CrossMark

SI: PLANT PATHOLOGY FOR INNOVATIVE AGROECOLOGY

The impact of high throughput sequencing on plant health diagnostics

Ian P. Adams · Adrian Fox · Neil Boonham · Sébastien Massart ·
Kris De Jonghe

Abstract High throughput sequencing informed diagnostics is revolutionising plant pathology. The application of this technology is most advanced in plant virology, where it is already becoming a front-line diagnostic tool and it is envisaged that for other types of pathogen and pests this will be the case in the near future. However, there are implications to deploying this technology due to a number of technical and scientific challenges. Firstly, interpretation of data and the assessment of plant health risk against a limited baseline of existing knowledge of the presence of pathogens in a given geographic region. Secondly, evidence of causality and the separation of pathogenic from commensal organisms in the sequence data, thirdly, the tension between the generation of a rapid sequence result with the necessary but laborious epidemiological characterisation in support of plant health risk assessment. Finally, the validation and accreditation of methods based on this rapidly evolving technology. These in turn present challenges for plant health policy and regulation. This review discusses the development of this technology, its application in plant health diagnostics, and explores the implications of applying this technology in the plant health setting.

Keywords High throughput sequencing · NGS ·
Diagnostics · Validation · Accreditation

I. P. Adams (✉) · A. Fox
Fera Science Ltd, Sand Hutton, York YO41 1LZ, UK
e-mail: Ian.adams@fera.co.uk

N. Boonham
IAFRI, Newcastle University, Newcastle upon Tyne, Tyne and
Wear NE1 7RU, UK

S. Massart
Laboratory of Integrated and Urban Phytopathology, TERRA,
Gembloux AgroBio Tech, Université de Liège, 5030 Gembloux,
Belgium

K. De Jonghe
Plant Sciences Unit, Flanders Research Institute for Agriculture,
Fisheries and Food (ILVO), Burgemeester Van Gansberghelaan
96, 9820 Merelbeke, Belgium

Introduction

Many diseases and pests can cause serious damage in plants, and thus result in enormous economic losses in agriculture and horticulture. Some of these organisms are widespread while others are still restricted to specific regions in the world and plant health describes legislative measures taken to control the spread of the latter into new regions. National Plant Protection Organisations (NPPOs) implement and enforce the legislation by controlling import and movement of plants and plant materials based on the outcomes of risk assessments. NPPOs are frequently looking to improve the availability of diagnostic tools and in some situations to improve the sensitivity, specificity, reliability and cost effectiveness of analysis. The rational use of these improved technologies (primarily based on serology and DNA amplification methods) has progressively improved our ability to efficiently detect and identify plant pests

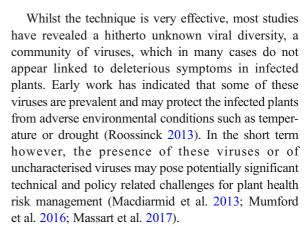Springer

and pathogens. For new diseases, diagnostic laboratories deploy investigational approaches typically including non-targeted methods such as culturing (in vitro and/or in vivo) and microscopy (optical or electron-microscopy) as well as using panels of targeted diagnostics (molecular or serological) to ascertain potential causes of disease.

High Throughput Sequencing (HTS) methods, also known as Next Generation Sequencing (NGS), are probably the most significant advances in molecular biology since the advent of the PCR process in the early 1980s. It enables the de novo sequencing of large amounts of nucleic acid for an ever-decreasing amount of time and cost, enabling the completion of the first thousand-dollar human genome in 2014 (Hayden 2014). The current maximal sequence throughput of these platforms is up to 6 billion reads (2x150bp) per run (<3 days) for the Illumina XTEN (www.illumina.com). The development of these platforms has had a significant impact in the accumulation of genomic data. Microbial genome sequencing in particular is becoming a routine analysis and using the latest technology it is possible to generate near complete bacterial genomes for under €100 (Land et al. 2015). In human health this has led to its routine use in disease outbreaks. The first use was during an outbreak of *Clostridium difficile* in a special care baby unit at the John Radcliffe Hospital in Oxford. Using bacterial genome sequencing it was possible to identify carriers of the infectious strain and map the outbreak through the hospital staff and patients, enabling rapid intervention (Eyre et al. 2013). The US Federal Drug Administration now use genome sequencing as their default microbial food poisoning outbreak tracking tool and have setup a dedicated network called Genometrakr (Chen et al. 2014).

The most developed HTS application in plant health is virus discovery where the technique is rapidly becoming routine for resolving the cause of new or unusual viral symptoms in a diagnostic context. In the short term HTS also holds great promise for the screening of propagation material for quarantine or certification purposes (Fox et al. 2015; Al Rwahnih et al. 2015) in particular for plant viruses where the benefits could be harnessed with only minor modifications to existing techniques yet with technical and quality control challenges (Massart et al. 2014). In the longer term it is expected that these techniques become more widespread for surveillance, screening and identification of other pests and pathogens.

Whilst the technique is very effective, most studies have revealed a hitherto unknown viral diversity, a community of viruses, which in many cases do not appear linked to deleterious symptoms in infected plants. Early work has indicated that some of these viruses are prevalent and may protect the infected plants from adverse environmental conditions such as temperature or drought (Roossinck 2013). In the short term however, the presence of these viruses or of uncharacterised viruses may pose potentially significant technical and policy related challenges for plant health risk management (Macdiarmid et al. 2013; Mumford et al. 2016; Massart et al. 2017).

This review describes how High Throughput Sequencing technologies are being used in plant health diagnostics today, how further developments may impact in the future, and the implications of using the technology for plant health policy.

Plant health related applications

Background

In 2007 the technique was used to identify a virus as potentially responsible for US bee colony losses (Cox-Foster et al. 2007) and in 2008 it was used to identify a novel Arenavirus responsible for killing a number of transplant patients (Palacios et al. 2008). Development of HTS methods in plant health has mirrored that of the wider community. In 2009 three different groups used differing metagenomics approaches to sequence genomes of novel plant viruses (Kreuze et al. 2009; Adams et al. 2009; Al Rwahnih et al. 2009), paving the way for a wide adoption of HTS in a range plant health related applications.

