technical expertise and manufacturer level of technical support.
- Expenses involved in the operation of a sequencing machine – purchase and maintenance (Rehm et al., 2013).

Sequencing platforms are regularly updated and the laboratory should closely monitor these updates and evaluate their potential impact on the HTS test results.

4.2.7 | Prevention of contamination

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 high risk of contamination within HTS tests comes from the multiple handling steps and the use of more reagents in the sample preparation process. There is also a high likelihood of detecting contaminants in HTS tests because of their broad range and specific detection. Contamination can occur at different steps of the laboratory protocol (i.e. sampling, nucleic acid extraction, library preparation, sequencing). Sources of contamination may include sample handling, laboratory surfaces and equipment/tools contamination, reagents and carry-over (Asplund et al., 2019; Champlot et al., 2019; Dickins et al., 2014; Gaafar & Ziebell, 2020; Rosseel et al., 2014).

Contamination between successive uses of a sequencing machine (i.e. 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 owing to traces of other samples or index-hopping between samples (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).

In addition, best practices for molecular laboratories should be applied (e.g. the use of 'clean' reagents, consumables, tools and equipment, frequent changes of disposable tools and frequent cleaning of benches, equipment and tools). The EPPO Standard PM 7/98 (appendix 2; 2021) provides guidance on how to avoid contamination in molecular laboratories. The physical separation of samples suspected to contain a high concentration of target organism(s) from other samples is also highly recommended.

Despite every precaution taken, some contamination can still occur, for instance cross-contamination owing to index-hopping with pooled samples. Therefore, the level of contamination should be monitored throughout the HTS test; see Massart et al. (2022) for more details.

4.3 | Bioinformatics

The bioinformatic analysis is a key element of the HTS test as it can generate false positive and/or false negative results. It consists of a combination of successive algorithms (often referred to as a pipeline) used to analyse the raw sequencing data.

Proper bioinformatic analysis relies on the appropriate selection of the 'bioinformatic triad', corresponding to (i) the algorithm(s), (ii) its (their) parameters and thresholds and (iii) the sequence database(s). Indeed, the results generated by the pipeline depend on each of the above, including the (version of) software used, the parameters and thresholds applied, and the accuracy and completeness of sequence database(s) used for sequence comparison(s). Some bioinformatic steps do not require a sequence database and are therefore influenced by the 'bioinformatic duet' corresponding to the algorithms and their parameters/thresholds. The impact of the bioinformatic pipeline on the correct identification of target(s) has been demonstrated by Massart et al. (2019) through a test performance study with 21 plant virology laboratories analysing 10 sRNA datasets. A similar observation was made in a study of 16S rRNA gene amplicon sequencing data for the estimation of the composition of a microbiome (O'Sullivan et al., 2021).

Whatever the bioinformatic strategy, many pipelines have been developed that can operate either on a Linux system or a web interface, as well as commercial packages or user friendly open-source software. The utilization of these pipelines requires competent personnel. However, a current general trend is to simplify the use and the parameterization of these tools, making them usable without extensive bioinformatic knowledge or, sometimes, as a 'one-click' solution. For such simplified pipelines, it is paramount that the personnel implementing and using them understand their basics and their limitations and can determine the most appropriate parameters and threshold to make correct interpretations for the intended use of the HTS test. An overview of the bioinformatic steps with in-depth data processing options and their tools for the detection of plant viruses is available (Kutnjak et al., 2021) as well as reference datasets for evaluating the pipeline (Tamisier et al., 2021).

The sub-steps needed for each of the three main bioinformatic steps and their position in the analysis pipeline should be defined during test development/adaptation/optimization along with their parameters and corresponding quality metrics and thresholds (Budowle et al., 2014; Hébrant et al., 2018; Weiss et al., 2013). The order of sub-steps can be modified, depending on the bioinformatics pipeline that is used, for example the elimination of low-quality reads and nucleotides can be carried out at any time during the process. If the fixed thresholds are not met, the decision concerning repeating (parts of) the HTS test or proceeding with the bioinformatic analysis should be documented and the reason for the failure should be investigated. The bioinformatics steps are illustrated in Figure 2 (first step) and Figure 3 (second and third steps).