HTS informed diagnostics/aetiology

The identification of novel disease causing viruses was one of the first areas of plant health impacted by the application of HTS technology, leading to the discovery of more than one hundred new plant viruses, new virus variants or new plant hosts for known viruses (Barba et al. 2014). HTS is now in use as part of routine virus diagnostic workflows in several diagnostic laboratories, to identify novel viruses from plant hosts, as illustrated by the increase in new virus species published in the literature on a monthly basis. The technology is also

being used in addition to conventional methods to inform diagnostic workflows in the identification of well characterised pathogens (Fox et al. 2016) or to identify pathogens following initial detection using targeted generic tests (Skelton et al. 2018).

HTS has also been used to identify and gain further information on isolates of bacteria and fungi, such as the identification of *Calonectria pseudonaviculata* as the cause of a new blight on *Sarcococca hookeriana* (Malapi-Wight et al. 2016). The technique can be used to type strains being used to "genotype by sequencing" strains of the plant-pathogenic fungi *Pyrenophora teres* and *Sphaerulina musiva* (Leboldus et al. 2015). Normally the sequencing starts with a pure culture but some years ago, Duan et al. (2009) demonstrated it was possible to reconstruct the whole genome sequence of the pathogenic bacteria 'Candidatus Liberibacter asiaticus' from an infected plant. This study indicated that rapid detection and identification of pathogens more complex than viruses based on sequencing directly from the infected plant material was possible. Adams et al. (2011) identified the presence of *Xanthomonas* causing disease in a Hedera (ivy) using a metagenomics approach.

Despite these successes identifying non-viral pathogens in metagenomes is problematic as many bacteria and higher organisms share significant genome conservation. The consequence of this being that similarities are often found to pathogenic organisms while the sequences may originate from non-pathogenic organism whose genome may not be available. This problem was exemplified by the metagenomics study of (Afshinnekoo et al. 2015a) where initial analysis indicated the presence of both *Yersinia pestis* (plague) and *Bacillus anthracis* on the New York subway. However, these claims were later retracted with the statement "our metagenomic analysis tools identified reads with similarity to *B. anthracis* and *Y. pestis* sequences, there is minimal coverage to the backbone genome of these organisms, and there is no strong evidence to suggest these organisms are in fact present" (Afshinnekoo et al. 2015b).

Development of improved targeted diagnostics

Following on from the use of HTS to identify novel diseases, these techniques deliver large quantities of genomic data which can be used to inform the development of targeted high throughput diagnostics such as

real time PCR. This approach was successfully applied to *Maize chlorotic mottle virus* (MCMV) in east Africa, where after HTS had been used to identify the pathogen, a real-time PCR assay was rapidly developed and deployed (Adams et al. 2013). This was necessary because the isolates of MCMV found in east Africa were highly divergent from those found in the USA and molecular and immunological assays developed to the USA isolates could not be used to detect the east African isolates (Mahuku et al. 2015).

Another application is the routine checking of virus populations to ensure that primers used in routine targeted diagnostics will detect known isolates of a virus. This has been developed by Agroscope (Switzerland) in support of PCR based testing for seed potato certification. On an annual basis extracted RNAs from all certification samples are pooled together (thousands of samples are pooled in a single bulked sample for sequencing). After sequencing at very high sequencing depth (hundreds millions of sequences using a total RNA protocol), the generated sequences are assembled in order to identify genome mutations within common target viruses (Schumpp et al. 2016), to evaluate and modify current PCR based diagnostics.

Sequencing whole bacterial genomes also enables comparative genomic approaches, one practical output of which is the identification of molecular markers associated with different phenotypic characteristics, these markers may be of practical benefit for the development of specific, targeted diagnostic assays. Pritchard et al. (2013) used 20 genome sequences of *Dickeya* to design a range of real-time PCR assays able to distinguish between the species. This approach has since been taken for the development of assays to detect a range of pathogens including *Pseudomonas coronafaciens*, the cause of halo blight in oats (An et al. 2015), *Erwinia amylovora* the cause of fire blight in fruit trees (Bühlmann et al. 2013b) and to discriminate subspecies of *Xanthomonas arboricola pv. pruni* (Bühlmann et al. 2013a).

Disease monitoring and source tracking (population genetics)

In the same way as the FDA is using whole genomes to monitor the causal agents of outbreaks of food poisoning, similar approaches are becoming established in plant pathology. Hubbard et al. (2015) sequenced the transcriptomes (messenger RNA) of leaves infected
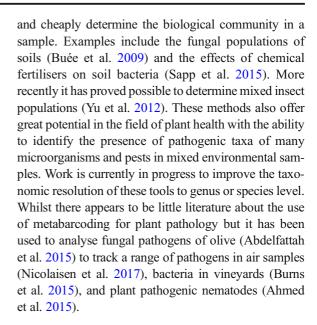
with *Puccinia striiformis* f. sp. *tritici* (yellow rust) in a method they called "field pathogenomics" and were able to track changes in genotypes within populations without the need to first isolate the pathogen. A similar approach has been applied to quickly recover the full genome sequences of virus from different isolates and this approach has been used to link isolates of *Maize chlorotic mottle virus* (MCMV) detected in Rwanda and Kenya as having a shared origin whereas *Sugarcane mosaic virus* (SCMV) in each country was very divergent and therefore likely to be from independent sources (Adams et al. 2014a).

For organisms such as fungi and invertebrate pests, whole genome sequencing of multiple individuals can still be impractical and a number of approaches have been explored to solve this problem. The "field pathogenomics" approach described above exploited mRNA (expressed genes) to avoid sequencing large genomes. Whilst Bonants et al. (2015) used an approach termed Complexity Reduction of Polymorphic Sequences (CRoPS), effectively sequencing restriction fragment length polymorphism (RFLP) fragments from 8 *Synchytrium endobioticum* isolates which enabled them to develop a real-time PCR assay able to distinguish between the common race D1 and non-race 1 pathotypes.