4.3.1 | Analysis of the raw reads

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

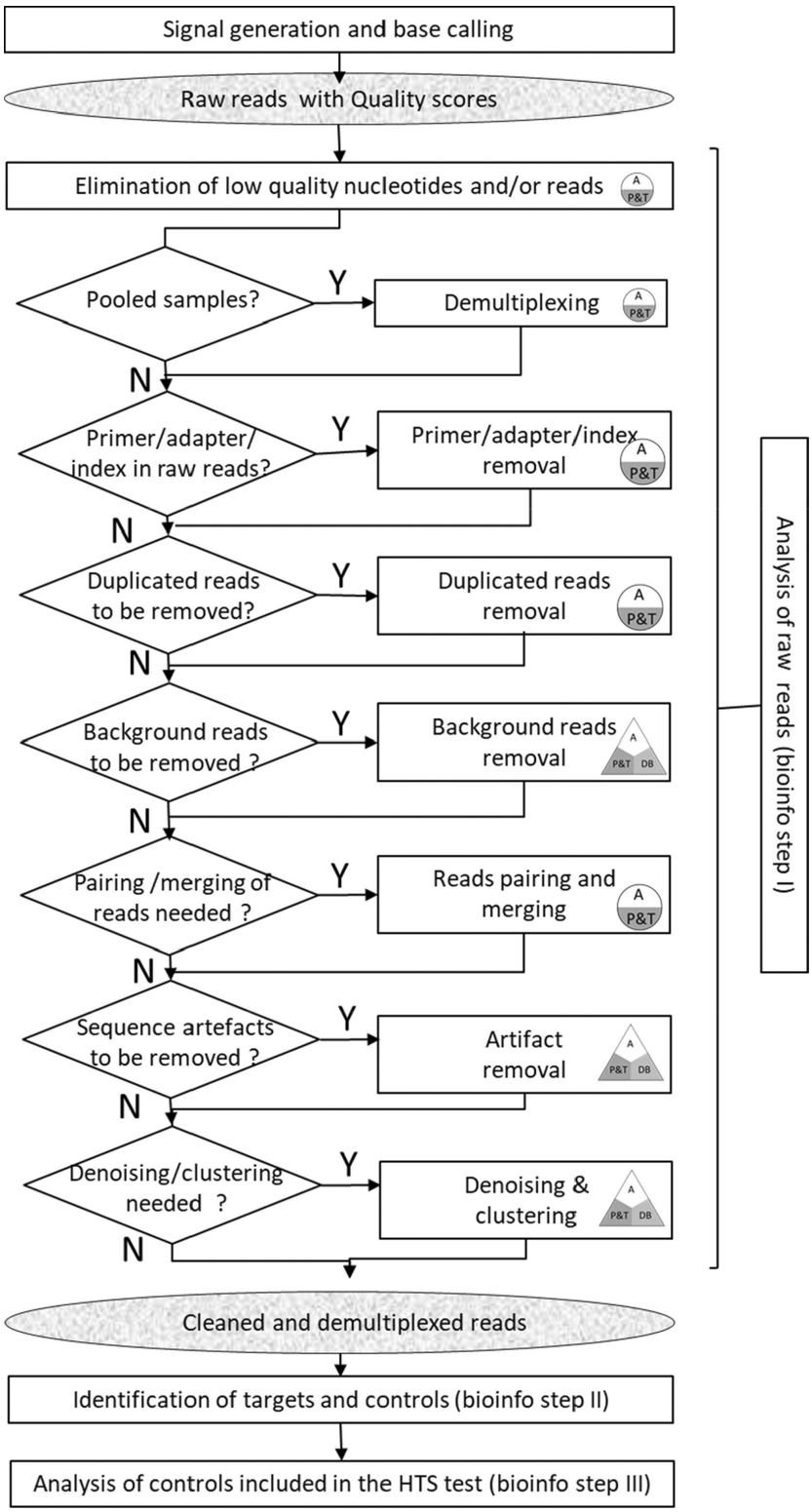

FIGURE 2 Example of the organization of the first step of the bioinformatic analyses, i.e. analysis of raw reads. The order of sub-steps can be modified, depending on the bioinformatics pipeline that is used. The rectangles correspond to operations/calculation while the grey ellipses correspond to file(s) containing sequence information and generated by the analysis. The triangle and circle represent the influence on the generated results of bioinformatics triad and duet respectively. A, Algorithm; P&T, parameters and thresholds; DB, database

run (e.g. cluster densities, quality profiles, number and length of reads) and the metrics' specifications. These metrics are platform dependent, and the most relevant metrics should be determined during the test validation with the setting of (a) minimum threshold(s) (Hébrant et al., 2018). Alternatively, the analysis of these metrics

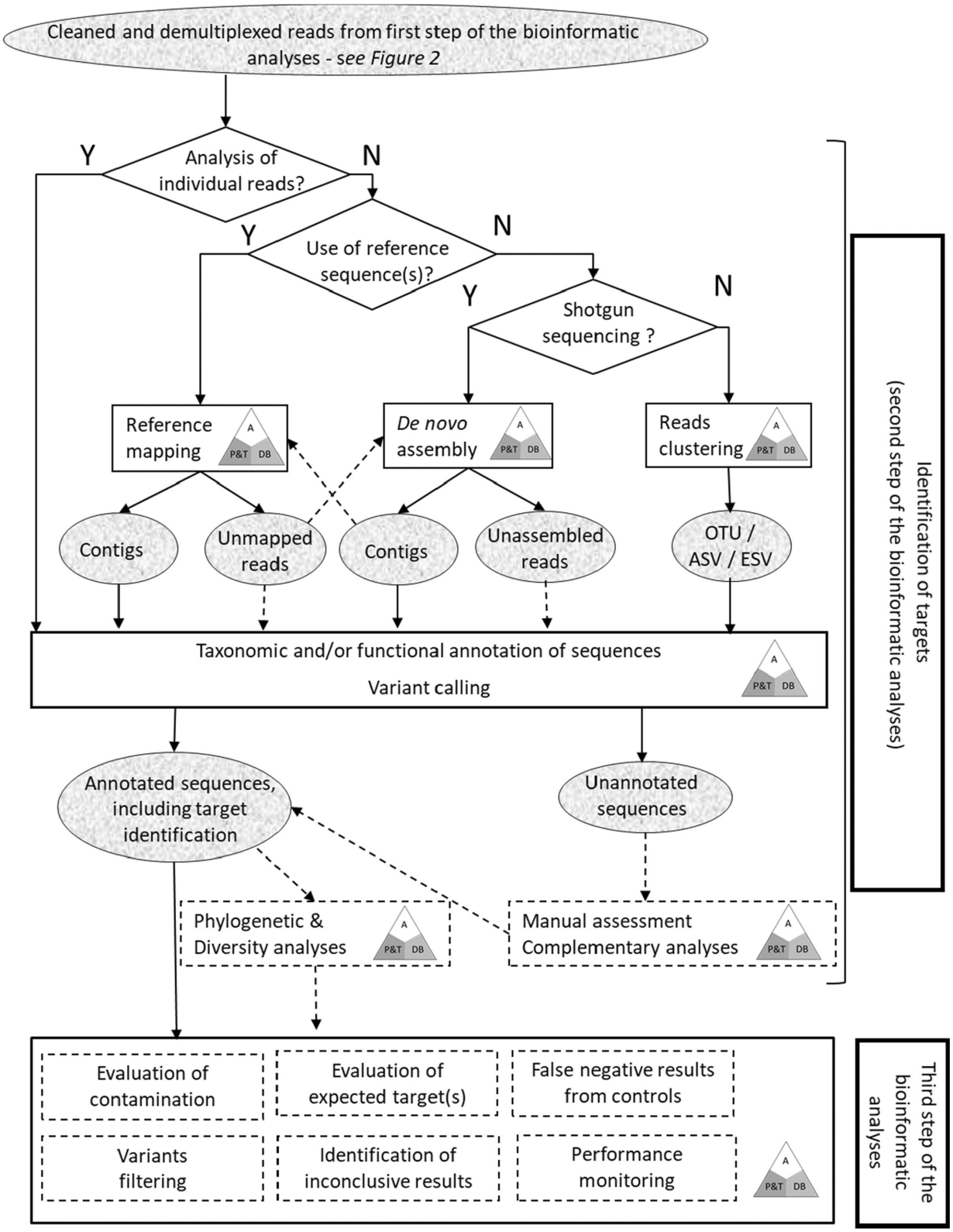

FIGURE 3 Overview of the second and third steps of the bioinformatic analyses: identification of target(s) and analysis of controls, respectively. The selection of the sub-steps depends on the bioinformatic pipeline that is used. Dashed arrows are alternative steps. The rectangles correspond to operations/calculation while the grey ellipses correspond to file(s) containing sequence information and generated by the analysis. The triangle represents the influence of the bioinformatics triad on the generated results. ASV, Amplicon sequence variants; ESV, exact sequence variants; OTU, operational taxonomic units and, for the bioinformatics triad; A, algorithm; P&T, parameters and thresholds; DB, database. More information about each of the steps can be found in the text

can be carried out after the trimming of primers, adapters and (when relevant) indexes.

Raw reads should also be analysed for their quality by checking the base quality scores (for example, Phred quality score), which depend on the sequencing platform used. Some nucleotides and/or the full sequences of reads whose quality does not meet an established threshold should be removed so that only sequences of appropriate quality are retained (Budowle et al., 2014; Hébrant et al., 2018; Weiss et al., 2013). The minimum threshold of the base quality scores should be defined during validation of the 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) and may depend on the scope of the HTS test.

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 misassignment 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). The risk of misassignment also depends on the type of indexes (e.g. long and very different indexes, single vs. dual indexes).
- *Primer, adapter and indexes removal* (also known as clipping or trimming). Primers, adapters and indexes (if used) contained 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 read removal*. Some sequences not related to the target(s), called here background reads (e.g. host sequences, ribosomal sequences, phage sequences, environmental contaminant sequences, reagent contaminants), 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 types of reads are mainly associated with shotgun sequencing strategies and their presence 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) using reference sequences and/or host control and/or no template control included in the run. The removal of background reads is particularly important when the target(s) is (are) present in low concentration(s) (Baizan-Edge et al., 2019). Caution should, however, be exercised when dealing with organisms that are capable of being completely or partially integrated in their host genome, since they may be removed during this process (e.g. pararetroviruses or unknown viruses in plants, bacteriophages in bacteria; Hohn et al., 2008; Massart et al., 2019; Sharma et al., 2017). There may also be a risk of removing the target reads during this process when high sequence similarities exist between the host and the target or if the quality of the reference genomes used for the removal of background reads is not appropriate. Some reference genomes are incomplete, and their annotations are still in progress and can contain target sequences or sequences from other organisms (i.e. endophytes or commensal organisms). Very different results can be obtained if keeping or removing background reads and therefore including such a step should be carefully considered during test development/optimization with parameters and thresholds settings based on the HTS test intended use (e.g. detection to species or genus level).
- *Duplicated read removal*. Duplicated reads originate from the same amplified fragments and have exactly identical sequences (redundancy). Their characteristics are common coordinates (e.g. the same start and end coordinates after mapping), the same sequencing direction (or mapped strand) and identical sequences. The presence of duplicated reads depends on the initial sequence complexity of the extracted nucleic acids, the library preparation procedure and the sequencing technology. They can be generated during a fragmentation or cleavage step or by an amplification-based technology (Hébrant et al., 2018; Maliogka et al., 2018). A dataset containing many duplicated reads might also be the result of a suboptimal library preparation, where too little input material was available. A high abundance of duplicated reads can limit the sensitivity of the HTS test as they can compete with low-abundance targets, despite the sequencing run itself producing a large total number of reads. It is therefore recommended to evaluate the proportion of duplicated reads. The elimination of duplicated reads depends on the HTS protocol and is not required in protocols that use the number of reads (sometimes identical) to estimate the relative abundance of a target such as amplicon sequencing for metabarcoding.
- *Merging and pairing 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 technologies, 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 pairing parameters (mismatch tolerance, minimal number of overlapping nucleotides) should take into account that the two sequences generating the consensus sequence might not be fully identical owing to sequencing errors.

- *Chimera / artefact removal*. Amplicon sequencing can generate chimeric sequences corresponding to a combination of different sequences from the original sample, leading to the formation of artefact sequences with the first part of the sequence coming from a target organism and the second part coming from another organism as a result of an amplicon accidentally acting as a primer during PCR. Similarly, whole genome amplification techniques such as multiple displacement amplification commonly used in low-input library preparation protocols for shotgun sequencing can produce chimeric sequences (Lasken & Stockwell, 2007; Quince et al., 2011). It is important to monitor and remove these sequences using appropriate tools before 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 using a nucleotide similarity threshold that ideally broadly approximates species boundaries (Mahé et al., 2015). Nevertheless, the optimal selection of threshold for clustering can vary across taxa and can result in over-clustering (grouping different species together in one cluster) or under-clustering (splitting one species into 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 to and will show a similar occurrence pattern to 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 ASVs or exact sequence variants.

4.3.2 | Identification of targets

The proper bioinformatic identification of the target(s) is important to avoid false positive (incorrect taxonomic assignment, gene annotation or variant detection) or false negative (absence of identification) results. Specific considerations related to the analytical specificity of an HTS test are developed elsewhere (Massart et al., 2022).

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, the algorithm applied and its parameters and the sequence database(s) used (see taxonomic position and functional assignment sub-steps below).
- 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 (at a predefined level) sequences on a defined portion of their length. The read assembly can be complex with very short reads (as for small RNA sequencing; Massart et al., 2019). The parameters for read assembly depend on the type of algorithm used and should be defined during test development/optimization. These include the percentage of identity between reads, the minimum overlap, the minimal length of contigs, the *k*-mer length or bubble size. For the genome sequencing of isolates of cellular organisms, the quality of assembly in contigs can be evaluated, for example by summarizing the length of the contigs using N_{50} or U_{50} values (Castro & Ng, 2017) or by comparing the contigs with related genomes and/or genes, 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). If some reads remain unused at the end of the de novo assembly process they 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 an 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 either partial or complete genome(s) (Budowle et al., 2014; Hébrant et al., 2018; Roy et al., 2018). Several reference sequences can be used for each target to account for genetic variability (Massart et al., 2019), increase the number of mapped reads and improve the annotation quality. The mapping parameters, such as number of mismatches or number/length of gaps allowed, the minimal percentage of identity or the minimal fraction of read mapping to the reference, are critical to avoid incorrect results. If the mapping parameters are less stringent, non-specific mapping to another species can happen, while too stringent mapping parameters can result in the failure to map reads from an isolate too distant from the reference sequence (Roy et al., 2018; Weiss et