Early surveillance

Reliable surveillance and monitoring programmes are of great significance in plant health, and effective eradication or containment post incursion relies on timely detection. In general, the greater the time between incursion and detection, the less effective and the more complicated any remedial action (Mastin et al. 2017; Cunniffe et al. 2016), although this is not always the case (Thompson et al. 2018). Non-targeted HTS approaches offer great promise for broad spectrum surveillance before entry or before emergence (Luvisi et al. 2016). Adams et al. (2009) noted the presence of western flower thrips (*Frankliniella occidentalis* (Pergande)) sequence in a plant viral metagenome and observation that was confirmed by discussion with glasshouse staff who had dealt with an outbreak in the glasshouse where the plants had been grown, suggesting a potential role for NTS in broad spectrum surveillance across pathogens and pests using the same datasets.

Similarly, meta-barcoding has revolutionised the field of molecular ecology with the ability to rapidly and cheaply determine the biological community in a sample. Examples include the fungal populations of soils (Buée et al. 2009) and the effects of chemical fertilisers on soil bacteria (Sapp et al. 2015). More recently it has proved possible to determine mixed insect populations (Yu et al. 2012). These methods also offer great potential in the field of plant health with the ability to identify the presence of pathogenic taxa of many microorganisms and pests in mixed environmental samples. Work is currently in progress to improve the taxonomic resolution of these tools to genus or species level. Whilst there appears to be little literature about the use of metabarcoding for plant pathology but it has been used to analyse fungal pathogens of olive (Abdelfattah et al. 2015) to track a range of pathogens in air samples (Nicolaisen et al. 2017), bacteria in vineyards (Burns et al. 2015), and plant pathogenic nematodes (Ahmed et al. 2015).

The challenges of applying HTS to plant health

Background

The application of Next-Generation Sequencing brings a step change in the ability to detect and identify previously uncharacterized pathogen-candidates. However, in applying these technologies there are also implications for plant health regulatory authorities in assessing the potential risks posed by previously unknown pathogens. This can be particularly problematic if findings are made in traded material subject to inspection. In particular, there are questions over novel findings if they are made in single samples, samples without symptoms or for poorly characterised organisms where little or no pathological or epidemiological information is known.

There have been several reviews examining the drivers of emerging infectious diseases of plants (EIDs). Anderson et al. (2004) identified that 47% of EIDs are viruses. The impact of improved diagnostics in recording the spread of these diseases is largely overlooked, but may be as much a driver of trends in pathogen detection as changes in trade or research focus (Fox and Mumford 2017). As plant virology is at the forefront in applying HTS to plant pathogen detection, the examples and cases given below are largely based on plant virus diagnostics. However, the issues discussed will become relevant to other disciplines as they apply these techniques in diagnostic protocols.

One area where there is an obvious clash between the application of a non-target test method such as HTS and conventional, targeted testing applied to plant biosecurity is the ability to find related or novel pathogens. Plant Health inspections are based on lists of regulated organisms (Jones and Baker (2007)) an approach that is fundamental to quarantine legislation and can be used to prioritise risks, as seen with the UK plant health risk register (Baker et al. 2014). However, one drawback of list-based legislation is that it can often lag behind advances in diagnostic technology as well as the speed with which pathogens can emerge and spread across borders. The ability of list based systems to deal with an ever increasing catalogue of often uncharacterised pathogens will inevitably be limited. Rodoni (2009) mentions the alarming rate of virus discovery even at the advent of metagenomics being applied in plant pathology, an increase in ICTV ratified plant viruses from <380 species (1991) to >900 named species (2005), as well as a further 2006 tentative virus species that had been detected. Rodoni (2009) also highlights that the rates of accumulation of uncharacterised pathogens was likely to continue, given the nature of short-term molecular based research projects compared to the more time consuming and laborious task of biological characterisation. There is the possibility that a finding from a metagenomics study may be the first identification of an endemic native pathogen, or the finding may be the inadvertent discovery of a virus with a 'persistent lifestyle' (Roossinck 2010).

Ultimately for a National Plant Protection Organisation the key question will be whether a novel finding is truly a new incursion rather than a pathogen that has been present in a region but previously unreported due to limitations in existing diagnostic technologies. This will put a far greater emphasis on baseline surveillance activity to demonstrate what pathogens are currently present in a region. These activities have become secondary in current plant health systems with the greatest focus on stopping incursions rather than demonstrating to what extent a pathogen may already be established. Whilst many inferences can be made about a novel pathogen by analogy to known genetically similar pathogens there is now, more than ever, a need to develop traditional skills to establish the biology of a novel pathogen e.g. host range, epidemiology, symptomology (Massart et al. 2017).

In 2014 Fera, in collaboration with Scientists from BecA, Kenya, sequenced the viral metagenomes of 114 diverse crop and weed plant species sampled from four maize mixed cropping farms, following the emergence of Maize Lethal Necrosis Disease (MLND) in the region. The aim was to explore potential reservoir hosts of the causal viruses, Table 1 details the viruses found during this study. Amongst the viruses found, 14 where previously characterised whilst a further 34 were newly identified to science. This clearly demonstrates the problem, which, if any, of these 34 new viruses might pose a future risk to agriculture in the region or are just new discoveries of endemic viruses.