al., 2013). One may also consider the inclusion of reference sequences of closely related non-target species possibly present in the sample as this could avoid possible false positive results. Important mapping result metrics include genome coverage, average read depth, the distribution of reads on the reference sequence and the 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 concentrations (Maliogka et al., 2018). The ordering of contigs along a genome (i.e. scaffolding) can improve downstream analyses such as taxonomic and/or 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 a different but closely related species mapped on the reference sequence used). In addition, the contigs generated from read assembly might need to be further independently annotated. For individual reads, clustered reads and contigs, taxonomic assignment should be determined using the latest taxonomic information, including up to date sequence-based demarcation criteria and appropriate sequence databases and software. Similarity searches performed from assembled contigs or reads using dedicated tools (e.g. AODP, BLAST, DIAMOND, EDNA, Mash, Kraken, KAIJU) provide indications on the taxonomic assignation, most often with a confidence threshold (Lambert et al., 2018; Maliogka et al., 2018; Massart et al., 2019). These similarity searches are continuously evolving (Budowle et al., 2014; Lefebvre et al., 2019; Rott et al., 2017; Ye et al., 2019). In addition to sequence similarity searches, some taxonomic classifiers, such as the RDP classifier, QIIME2's q2-feature-classifier, Metaxa2, SINTAX or TAG.ME, also consider other similar sequences in the reference sequence database and provide a confidence score using approaches such as bootstrapping. The level of certainty of similarity searches should always be retained and mentioned (e.g. E-value, bit score, bootstrap score) together with the tool and sequence database (version) used. Expert judgement may be needed to evaluate the result of a taxonomic assignment (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 complete 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, the more reliable the taxonomic assignment). This is particularly important when identifying sequences from environmental samples. Additional analyses such as phylogenetic analyses may also be required. For amplicon sequencing, the resolution of the taxonomic assignment of the OTUs or ASVs depends on different factors with the main ones being the chosen barcode, the completeness and accuracy of the reference database and the algorithm used to identify the taxonomic position. Currently barcodes are relatively short (a few hundred nucleotides), and hence can provide only a limited taxonomic resolution. A broad range of classification methods such as naive Bayesian classifiers, lowest common ancestor-based methods and phylogenetic placement methods can be used according to the type of sequence used. Combined with short barcodes, most classification methods do not lead 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 development/optimization, to verify whether the barcode is suited to detect the target organism(s) at a satisfactory taxonomic level.
- *Functional assignment*. The determination of the (potential) function of genes, the (prediction of) genomic features related to pathogenicity, resistance to antibiotics or to pesticides (Sundin & Wang, 2018), 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 (near) complete genome of pests*. Obtaining a (near) complete genome sequence may be required to validate the taxa identified, to gain information on the gene content and population diversity, or to properly resolve the epidemiology and origin of an outbreak. Obtaining (near) complete genome sequences for viruses is relatively easy because of their small genome sizes. The ability to recover a (near) complete genome becomes more complex with pests with larger genomes such as bacteria, fungi and phytoplasmas. When a (near) complete genome is required, a 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 may arise from SNP or indels or by the integration/deletion of entire genes compared with a reference sequence or compared with the consensus contigs generated (for example, the quasi-species in a virus population). The identification of

SNPs and indels relies on dedicated algorithms with specific parameters, for example strand bias, mapping quality, base calling quality (Gargis et al., 2015; Roy et al., 2018; Weiss et al., 2013). Variant identification is important for some applications as it can impact the pathogenicity of an organism (e.g. pathogenicity island for bacteria, resistance breaking mutations) or indicate the presence of a divergent isolate of a species.

- *Unused quality reads*. A number of reads that have passed all of the quality checks may still not be assembled, mapped or annotated after the various bioinformatic analyses. These reads, also called unused reads or unmapped reads, can be gathered as a separate output during the analysis and their number or proportion calculated. Depending on the purpose of the HTS 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 the presence of unforeseen organisms among them. Some individual sequences or some contigs may still not be annotated after a second round 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 databases allow their annotation (Gasc et al., 2015; Solden et al., 2016).

4.3.3 | Analysis of controls

The third and last step of the bioinformatic analysis is important to identify potential false positive and/or negative results. False negative results can come from several origins, for example because the target concentration in the plant extract is below the limit of detection (see the special section on analytical sensitivity elsewhere; Massart et al., 2022) or because of sample degradation, the inhibition of enzymatic reactions or the generation of an insufficient number of reads or an inappropriate bioinformatic triad.

False positive results may come from contamination during different sub-steps of the laboratory phase or an inappropriate bioinformatic triad. They may be due to improper handling of samples and/or pooling of libraries. To address false positive and/or negative results, different controls can be included to monitor the different stages of an HTS test (see Massart et al., 2022 for details). The origin of false positive and/or false negative results should be investigated and addressed, and the decision on whether to repeat (parts of) the HTS test should be documented.

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

- *Evaluation of contamination*. Although the contamination rate has been decreasing with the improvement of laboratory protocols and sequencing platforms, there is still a need to monitor it, both qualitatively and quantitatively. To check for contamination at different stages of the HTS test, different controls (e.g. alien controls, negative controls, positive controls and internal control) can be used (Massart et al., 2022).
- *Evaluation of the ability to detect (the) expected target(s)*. This can be carried out using appropriate controls (see Massart et al., 2022 for details). 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 relevant reference sequences).
- *False negative results from controls*. False negative results can be expected when one of the targets from the control(s) (Massart et al., 2022) is not detected in the sequence data. The result metrics for reference mapping 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, false variants generated owing 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. 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 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 dataset can be used to check whether 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. Independently of the laboratory's decision, the process should be documented as part of quality assurance.
- *Performance monitoring*. The performance of HTS tests may be checked routinely by including appropriate controls (see Massart et al., 2022). For example, for HTS tests used for the detection of quarantine pests, a positive control close to the limit of detection should be included in each sequencing run and the control results monitored over time.

5 | CONCLUDING REMARKS

HTS technologies have brought a unique opportunity to improve the detection of any pest present in any sample without any previous information. They have been adopted in research for plant pest detection and

identification for more than a decade and have led to significant advances, including the discovery of previously unknown pests or the detection of pests, sometimes unexpectedly, during quarantine or post-entry quarantine evaluation. These advances, combined with the cost reduction of these technologies and their improvement in reliability, now create a momentum for their progressive adoption by plant health diagnostics laboratories.

However, adopting HTS technologies for plant pest detection and identification represents a disruptive transition raising complex challenges for any laboratory. The biggest challenge corresponds to the bioinformatic component of HTS test, most specifically the need to properly manage the huge quantity of generated sequence data and the complexity of the calculations required for their analysis. Plant health laboratories are not used to handling such large amounts of information, and this requires completely new skills, equipment and protocols. High-throughput sequencing testing therefore requires a strong investment in information management technologies and new expertise. For the laboratory component, most protocols for HTS testing rely on classical molecular biology reactions: fragmentation of nucleic acids, ligation of nucleic acids, end-repair, reverse transcription, amplification, etc. There, the challenges correspond to the number of steps needed which is far higher than for other molecular tests such as (RT-)PCR or LAMP. Behind the analytical part, adopting HTS testing will also have an impact on all the support processes of a laboratory (quality management, purchase of reagents or service, information technology, human resource management, etc.).

In this context, the present paper describes general and technical recommendations for a laboratory, active in plant health diagnostics but also in research, to prepare for the adoption of HTS testing. These recommendations identify the key elements to take into account for the laboratory and bioinformatic components and provide a guide for preparing to develop and practically implement HTS testing.

The complexity of the HTS process with its continuously evolving technologies requires guidelines with enough flexibility to remain up to date and, at the same time, that provide sufficient information to support laboratories embarking on the setup of HTS diagnostics. The present guidelines have been also designed in such a way that they can be applied to the detection and identification of any plant pest from any type of matrix.

The recommendations described in this paper are complemented by a second paper (Massart et al., 2022) which describes guidelines for the reliable use of HTS testing by a laboratory. Both publications were written in the frame work of the EU-funded VALITEST project (grant no. 773 139) and aim to be the basis of international standards on HTS testing. Internationally accepted guidelines are indeed essential to adopt HTS tests for plant pest diagnostics, to facilitate trade and to increase confidence in the results and their interpretation, and this work was used to draft an EPPO standard currently in the approval stage.