It is also important to note that not all newly discovered viruses will induce disease symptoms. Sifting the high-risk pathogens from viruses with persistent lifestyles without the ability to refer to biological context will present the greatest challenge to pathologists and policy makers alike. Traditionally, plant pathology has started with a symptom and tried to identify the causal agent of the symptom, an approach which has used a combination of biological techniques, morphology and targeted diagnostics with species or genus specific antisera (ELISA) or primers (PCR based methods). However, there are limitations to the range and scope of each of these diagnostic approaches which can subsequently lead to erroneous conclusions. A good example of this is the search for the causal agent of lettuce big-vein disease. The association of big-vein disease with infection by a fungal pathogen was first reported in 1934 (Jagger and Chandler 1934). However, 50 years passed before virus-like particles of *Lettuce big vein 'associated' varicosavirus* (LBVaV, formerly 'lettuce big vein varicosavirus') were observed from infected plants (Kuwata et al. 1983). A decade later, a second virus was identified from affected lettuce and associated with the disease, *Mirafiori lettuce big-vein ophiovirus* (MLBVV) (Roggero et al. 2000). Subsequent investigations have shown that MLBVV is the more probable cause of big-vein disease (Lot et al. 2002), but that LBVaV may still be associated with other symptoms such as localised necrosis (Verbeek et al. 2013). This eight-decade arc of investigation moved on with advances in diagnostic methods, but ultimately progress was hampered by the limitations of these techniques: Electron microscopy can only observe the morphology of viral particles present, but cannot be used to give a conclusive diagnosis at species level; Biological indexing can give an indication of the pathogens present in a sample, but is prone to failure where a virus is labile or is not amenable to mechanical transmission; Targeted methods can only detect the targets the assays have been

Table 1 Viruses sequenced in Kenyan maize farms

	Crop	Samples	Known plant viruses	Novel plant viruses	totals
Farm 1	maize	4	Maize chlorotic mottle virus, Sugarcane mosaic virus, Potato virus S, Bean common mosaic virus, Maize yellow dwarf mosaic virus	Tombusvirus, Carmovirus, Foveavirus, Closterovirus, Betaflexivirus, positive strand ssRNA virus	30
	others	26	Bean common mosaic virus, Beet pseudoyellows virus, Maize yellow dwarf mosaic virus, SCMV, Potato virus S	Caulimoviridae virus, Chrysovirus, Crinivirus, Potyvirus (es), Tombusvirus, unclassified positive strand ssRNA virus, Varicosavirus	
Farm 2	maize	9	Maize chlorotic mottle virus	Chrysovirus, Luteovirus, Carmovirus, tombusvirus, virus, positive strand ssRNA virus, unclassified virus	29
	others	20	Potato virus S, shallot latent virus, Cauliflower mosaic virus,	Chrysovirus, Crinivirus, Cytorhabdovirus, Waikavirus, Varicosavirus Polerovirus, Polerovirus associated RNA, Tymoviridae virus, positive strand ssRNA virus	
Farm 3	maize	6	Maize chlorotic mottle virus, Maize yellow dwarf mosaic virus	Badnavirus, Polerovirus associated RNA, Tymoviradae virus	26
	others	20	Turnip mosaic virus,	Badnavirus, Chrysovirus, Cytorhabdovirus, positive strand ssRNA virus	
Farm 4	maize	4	Maize chlorotic mottle virus, Maize yellow dwarf mosaic virus	none	30
	others	26	Banana streak virus, Apple stem grooving virus, Citrus tristeza virus, Potato virus S,	Badnavirus, Potyvirus (es), Tombusvirus, Rhabdoviridae virus, positive strand ssRNA virus, unclassified virus, Varicosavirus	
				total	115

designed against, at best a broad number of species in a given genus (Adams et al. 2013; De Clerck et al. 2017; Rott et al. 2017).

Causality

The application of HTS to phytopathology provides the opportunity to circumvent the extensive time taken on investigational virus discovery, as exemplified in the lettuce example above. However, although the pathogen discovery phase of the work would be accelerated, the key aspect of the work still remains, the ability to demonstrate a causal relationship between the virus and the disease. Traditionally, in plant pathology as in animal/human pathology the approach is to fulfil 'Koch's Postulates' (Evans 1976), the basic tenets of which have been accepted as the benchmark for demonstration of causation. However, fulfilling Koch's postulates can be time consuming especially when working with infections by obligate pathogens, or with disease caused by multiple pathogens; or where symptom expression is also under the influence of environmental factors such as temperature (Dahal et al. 1998) or nitrogen stress (Talbot et al. 1997).

In a plant health context this time delay is problematic as considerable spread of the disease can occur whilst fulfilling Koch postulates, resulting in delays to preventative action being taken. Given the adage that 'Correlation does not equal causation' the medical statistician Sir Austin Bradford Hill discussed a pragmatic approach for inferring causation from analysis of epidemiological data and lists nine factors (strength of the association; consistency; specificity; temporality; biological gradient; plausibility; coherence; experimental evidence and judging by analogy) which should be considered when assessing whether a disease is the result of a given set of circumstances (Hill 1965). Technological advances, not least the ability to detect and identify nucleic acids from samples, have led to an increasing array of methods available to the plant pathologist. This resulted in a growing reliance on the identification of genotypes for diagnosis of disease. With these advances in mind, Fredericks and Relman (1996) developed guidelines to allow sequence based microbial identification to be incorporated into assessments of causation. These guidelines include reference to 'copy number', or the relative quantification of sequences from samples of infected and non-infected

hosts, to allow the pathologist to account for possible latent infections. In each case above these authors were keen to point out that their criteria were guidelines, rather than rigid rules, and were intended to encourage scientific rigour.

If we take these approaches into account, employing a rigorous experimental design, it should be possible to use HTS based identification, potentially supported with conventional molecular testing (e.g. PCR), to infer a casual association between a novel pathogen and an existing disease of unknown aetiology. Whilst Koch's postulates would still be accepted as 'the gold standard', there should be enough supporting evidence from this type of association to allow preventative action. Such an approach was utilised by Adams et al. (2014b) whilst investigating a disease causing carrot root necrosis. This study had to account for issues raised here as well as other practical considerations: the presence of multiple co-infections; the need to separate the pathogenic from the commensal or even mutually beneficial; the inability to experimentally demonstrate causation due to time-scale and inability to transmit the putative pathogen; the presence of asymptomatic infected individuals; symptom incidence not evident until post-harvest cutting of the infected host; as well as the limited availability of conventional test methods. A statistical approach was taken to allow affected and unaffected individuals to be collected post-harvest, and these could be screened for known pathogens and tested using HTS for novel or unusual pathogens. The strength of association between both single and multiple pathogen infections and the presence of symptom could then be calculated. The approach taken gave at least an indication of the putative cause of disease. At a broader scale, a framework of scaled biological characterization and risk assessment for new viral species has been recently proposed (Massart et al. 2017).