FUNDING INFORMATION

This article is based upon work from the work package 2 of the project VALITEST (https://www.valit est.eu/), supported by the European Union's Horizon 2020 research and innovation programme under grant agreement no. 773139.

REFERENCES

Abdelfattah A, Sanzani SM, Wisniewski M, Berg G, Cacciola SO, Schena L (2019) Revealing cues for fungal interplay in the plant-air interface in vineyards. *Frontiers in Plant Science* 10, 922.

Adams IP, 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.

Adams IP, 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.

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.

Ahmed M, Back MA, 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.

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.

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.

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.

Aritua V, Musoni A, Kabeja A, Butare L, Mukamuhirwa F, Gahakwa D, Kato F, Abang MM, Buruchara R, Sapp M, Harrison J, Studholme DJ, 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.

Asplund M, Kjartansdóttir KR, Mollerup S, Vinner L, Fridholm H, Herrera JAR, Friis-Nielsen J, Hansen TA, Jensen RH, Nielsen IB, Richter SR, Rey-Iglesia A, Matey-Hernandez ML, Alquezar-Planas DE, Olsen PVS, Sicheritz-Pontén T, Willerslev E, Lund O, Brunak S, Mourier T, Nielsen LP, Izarzugaza JMG, Hansen AJ (2019) Contaminating viral sequences in high-throughput sequencing viromics: a linkage study of 700 sequencing libraries. *Clinical Microbiology and Infection* 25(10), 1277–1285.

Aziz N, Zhao Q, Bry L, Driscoll DK, Funke B, Gibson JS, Grody WW, Hegde MR, Hoeltge GA, Leonard DGB, Merker JD, Nagarajan R, Palicji LA, Robetorye RS, Schrijver I, Weck KE, Voelkerding KV (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.

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(3), 533–542.

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

Batovska J, Piper AM, Valenzuela I, Cunningham JP, Blacket MJ (2021) Developing a non-destructive metabarcoding protocol for detection of pest insects in bulk trap catches. *Scientific Reports* 11, 7946.

Bester R, Cook G, Breytenbach JH, Steyn C, De Bruyn R, Maree HJ (2021) Towards the validation of high-throughput sequencing (HTS) for routine plant virus diagnostics: measurement of variation linked to HTS detection of citrus viruses and viroids. *Virology Journal* 18(1), 1–9.

Bonants PJM, van Gent-Pelzer MPE, van Leeuwen GCM, van der Lee TAJ (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.

Braukmann TWA, Ivanova NV, Prosser SWJ, Elbrecht V, Steinke D, Ratnasingham S, de Waard JR, Sones JE, Zakhhharov EV, Hebert PDN (2019) Metabarcoding a diverse arthropod mock community. *Molecular Ecology Resources* 19, 711–727.

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 BB, Matthijnssens J, Deboutte W, Ellis RJ, 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 VM, Drosten C, Jones TC, Hendriksen RS, Aarestrup FM, 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.

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.

Budowle B, Donnell ND, Bielecka-Oder A, Colwell RR, Corbett CR, Fletcher J, Forsman M, Kadavy DR, Markotic A, Morse SA, Murch RS, Sajantila A, Schmedes SE, Ternus KL, Turner SD, Minot S (2014) Validation of high throughput sequencing and microbial forensics applications. *Investigative Genetics*, 5, 9.

Buschmann T, Zhang R, Brash DE, Bystrykh LV (2014). Enhancing the detection of barcoded reads in high throughput DNA sequencing data by controlling the false discovery rate. *BMC Bioinformatics* 15, 264.

Cai W, Nunziata S, Rascoe J, Stulberg MJ (2019). SureSelect targeted enrichment, a new cost effective method for the whole genome sequencing of Candidatus Liberibacter asiaticus. *Scientific reports* 9(1), 1–8.

Candresse T, Filloux D, Muhire B, Julian C, Galzi S, Fort G, Bernardo P, Daugrois JH, Fernandez E, Martin DP, 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.

Carew ME, Coleman RA, Hoffmann AA (2018) Can non-destructive DNA extraction of bulk invertebrate samples be used for metabarcoding? *PeerJ* 6, e4980.

Castro CJ, Ng TFF (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.

Catara V, Cubero J, Pothier JF, Bosis E, Bragard C, Đermić E, Holeva MC, Jacques MA, Petter F, Pruvost O, Robène I, Studholme DJ, Tavares F, Vicente JG, Koebnik R, Costa J (2021). Trends in molecular diagnosis and diversity studies for phytosanitary regulated Xanthomonas. *Microorganisms* 9(4), 862.

Cella E, Angeletti S, Fogolari M, Bazzardi R, de Gara L, Ciccozzi M (2018) Two different *Xylella fastidiosa* strains circulating in Italy: phylogenetic and evolutionary analyses. *Journal of Plant Interactions* 13, 428–432.

Champlot S, Berthelot C, Pruvost M, Bennett EA, Grance T, Geigl E-M (2019) An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. *PLoS ONE* 5(9), e13042.

Chandelier A, Hulin J, San Martin G, Debode F, Massart S (2021) Comparison of qPCR and metabarcoding methods as tools for the detection of airborne inoculum of forest fungal pathogens. *Phytopathology* 111(3) 570–81.

Davis MPA, van Dongen S, Abreu-Goodger C, Bartonicek N, Enright AJ (2013) Kraken: a set of tools for quality control and analysis of high-throughput sequence data. *Methods* 63, 41–49.

Davis JJ, Boisvert S, Brettin T, Kenyon RW, Mao C, Olson R, Overbeek R, Santerre J, Shukla M, Wattam AR, Will R, Xia F, Stevens R (2016) Antimicrobial resistance prediction in PATRIC and RAST. *Scientific Reports* 6, 27030.

Del Fabbro C, Scalabrin S, Morgante M, Giorgi FM (2013) An extensive of read trimming effects on Illumina NGS data analysis. *PLoS ONE* 8(12), e85024.

Denman S, Doonan J, Ransom-Jones E, Broberg M, Plummer S, Kirk S, Scarlett K, Griffiths AR, Kaczmarek M, Forster J, Peace A, Golyshin PN, Hassard F, Brown N, Kenny JG, McDonald JE (2018) Microbiome and infectivity studies reveal complex polyspecies tree disease in Acute oak decline. *The ISME Journal* 12, 386–399.

Dickins B, Rebolledo-Jaramillo B, Su MS-W, Paul IM, Blankenberg D, Stoler N, Makova KD, Nekrutenko A (2014) Controlling for contamination in re-sequencing studies with a reproducible web-based phylogenetic approach. *Biotechniques* 56(3), 134–141.

Dormontt EE, Van Dijk KJ, Bell KL, Biffin E, Breed MF, Byrne M, Caddy-Retalic S, Encinas-Viso F, Nevill PG, Shapcott A, Young JM (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.