Data interpretation

The genome databases are growing daily which means that more and more sequences will be available from pathogens but also from commensal and/or closely related non-pathogenic species. This will improve the ability to specifically detect a plant pathogen from the generated sequences thanks to the more precise and robust identification of specific genome regions. Little work has been done within countries or regions establishing a baseline of viruses/pathogens present in its territory, the lack of this information, makes it difficult to make informed decisions when viruses/pathogens are detected by HTS on imported material. Additionally, there is the problem of differentiation between bacterial/fungal pathogens and their taxonomically close relatives (not pathogenic) in a metagenomics approach. Accurate taxonomic assignment at the species level of DNA sequences from bacterial and fungal microbiota is a challenging and yet unsolved problem. The lack of clear demarcation between species and incomplete or inaccurate reference databases and the resolution of current analysis tools often limit identification to the genus or family levels.

Validation

The validation process of the HTS technology in diagnostics has first, as expected, been subject of discussion in a clinical setting (Mccourt et al. 2013; Frampton et al. 2013; Mattocks et al. 2010; Salto-Tellez and Gonzalez De Castro 2014). In plant pathology, the progress made in the clinical environment will help with our adoption of HTS as routine technology in plant health diagnostics. Many plant health laboratories currently use methods accredited under the ISO17025 standard and are moving towards a flexible scope of accreditation due to the number of pathogens/pests, hosts and matrices, and therefore methods, that require accreditation. The key differences between HTS and conventional tests within a quality framework is the non-targeted nature and under current practice the lack of controls to enable test performance to be effectively monitored. To satisfy the requirements of accreditation bodies it will be necessary to develop approaches to monitor the performance of HTS on a run-by-run basis. As a result, accreditation may be more straightforward for screening applications (i.e. testing of a number of targets in parallel) than for virus discovery where only some parts of the process may be monitored to the appropriate level.

EPPO (the European and Mediterranean Plant Protection Organization) is currently revising their diagnostic standards PM7/84 (OEPP/EPPO 2007) and PM7/98 (OEPP/EPPO 2014) in view of the minimum validation requirements in a flexible scope setting. Many of the recommendations for the validation of molecular diagnostics can be applied to HTS but it will be important to agree a uniform interpretation of accreditation standards.

The validation process will also have to cope with the rapidly evolving nature of the technology. There has been a constant introduction of new sequencers and improved models. This is mirrored by improvements in the consumables run on the machines and crucially, the techniques and software used to analyse the data. The validation/accreditation needs to be flexible enough to allow laboratories to use the most appropriate tools without fixing outdated protocols in place. However, we are still facing many technical, scientific and regulatory challenges. To meet the quality assurance technical requirements, standards on managing false positives, especially those caused by contamination or biologically inactive pathogens, as well as interfacing HTS with other technologies (at least for confirmation purposes) should be drafted and implemented.

Future prospects

Currently, novel findings may be dealt with between the laboratory and the NPPO through ad-hoc consultation, where HTS findings are discussed in relation to likely effects based on similarity to well characterised pathogens and the risk presented by the particular commodity or pathway on which the finding was intercepted. This approach is largely about identifying which pathogens are likely to present a risk and which novel findings are likely to be viruses with persistent lifestyles and therefore of limited plant health interest. At present there are caveats applied to the judgements made on such novel findings, largely due to the uncertainty associated with background knowledge and the limited contextual data on any given sample. As the technology moves more widely from an R&D based support tool to frontline diagnostic applications there will be a need to formalise the framework of this flow of information and to support this with a greater emphasis on gathering supporting biological data.

Currently most labs using HTS are doing so as the first step in a pipeline of methods seeking to identify the causal agent of a disease; follow on testing is done for confirmation of results and a framework for doing so has been recently published (Massart et al. 2017). Confirmatory testing is done for several different reasons. Currently most people are using Illumina platforms which are known to have a significant problem with sample-to-sample contamination due to the internal architecture of the instruments. In some cases, this means

without follow on testing it is not possible to be certain about which sample is infected. In addition, confirmatory testing is frequently done in a regulatory context to provide certainty to policy makers about findings. As the HTS technology improves for diagnostic use and contamination presumably becomes less of an issue, there may be some circumstances where confirmatory testing, performed due to contamination becomes unnecessary.

The international scientific community, as well as plant health policy makers are well aware of the enormous benefits the HTS technology offer and if, the above listed challenges are solved, and the large amounts of data are used wisely with relevant interpretation at bioinformatics, scientific, regulatory, and commercial level, the technique will be of enormous benefit to plant health and will contribute to more sustainable agriculture and safer trade in plants and plant products.

References

Abdelfattah, A., Li Destri Nicosia, M. G., Cacciola, S. O., Droby, S., & Schena, L. (2015). Metabarcoding analysis of fungal diversity in the Phyllosphere and Carposphere of olive (Olea europaea). *PLoS One, 10*, e0131069.

Adams, I. P., Glover, R. H., Monger, W. A., et al. (2009). Next-generation sequencing and metagenomic analysis: A universal diagnostic tool in plant virology. *Molecular Plant Pathology, 10*, 537–545.

Adams, I. P., Glover, R. H., Monger, W. A., et al. (2011). Next-generation sequencing and metagenomic analysis: a universal diagnostic tool in plant pathology. In F. J. De Bruijn (Ed.), *Handbook of Molecular Microbial Ecology II: Metagenomics in Different Habitats* (pp. 63–71). New York: Wiley.

Adams, I. P., Miano, D. W., Kinyua, Z. M., Wangai, A., Kimani, E., Phiri, N., Reeder, R., Harju, V., Glover, R., Hany, U., Souza-Richards, R., Deb Nath, P., Nixon, T., Fox, A., Barnes, A., Smith, J., Skelton, A., Thwaites, R., Mumford, R., & Boonham, N. (2013). Use of next-generation sequencing for the identification and characterization of Maize chlorotic

mottle virus and sugarcane mosaic virus causing maize lethal necrosis in Kenya. *Plant Pathology, 62*, 741–749.

Adams IP, Harju VA, Hodges T, Hany U, Skelton A, Rai S, Deka MK, Smith J, Fox A, Uzayisenga B, Ngaboyisonga C, Uwumukiza B, Rutikanga A, Rutherford M, Ricthis B, Phiri N, Boonham N, (2014a). First report of maize lethal necrosis disease in Rwanda. *New Disease Reports* 29. 22.

Adams, I. P., Skelton, A., Macarthur, R., Hodges, T., Hinds, H., Flint, L., Nath, P. D., Boonham, N., & Fox, A. (2014b). Carrot yellow leaf virus is associated with carrot internal necrosis. *PLoS One, 9*, e109125.

Afshinnekoo, E., Meydan, C., Chowdhury, S., Jaroudi, D., Boyer, C., Bernstein, N., Maritz, J. M., Reeves, D., Gandara, J., Chhangawala, S., Ahsanuddin, S., Simmons, A., Nessel, T., Sundaresh, B., Pereira, E., Jorgensen, E., Kolokotronis, S. O., Kirchberger, N., Garcia, I., Gandara, D., Dhanraj, S., Nawrin, T., Saletore, Y., Alexander, N., Vijay, P., Hénaff, E. M., Zumbo, P., Walsh, M., O'Mullan, G. D., Tighe, S., Dudley, J. T., Dunaif, A., Ennis, S., O'Halloran, E., Magalhaes, T. R., Boone, B., Jones, A. L., Muth, T. R., Paolantonio, K. S., Alter, E., Schadt, E. E., Garbarino, J., Prill, R. J., Carlton, J. M., Levy, S., & Mason, C. E. (2015a). Geospatial resolution of human and bacterial diversity with city-scale metagenomics. *Cell Systems, 1*, 72–87.

Afshinnekoo, E., Meydan, C., Chowdhury, S., Jaroudi, D., Boyer, C., Bernstein, N., Maritz, J. M., Reeves, D., Gandara, J., Chhangawala, S., Ahsanuddin, S., Simmons, A., Nessel, T., Sundaresh, B., Pereira, E., Jorgensen, E., Kolokotronis, S. O., Kirchberger, N., Garcia, I., Gandara, D., Dhanraj, S., Nawrin, T., Saletore, Y., Alexander, N., Vijay, P., Hénaff, E. M., Zumbo, P., Walsh, M., O'Mullan, G. D., Tighe, S., Dudley, J. T., Dunaif, A., Ennis, S., O'Halloran, E., Magalhaes, T. R., Boone, B., Jones, A. L., Muth, T. R., Paolantonio, K. S., Alter, E., Schadt, E. E., Garbarino, J., Prill, R. J., Carlton, J. M., Levy, S., & Mason, C. E. (2015b). Erratum: Geospatial resolution of human and bacterial diversity with City-scale metagenomics. *Cell Systems, 1*, 97–97.e3.

Ahmed, M., Sapp, M., Prior, T., Karssen, G., & Back, M. (2015). Nematode taxonomy: From morphology to metabarcoding. *SOIL Discuss., 2015*, 1175–1220.

Al Rwahnih, M., Daubert, S., Golino, D., & Rowhani, A. (2009). Deep sequencing analysis of RNAs from a grapevine showing Syrah decline symptoms reveals a multiple virus infection that includes a novel virus. *Virology, 387*, 395–401.

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. *Plant Pathology Journal, 31*, 25–32.

Anderson, P. K., Cunningham, A. A., Patel, N. G., Morales, F. J., Epstein, P. R., & Daszak, P. (2004). Emerging infectious diseases of plants: Pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology & Evolution, 19*, 535–544.

Baker, R. H. A., Anderson, H., Bishop, S., Macleod, A., Parkinson, N., & Tuffen, M. G. (2014). The UK Plant health risk register: A tool for prioritizing actions. *EPPO Bulletin, 44*, 187–194.

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

Bonants, P. J. M., Gent-Pelzer, M. P. E., Leeuwen, G. C. M., & Lee, T. J. (2015). A real-time TaqMan PCR assay to discriminate between pathotype 1 (D1) and non-pathotype 1 (D1) isolates of Synchytrium endobioticum. *European Journal of Plant Pathology, 143*, 495–506.

Buée, M., Reich, M., Murat, C., et al. (2009). 454 Pyrosequencing analyses of forest soils reveal an unexpectedly high fungal diversity. *New Phytologist, 184*, 449–456.

Bühlmann, A., Pothier, J., Tomlinson, J., et al. (2013a). Genomics-informed design of loop-mediated isothermal amplification for detection of phytopathogenic Xanthomonas arboricola pv. Pruni at the intraspecific level. *Plant Pathology, 62*, 475–484.

Bühlmann, A., Pothier, J. F., Rezzonico, F., Smits, T. H. M., Andreou, M., Boonham, N., Duffy, B., & Frey, J. E. (2013b). Erwinia amylovora loop-mediated isothermal amplification (LAMP) assay for rapid pathogen detection and on-site diagnosis of fire blight. *Journal of Microbiological Methods, 92*, 332–339.

Burns, K. N., Kluepfel, D. A., Strauss, S. L., Bokulich, N. A., Cantu, D., & Steenwerth, K. L. (2015). Vineyard soil bacterial diversity and composition revealed by 16S rRNA genes: Differentiation by geographic features. *Soil Biology and Biochemistry, 91*, 232–247.

Chen, P., Jeannotte, R., & Weimer, B. C. (2014). Exploring bacterial epigenomics in the next-generation sequencing era: A new approach for an emerging frontier. *Trends in Microbiology, 22*, 292–300.