Elbrecht V, Braukmann TWA, Ivanova NV, Prosser SWJ, Hajibabaei M, Wright M, Zakharov EV, Hebert PDN, Steinke D (2019) Validation of COI metabarcoding primers for terrestrial arthropods. *PeerJ* 7, e7745.

EPPO PM 7/77 (3) (2019) Documentation and reporting on a diagnosis. *EPPO Bulletin* 49(3), 527–529.

EPPO PM 7/98 (5) (2021) Specific requirements for laboratories preparing accreditation for a plant pest diagnostic activity. *EPPO Bulletin* 51; 468–498.

EPPO PM7/129 (2) (2021) DNA barcoding as an identification tool for a number of regulated pests. *Bulletin OEPP/EPPO Bulletin* 51(1), 100–143.

Fox A, Fowkes AR, Skelton A, Harju V, Buxton-Kirk A, Kelly M, Forde SMD, 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.

Franco Ortega S, Ferrocino I, Adams I, Silvestri S, Spadaro D, Gullino ML, Boonham N (2020) Monitoring and surveillance of aerial mycobiota of rice paddy through DNA metabarcoding and qPCR. *Journal of Fungi* 6(4), 372. https://doi.org/10.3390/jof6040372

Gaafar YZA, 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.

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.

Gargis AS, Kalman L, Bick DP, da Silva C, Dimmock DP, Funke BH, Gowrisankar S, Hegde MR, Kulkarni E, Mason CE, Nagarajan R, Voelkerding KV, Worthey DA, Aziz N, Barnes J, Bennett SF, Bisht H, Church FM, Dimitrova Z, Gargis SR, Hafez N, Hambuch T, Hyland FCL, Lunna RA, MacCannell D, Mann T, McCluskey MR, McDaniel TK, Ganova-Raeva LM, Rehm HL, Reid J, Campo DS, Resnick RB, Ridge PG, Salit ML, Skums P, Wong L-JC, Zehnbauer BA, Zook JM, Lubin IM (2015) Good laboratory practice for clinical next-generation sequencing informatics pipelines. *National Biotechnologies* 33(7), 689–693.

Gasc C, Ribière C, Parisot N, Beugnot R, Defois C, Petit-Biderre C, Boucher D, Peyretaillade E, Peyret P (2015) Capturing prokaryotic dark matter genomes. *Research in microbiology* 166(10), 814–830.

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

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

Hamelin RC, Roe AD (2019) Genomic biosurveillance of forest invasive alien enemies: a story written in code. *Evolutionary Applications* 13(1), 95–115.

Hanshew AS, Mason CJ, Raffa KF, Currie CR (2013) Minimization of chloroplast contamination in 16S rRNA gene pyrosequencing of insect herbivore bacterial communities. *Journal of Microbiological Methods* 95, 149–155.

Haro C, Anguita-Maeso M, Metsis M, Navas-Cortés JA, Landa BB (2021) Evaluation of established methods for DNA extraction and primer pairs targeting 16S rRNA gene for bacterial microbiota profiling of olive xylem sap. *Frontiers in plant science* 12, 296.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences* 270, 313–321.

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.

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

IPPC Secretariat (2019) Preparing to use high-throughput sequencing (HTS) technologies as a diagnostic tool for phytosanitary purposes. Commission on Phytosanitary Measures Recommendation No. 8. Rome. Published by FAO on behalf of the Secretariat of the International Plant Protection Convention (IPPC). https://www.ippc.int/en/publications/87199/ [last accessed on 10 October 2021].

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.

Katsiani A, Maliogka VI, Katis N, Svanella-Dumas L, Olmos A, Ruiz-García AB, 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.

Kikuchi T, Cotton JA, Dalzell JJ, Hasegawa K, Kanzaki N, McVeigh P, Takanashi T, Tsai IJ, Assefa SA, Cock PJA, Otto TD, Hunt M, Reid AJ, Sanchez-Flores A, Tsuchihara K, Yokoi T, Larsson MC, Miwa J, Maule AG, Sahashi N, Jones JT, Berriman M (2011) Genomic insights into the origin of parasitism on the emerging plant pathogen *Bursaphelenchus xylophilus*. *PLoS Pathogens* 7(9), e1002219.

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

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.

Kutnjak D, Tamisier L, Adams I, Boonham N, Candresse T, Chiumenti M, De Jonghe K, Kreuze JF, Lefebvre M, Silva G, Malapi-Wight M, Margaria P, Mavrič Pleško I, McGreig S, Miozzi L, Remenant B, Reunard JS, Rollin J, Rott M, Schumpp O, Massart S, Haegeman A (2021) A primer on the analysis of high-throughput sequencing data for detection of plant viruses. *Microorganisms* 9(4), 841.

Laehnemann D, Borkhardt A, McHardy AC (2016) Denoising DNA deep sequencing data—high-throughput sequencing errors and their correction. *Briefings in Bioinformatics* 17(1), 154–179.

Lambert C, Braxton C, Charlebois RL, Deyati A, Duncan P, La Neve F, Malicki HD, Ribrioux S, Rozelle DK, Michaels B, Sun W, Yang Z, Khan AS (2018) Considerations for optimisation of high-throughput sequencing bioinformatic pipelines for virus detection. *Viruses* 10, 528.

Lasken RS, Stockwell TB (2007) Mechanism of chimera formation during the multiple displacement amplification reaction. *BMC Biotechnology* 7, 19.

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*, 3(4), 256–259.

Leifert WR, Glatz RV, Siddiqui MS, Collins SR, Taylor PW, 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 [last accessed on 28 February 2022].

Lu N, Li J, Bi C, Guo J, Tao Y, Luan K, Tu J, Lu Z (2019) Chimera Miner: an improved chimeric read detection pipeline and its application in single cell sequencing. *International Journal of Molecular Sciences* 20(8), 1953.

Lundberg DS, Yourstone S, Mieczkowski P, Jones CD, Dangl JL (2013) Practical innovations for high-throughput amplicon sequencing. *Nature Methods* 10, 999–1002.

MacConaill LE, Burns RT, Nag A, Coleman HA, Slevin MK, Giorda K, Light M, Lai K, Jarosz M, McNeill MS, Ducar MD, Meyerson M, Thorner AR (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.

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

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

Malapi-Wight M, Adhikari B, Zhou J, Hendrickson L, Maroon-Lango CJ, McFarland C, Foster JA, Hurtado-Gonzales OP (2021) HTS-based diagnostics of sugarcane viruses: seasonal variation and its implications for accurate detection. *Viruses* 13(8), 1627.

Maliogka VI, Minafra A, Saldarelli P, Ruiz-García AB, 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.

Maljkovic Berry I, Melendrez MC, Bishop-Lilly KA, Rutvisuttinunt W, Pollett S, Talundzic E, Morton L, Jarman RG (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.