Cox-Foster, D. L., Conlan, S., Holmes, E. C., Palacios, G., Evans, J. D., Moran, N. A., Quan, P. L., Briese, T., Hornig, M., Geiser, D. M., Martinson, V., vanEngelsdorp, D., Kalkstein, A. L., Drysdale, A., Hui, J., Zhai, J., Cui, L., Hutchison, S. K., Simons, J. F., Egholm, M., Pettis, J. S., & Lipkin, W. I. (2007). A metagenomic survey of microbes in honey bee colony collapse disorder 99. *Science, 318*, 283–287.

Cunniffe, N. J., Cobb, R. C., Meentemeyer, R. K., Rizzo, D. M., & Gilligan, C. A. (2016). Modeling when, where, and how to manage a forest epidemic, motivated by sudden oak death in California. *Proceedings of the National Academy of Sciences, 113*, 5640–5645.

Dahal, G., Hughes, J. D. A., Thottappilly, G., & Lockhart, B. (1998). Effect of temperature on symptom expression and reliability of banana streak badnavirus detection in naturally infected plantain and banana (Musa spp.). *Plant Disease, 82*, 16–21.

De Clerck, C., Crew, K., Mcmichael, L., et al. (2017). Lessons learned from the virus indexing of Musa germplasm: Insights from a multiyear collaboration. *Annals of Applied Biology, 171*, 15–27.

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

Evans, A. S. (1976). Causation and disease: The Henle-Koch postulates revisited. *The Yale Journal of Biology and Medicine, 49*, 175–195.

Eyre, D. W., Cule, M. L., Wilson, D. J., Griffiths, D., Vaughan, A., O'Connor, L., Ip, C. L. C., Golubchik, T., Batty, E. M., Finney, J. M., Wyllie, D. H., Didelot, X., Piazza, P., Bowden, R., Dingle, K. E., Harding, R. M., Crook, D. W., Wilcox, M. H., Peto, T. E. A., & Walker, A. S. (2013). Diverse sources of C. Difficile infection identified on whole-genome sequencing. *New England Journal of Medicine, 369*, 1195–1205.

Fox, A., & Mumford, R. (2017). Plant viruses and viroids in the United Kingdom: An analysis of first detections and novel discoveries from 1980 to 2014. *Virus Research, 241*, 10–18.

Fox, A., Adams, I., Hany, U., et al. (2015). The application of next-generation sequencing for screening seeds for viruses and viroids. *Seed Science and Technology, 43*, 531–535.

Fox A, Fowkes A, Buxton-Kirk A, et al., (2016). First report of Potato aucuba mosaic virus in *Solanum jasminoides* in the United Kingdom. *New Disease Reports* 34, 32.

Frampton, G. M., Fichtenholtz, A., Otto, G. A., Wang, K., Downing, S. R., He, J., Schnall-Levin, M., White, J., Sanford, E. M., An, P., Sun, J., Juhn, F., Brennan, K., Iwanik, K., Maillet, A., Buell, J., White, E., Zhao, M., Balasubramanian, S., Terzic, S., Richards, T., Banning, V., Garcia, L., Mahoney, K., Zwirko, Z., Donahue, A., Beltran, H., Mosquera, J. M., Rubin, M. A., Dogan, S., Hedvat, C. V., Berger, M. F., Pusztai, L., Lechner, M., Boshoff, C., Jarosz, M., Vietz, C., Parker, A., Miller, V. A., Ross, J. S., Curran, J., Cronin, M. T., Stephens, P. J., Lipson, D., & Yelensky, R. (2013). Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nature Biotechnology, 31*, 1023–1031.

Fredericks, D. N., & Relman, D. A. (1996). Sequence-based identification of microbial pathogens: A reconsideration of Koch's postulates. *Clinical Microbiology Reviews, 9*, 18–33.

Hayden, E. (2014). Technology: The $1,000 genome. *Nature, 507*, 294–295.

Hill, A. (1965). The Envrionment and disease: Association or causation? *Proceedings of the Royal Society of Medicine, 58*, 295–300.

Hubbard, A., Lewis, C. M., Yoshida, K., Ramirez-Gonzalez, R. H., de Vallavieille-Pope, C., Thomas, J., Kamoun, S., Bayles, R., Uauy, C., & Saunders, D. (2015). Field pathogenomics reveals the emergence of a diverse wheat yellow rust population. *Genome Biology, 16*, 23.

Jagger, I. C., & Chandler, N. (1934). Big vein, a disease of lettuce. *Phytopathology, 24*, 1253–1256.

Jones, D. R., & Baker, R. H. A. (2007). Introductions of non-native plant pathogens into Great Britain, 1970–2004. *Plant Pathology, 56*, 891–910.

Kreuze, J. F., Perez, A., Untiveros, M., Quispe, D., Fuentes, S., Barker, I., & Simon, R. (2009). Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: A generic method for diagnosis, discovery and sequencing of viruses. *Virology, 388*, 1–7.

Kuwata, S., Kubo, S., Yamashita, S., & Doi, Y. (1983). Rod-shaped particles, a probable entity of lettuce big vein virus. *Annals of the Phytopathological Society of Japan, 49*, 246–251.

Land, M., Hauser, L., Jun, S.-R., Nookaew, I., Leuze, M. R., Ahn, T. H., Karpinets, T., Lund, O., Kora, G., Wassenaar, T., Poudel, S., & Ussery, D. W. (2015). Insights from 20 years of bacterial genome sequencing. *Functional & Integrative Genomics, 15*, 141–161.

Leboldus, J. M., Kinzer, K., Richards, J., Ya, Z., Yan, C., Friesen, T. L., & Brueggeman, R. (2015). Genotype-by-sequencing of the plant-pathogenic fungi Pyrenophora teres and Sphaerulina musiva utilizing ion torrent sequence technology. *Molecular Plant Pathology, 16*, 623–632.

Lot, H., Campbell, R. N., Souche, S., Milne, R. G., & Roggero, P. (2002). Transmission by Olpidium brassicae of Mirafiori lettuce virus and lettuce big-vein virus, and their roles in lettuce big-vein etiology. *Phytopathology, 92*, 288–293.

Luvisi, A., Ampatzidis, Y. G., & De Bellis, L. (2016). Plant pathology and information technology: Opportunity for management of disease outbreak and applications in regulation frameworks. *Sustainability, 8*, 831.

Macdiarmid, R., Rodoni, B., Melcher, U., Ochoa-Corona, F., & Roossinck, M. (2013). Biosecurity implications of new technology and discovery in plant virus research. *PLoS Pathogens, 9*, e1003337.

Mahuku, G., Lockhart, B. E., Wanjala, B., Jones, M. W., Kimunye, J. N., Stewart, L. R., Cassone, B. J., Sevgan, S., Nyasani, J. O., Kusia, E., Kumar, P. L., Niblett, C. L., Kiggundu, A., Asea, G., Pappu, H. R., Wangai, A., Prasanna, B. M., & Redinbaugh, M. G. (2015). Maize lethal necrosis (MLN), an emerging threat to maize-based food security in sub-Saharan Africa. *Phytopathology, 105*, 956–965.

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*, 1093–1100.

Massart, S., Olmos, A., Jijakli, H., & Candresse, T. (2014). Current impact and future directions of high throughput sequencing in plant virus diagnostics. *Virus Research, 188*, 90–96.

Massart S, Candresse T, Gil J, et al., 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.

Mastin, A. J., Van Den Bosch, F., Gottwald, T. R., Chavez, V. A., & Parnell, S. R. (2017). A method of determining where to target surveillance efforts in heterogeneous epidemiological systems. *PLoS Computational Biology, 13*, e1005712.

Mattocks, C. J., Morris, M. A., Matthijs, G., et al. (2010). A standardized framework for the validation and verification of clinical molecular genetic tests. *European Journal of Human Genetics, 18*, 1276–1288.

Mccourt, C. M., Mcart, D. G., Mills, K., et al. (2013). Validation of next generation sequencing technologies in comparison to current diagnostic gold standards for BRAF, EGFR and KRAS mutational analysis. *PLoS One, 8*, e69604.

Mumford, R., Macarthur, R., & Boonham, N. (2016). The role and challenges of new diagnostic technology in plant biosecurity. *Food Security, 8*, 103–109.

Nicolaisen, M., West, J. S., Sapkota, R., Canning, G. G., Schoen, C., & Justesen, A. F. (2017). Fungal communities including plant pathogens in near surface air are similar across northwestern Europe. *Frontiers in Microbiology, 8*, 1729.

Oepp/Eppo. (2007). PM 7/84 basic requirements for quality management in plant pest diagnosis laboratories. *Bulletin OEPP/EPPO Bulletin, 37*, 580–588.

Oepp/Eppo. (2014). PM7/98 specific requirements for laboratories preparing accreditation for a plant pest diagnostic activity. *Bulletin OEPP/EPPO Bulletin, 44*, 117–147.

Palacios G, Druce J, Du L, et al., (2008). A new arenavirus in a cluster of fatal transplant-associated diseases. *A New England Journal of Medicine, 358*, 991–998.

Pritchard, L., Humphris, S., Saddler, G. S., Parkinson, N. M., Bertrand, V., Elphinstone, J. G., & Toth, I. K. (2013). Detection of phytopathogens of the genus Dickeya using a PCR primer prediction pipeline for draft bacterial genome sequences. *Plant Pathology, 62*, 587–596.

Rodoni, B. (2009). The role of plant biosecurity in preventing and controlling emerging plant virus disease epidemics. *Virus Research, 141*, 150–157.

Roggero, P., Ciuffo, M., Vaira, A. M., Accotto, G. P., Masenga, V., & Milne, R. G. (2000). An Ophiovirus isolated from lettuce with big-vein symptoms. *Archives of Virology, 145*, 2629–2642.

Roossinck, M. J. (2010). Lifestyles of plant viruses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 365*, 1899–1905.

Roossinck, M. J. (2013). Plant virus ecology. *PLoS Pathogens, 9*, e1003304.

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.

Salto-Tellez, M., & Gonzalez De Castro, D. (2014). Next-generation sequencing: A change of paradigm in molecular diagnostic validation. *The Journal of Pathology, 234*, 5–10.

Sapp, M., Harrison, M., Hany, U., Charlton, A., & Thwaites, R. (2015). Comparing the effect of digestate and chemical fertiliser on soil bacteria. *Applied Soil Ecology, 86*, 1–9.

Schumpp O, Dupuis B, Bréchon A, Al. E, (2016). Diagnostic moléculaire à haut debit pour détecter les viroses des plants de pommes de terre. *Recherche Agronomique Suisse, 7*, 456–465.

Skelton A, Uzayisenga B, Fowkes A, et al., (2018). First report of pepper veinal mottle virus, pepper yellows virus and a novel enamovirus in chilli pepper (*Capsicum* sp.) in Rwanda. *New Disease Reports* 37, 5.

Talbot, N., Mccafferty, H., Ma, M., Moore, K., & Hamer, J. (1997). Nitrogen starvation of the rice blast fungusMagnaporthe griseamay act as an environmental cue for disease symptom expression. *Physiological and Molecular Plant Pathology, 50*, 179–195.

Thompson, R. N., Gilligan, C. A., & Cunniffe, N. J. (2018). Control fast or control smart: When should invading pathogens be controlled? *PLoS Computational Biology, 14*, e1006014.

Verbeek, M., Dullemans, A. M., Van Bekkum, P. J., & Van Der Vlugt, R. A. (2013). Evidence for lettuce big-vein associated virus as the causal agent of a syndrome of necrotic rings and spots in lettuce. *Plant Pathology, 62*, 444–451.

Yu, D. W., Ji, Y., Emerson, B. C., Wang, X., Ye, C., Yang, C., & Ding, Z. (2012). Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution, 3*, 613–623.