Marquina D, Ronquist F, Łukasik P (2021) The effect of ethanol concentration on the morphological and molecular preservation of insects for biodiversity studies. *PeerJ* 9, e10799.

Martoni F, Nogarotto E, Piper AM, Mann R, Valenzuela I, Eow L, Rako L, Rodoni BC, Blacket MJ (2021) Propylene glycol and non-destructive DNA extractions enable preservation and isolation of insect and hosted bacterial DNA. *Agriculture* 11(1), 77.

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.

Massart S, Candresse T, Gil J, Lacomme C, Predajna L, Ravnikar M, Reynard J-S, Rumbou A, Saldarelli P, Škorić D, Vainio EJ, Valkonen JPT, 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.

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 HJ, Olivier T, Olmos A, Pooggin MM, Reynard J-S, Ruiz-García AB, Safarova D, Schneeberger PHH, Sela N, Turco S, Vainio EJ, 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. *Phytopathology* 109, 488–497.

Massart S, Adams I, Al Rwahnih M, Baeyen S, Bilodeau G, Blouin A, Boonham N, Bruinsma M, Candresse T, Chandelier A, De Jonghe K, Fox A, Gaafar Y, Gentit P, Haegemans A, Ho W, Hurtado-Gonzales O, Jonkers W, Kreuze J, Kutjnak D, Landa B, Leite Vicente C, Liu M, Maclot F, Malapi-Wight M, Maree H, Martoni F, Mehle N, Minafra A, Mollov D, Moreira A, Nakhla M, Petter F, Piper A, Ponchart J, Rae R, Remenant B, Rivera Y, Rodoni B, Roenhorst A, Rollin J, Saldarelli P, Santala J, Souza-Richards R, Spadaro D, Studholme D, Sultmanis S, van der Vlugt R, Tamisier L, Trontin C, Van Vaerenbergh J, Wetzel T, Ziebell H, Lebas BSM (2022) Guidelines for the reliable use of high throughput sequencing technologies to detect plant pathogens and pests. Manuscript in preparation.

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.

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.

McInerney P, Adams P, Hadi MZ (2014) Error rate comparison during polymerase chain reaction by DNA polymerase. *Molecular Biology International* 2014, 287430.

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.

Moreau C, Wray B, Czekanski-Moir J, Rubin B (2013) DNA preservation: a test of commonly used preservatives for insects. *Invertebrate Systematics* 27, 81.

Nicolaisen M, West JS, Sapkota R, Canning GGM, Schoen C, Justesen AF (2017) Fungal communities including plant pathogens in near surface air are similar across Northwestern Europe. *Frontiers in Microbiology* 8, 1729.

Nielsen M, Gilbert MTP, 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.

Nilsson RH, Anslan S, Bahram M, Wurzbacher C, Baldrian P, Tedersoo L (2019) Mycobiome diversity: high-throughput sequencing and identification of fungi. *Microbiology* 17, 95–109.

Núñez A, Amo de Paz G, Ferencova Z, Rastrojo A, Guantes R, García AM, Alcami A, Montserrat Gutiérrez-Bustillo A, Moreno DA (2017) Validation of the hirst-type spore trap for simultaneous monitoring of prokaryotic and eukaryotic biodiversity in urban air samples by NGS. *Applied and Environmental Microbiology*, 83(13), e00472-17.

Olmos A, Boonham N, Candresse T, Gentit P, Giovani B, Kutnjak D, Liefting L, Maree HJ, Minafra A, Moreira A, Nakhla MK, Petter F, Ravnikar M, Rodoni B, Roenhorst JW, Rott M, Ruiz-García AB, 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), 219–24.

O'Sullivan DM, Doyle RM, Temisak S, Redshaw N, Whale AS, Logan G, Huang J, Fischer N, Amos GC, Preston MD, Marchesi JR (2021). An inter-laboratory study to investigate the impact of the bioinformatics component on microbiome analysis using mock communities. *Scientific Reports* 11(1), 1–14.

Ovaskainen O, Abrego N, Somervuo P, Palorinne I, Hardwick B, Pitkänen J-M, Andrew NR, Niklaus PA, Schmidt NM, Seibold S, Vogt J, Zakharov EV, Hebert PDN, Roslin T, Ivanova NV (2020) Monitoring fungal communities with the global spore sampling project. *Frontiers in Ecology and Evolution* 7, 511.

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.

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.

Pantaleo V, Chiumenti M (Eds.) (2018) *Viral Metagenomics*. Springer New York. https://doi.org/10.1007/978-1-4939-7683-6.

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW (2014) Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research* 25, 1043–1055.

Piombo E, Abdelfattah A, Droby S, Wisniewski M, Spadaro D, Schena L (2021) Metagenomics approaches for the detection and surveillance of emerging and recurrent plant pathogens. *Microorganisms* 9(1), 188.

Piper AM, Batovska J, Cogan NOI, Weiss J, Cunningham JP, Rodoni BC, Blacket MJ (2019) Prospects and challenges if implementing DNA metabarcoding for high-throughput insect surveillance. *GigaScience* 8, 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.

Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK (2016) Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Analytical Methods* 8, 12–24.

Quail MA, Smith M, Jackson D, Leonard S, Skelly T, Swerdlow HP, Gu Y, Ellis P (2014) SASI-Seq: sample assurance Spike-Ins, and highly differentiating 384 barcoding for Illumina sequencing. *BMC Genomics* 15, 110.

Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ (2011) Removing noise from pyrosequenced amplicons. *BMC bioinformatics* 12(1), 38.

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

Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, Friez MJ, Funke BH, Hedge MR, Lyon Y (2013) ACMG clinical laboratory standards for next-generation sequencing. *Genetics in Medicine* 15(9), 733–747.

Ritari J, Salojärvi J, Lahti, L, de Vos WM (2015) Improved taxonomic assignment of human intestinal 16S rRNA sequences by a dedicated reference database. *BMC Genomics* 16, 1056.

Ritter CD, Häggqvist S, Karlsson D, Sääksjärvi IE, Muasya AM, Nilsson RH, 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.

Rivarez MPS, Vučurović A, Mehle N, Ravnikar M, Kutnjak D (2021) Global advances in tomato virome research: current status and the impact of high-throughput sequencing. *Frontiers in Microbiology* 12, 671925.

Robinson CV, Porter TM, Wright MTG, Hajibabaei M (2021) Propylene glycol-based antifreeze is an effective preservative for DNA metabarcoding of benthic arthropods. *Freshwater Science* 40(1), 77–87.

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.

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, https://doi.org/10.1094/PDIS-03-17-0306-RE.

Roy S, Coldren C, Karunamurthy A, Kip NS, Klee EW, Lincoln SE, Leon A, Pullambhatla M, Temple-Smolkin RL, Voelkerding KV, Wang C, Carter AB (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.

Sahlin K, Chikhi R, Arvestad L (2016) Assembly scaffolding with PE-contaminated mate-pair libraries. *Bioinformatics* 32(13), 1925–1932.

Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW (2014) Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology* 12, 87.

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.

Seppey M, Manni M, Zdobnov EM (2019) BUSCO: assessing genome assembly and annotation completeness. *In*: Kollmar (eds.) *Gene Prediction. Methods in Molecular Biology*, Humana, New York, vol. 1962, 227–245.

Sharma S, Chatterjee S, Datta S, Prasad R, Dubey D, Prasad RK, Vairale MG (2017) Bacteriophages and its applications: an overview. *Folia Microbiology* 62, 17–55.

Simpson AJG, Reinach FC, Arruda P, Abreu FA, Acencio M, Alvarenga R, Alves LMC, Araya JE, Baia GS, Baptista CS, Barros MH, Bonaccorsi ED, Bordin S, Bové JM, Briones MRS, Bueno MRP, Camargo AA, Camargo LEA, Carraro DM, Carrer H, Colauto NB, Colombo C, Costa FF, Costa MCR, Costa-Neto CM, Coutinho LL, Cristofani M, Dias-Neto E, Docena C, El-Dorry H, Facincani AP, Ferreira AJS, Ferreira VCA, Ferro JA, Fraga JS, França SC, Franco MC, Frohme M, Furlan LR, Garnier M, Goldman GH, Goldman MHS, Gomes SL, Gruber A, Ho PL, Hoheisel JD, Junqueira ML, Kemper EL, Kitajima JP, Krieger JE, Kuramae EE, Laigret F, Lambais MR, Leite LCC, Lemos EGM, Lemos MVF, Lopes SA, Lopes CR, Machado JA, Machado MA, Madeira AMBN, Madeira HMF, Marino CL, Marques MV, Martins EAL, Martins EMF, Matsukuma AY, Menck CFM, Miracca EC, Miyaki CY, Monteiro-Vitorello CB, Moon DH, Nagai MA, Nascimento ALTO, Netto LES, Nhani Jr A, Nobrega FG, Nunes LR, Oliveira MA, de Oliveira MC, de Oliveira RC, Palmieri DA, Paris A, Peixoto BR, Pereira GAG, Pereira Jr HA, Pesquero JB, Quaggio RB, Roberto PG, Rodrigues V, De M Rosa AJ, De Rosa Jr VE, De Sá RG, Santelli RV, Sawasaki HE, Da Silva ACR, Da Silva AM, Da Silva FR, Silva Jr WA, Da Silveira JF, Silvestri MLZ, Siqueira WJ, De Souza AA, De Souza AP, Terenzi MF, Truffi D, Tsai SM, Tsuhako MH, Vallada H, Van Sluys MA, Verjovski-Almeida S, Vettore AL, Zago MA, Zatz M, Meidanis J, Setubal JC (2000) The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature* 406, 151–159.

Solden L, Lloyd K, Wrighton K (2016) The bright side of microbial dark matter: lessons learned from the uncultivated majority. *Current Opinion in Microbiology* 31, 217–226.

Soltani N, Stevens KA, Klaassen V, Hwang MS, Golino DA, Al Rwahnih M (2021) Quality assessment and validation of high-throughput sequencing for grapevine virus diagnostics. *Viruses* 13, 1130.

Sundin GW, Wang N (2018) Antibiotic resistance in plant-pathogenic bacteria. *Annual Review of phytopathology* 56, 161–80.

Tamisier L, Haegeman A, Foucart Y, Fouillien N, Al Rwahnih M, Buzkan N, Candresse T, Chiumenti M, De Jonghe K, Lefebvre M, Margaria P, Reynard JS, Stevens K, Kuxtnjak D, Massart, S (2021). Semi-artificial datasets as a resource for validation of bioinformatics pipelines for plant virus detection. *Peer Community Journal* 1, https://doi.org/10.24072/pcjournal.62 [last accessed on 28 February 2022].

Taylor GS, Martoni F (2020) 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.

Tremblay ED, Duceppe M-O, Bérubé JA, Kimoto T, Lemieux C, Bilodeau GJ (2018) Screening for exotic forest pathogens to increase survey capacity using metagenomics. *Phytopathology* 108(12), 1509–1521.

Tremblay ED, Duceppe M-O, Thurston GB, Gagnon M-C, Côté M-J, Bilodeau GJ (2019) High-resolution biomonitoring of plant pathogens and plant species using metabarcoding of pollen pellet contents collected from a honeybee hive. *Environmental DNA* 1, 155–175.

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* 20(5), 1171–1181.

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.

Villamor DEV, Keller KE, Martin R, Tzanetakis I E (2021) Comparison of high throughput sequencing to standard protocols for virus detection in berry crops. *Plant Disease*, https://doi.org/10.1094/PDIS-05-21-0949-RE [last accessed on 13 October 2021].

Vink CJ, Thomas SM, Paquin P, Hayashi CY, Hedin M (2005). The effects of preservatives and temperatures on arachnid DNA. *Invertebrate Systematics* 19, 99–104.

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.

Weiss MM, Van der Zwaag B, Jongbloed JDH, Vogel MJ, Bruggenwirth HT, Lekanne Derez RH, Mook O, Ruivenkamp CAL, 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.

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.

Whitehurst LE, Cunard CE, Reed JN, Worthy SJ, Marsico TD, Lucardi RD, Burgess KS (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* 22(5), 1585–1606.

Wilcox TM, Zarn KE, Piggott MP, Young MK, McKelvey KS, Schwartz MK (2018) Capture enrichment of aquatic environmental DNA: a first proof of concept. *Molecular Ecology Resources* 18(6), 1392–1401.

Wright ES, Vetsigian KH (2016). Quality filtering of Illumina index reads mitigates sample cross-talk. *BMC Genomics* 17, 876.

Xu J, Wang N (2019) Where are we going with genomics in plant pathogenic bacteria? *Genomics* 111(4), 729–36.

Yang X, Chockalingam SP, Aluru S (2012) A survey of error-correction methods for next-generation sequencing. *Briefings in bioinformatics* 14(1), 56–66.

Ye SH, Siddle KJ, Park DJ, Sabeti PC (2019) Benchmarking metagenomics tools for taxonomic classification. *Cell* 178, 779–794.

Zhao L, Zhang H, Kohnen MV, Prasad KV, Gu L, Reddy AS (2019) Analysis of transcriptome and epitranscriptome in plants using PacBio Iso-Seq and nanopore-based direct RNA sequencing. *Frontiers in Genetics* 10, 253.

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 graminearum*. *Scientific Reports* 5, 8248.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Lebas, B., Adams, I., Al Rwahnih, M., Baeyen, S., Bilodeau, G.J. & Blouin, A.G. et al. (2022) Facilitating the adoption of high-throughput sequencing technologies as a plant pest diagnostic test in laboratories: A step-by-step description. *EPPO Bulletin*, 52, 394–418. Available from: https://doi.org/10.1111/epp.12863