High Throughput Sequencing For Plant Virus Detection and Discovery

D. E. V. Villamor,¹ T. Ho,¹ M. Al Rwahnih,² R. R. Martin,³ and I. E. Tzanetakis^{1,†}

¹Department of Plant Pathology, Division of Agriculture, University of Arkansas System, Fayetteville, AR 72701; ²Department of Plant Pathology, University of California, Davis 95616; and ³Horticulture Crops Research Unit, U.S. Department of Agriculture-Agricultural Research Service, Corvallis, OR 97330

Accepted for publication 20 February 2019.

ABSTRACT

Over the last decade, virologists have discovered an unprecedented number of viruses using high throughput sequencing (HTS), which led to the advancement of our knowledge on the diversity of viruses in nature, particularly unraveling the virome of many agricultural crops. However, these new virus discoveries have often widened the gaps in our understanding of virus biology; the forefront of which is the actual role of a new virus in disease, if any. Yet, when used critically in etiological studies, HTS is a powerful tool to establish disease causality between the virus and its host. Conversely, with globalization, movement of plant material is increasingly more common and often a point of dispute between countries. HTS could potentially resolve these issues given its capacity to detect and discover. Although many pipelines are available for plant virus discovery, all share a common backbone. A description of the process of plant virus detection and discovery from HTS data are presented, providing a summary of the different pipelines available for scientists' utility in their research.

Keywords: certification, clean plant, disease control and pest management, etiology, high throughput sequencing, techniques, virology, virus detection pipelines, virus discovery.

From a historical perspective, new virus discoveries are often outputs of etiological investigations of economically important virus and virus-like diseases of agricultural crops. What normally follows virus identification and characterization is the development of reliable detection methods. With the application of polymerase chain reaction (PCR) in plant virology, these two activities, (i) identification and characterization and (ii) development of detection methods, are undertaken almost simultaneously. Subsequently, as the population structure of the virus is studied, the detection method is further improved to accommodate detection of diverse isolates and strains, potentially allowing for the detection of all or the vast majority of variants that circulate in the agricultural and native hosts.

Detection is a critical component of disease management, particularly in the context of virus diseases because of the lack of commercially available chemistries against plant viruses. In the majority of cases once infected the plant remains infected throughout its lifetime. Yet there are exceptions; there are viruses that may cause disease but not move systemically such as blueberry necrotic ring blotch virus (BNRBV) (Robinson et al. 2016). Apart from breeding for virus

resistance, the use of virus-tested planting material; screened by reliable, sensitive detection protocols is the most critical component of any virus management program. Reliable detection also aides in the effective implementation of virus management strategies aiming the virus vector at the onset of infection, an especially important tool in the high value perennial crops.

A new method is poised to replace all previously used detection technologies. Referred to as high throughput sequencing (HTS), next generation sequencing (NGS), deep sequencing or large scale sequencing, this technology has revolutionized the nucleic acid sequencing process since it allows the sequencing of millions of nucleotides in a short period of time at a very high redundancy (depth of sequencing). Presently, HTS supersedes all previous sequencing technologies. When combined with specific bioinformatics tools, HTS can be used for detection of known viruses and discovery of new viruses (Adams et al. 2009; Al Rwahnih et al. 2009, 2015; Bag et al. 2015; Donaire et al. 2009; Elbeaino et al. 2014a, b; Giampetruzzi et al. 2012; Ho and Tzanetakis 2014; Kreuze et al. 2009; Marais et al. 2015a, b; Rott et al. 2017; Villamor et al. 2016, 2017). In addition, HTS has applications in genetic diversity, small RNA/gene expression and epidemiological studies. This chapter highlights the use of HTS in virus detection and discovery, as well as its concomitant use in genetic diversity and etiological studies, its impact in certification and movement of planting materials across the globe. The bioinformatics pipelines available for data analysis are also discussed.

[†]Corresponding author: I. Tzanetakis; itzaneta@uark.edu

The author(s) declare no conflict of interest.

HTS BACKGROUND AND APPLICATIONS

Background. The seminal work of Frederick Sanger in the 1970s using chain-terminating dideoxy nucleotides to sequence bacteriophage ϕ X174 (Sanger et al. 1977) provided the foundational principle of pre-HTS DNA sequencing, better known as Sanger sequencing. This method, improved by the replacement of radio-labeling with fluorescent labeling of nucleotides (Smith et al. 1986), slab gel with capillary electrophoresis systems (Marsh et al. 1997) and availability of robotic instruments, facilitated great advances in biology, reaching its pinnacle in the sequencing of the human genome (Lander et al. 2001; Venter et al. 2001).

Considered as first-generation sequencing technology, it could deliver 96 or 384 sample sequences per instrument with a read length ranging from 600 to over 1,000 nucleotides (nt). At the turn of the century, this technology underwent an unprecedented change, paving the way for the arrival of second and third-generation sequencing technologies, which are generally referred to as HTS technology. In its essence, HTS is a term used to describe methodologies used to generate millions (the PromethION instrument of Oxford Nanopore) to trillions (in the case of NovaSeq 6000 instrument of the Illumina platform) of nucleotide sequences in a single instrument run. The main difference between second and third-generation technologies is that the former requires template amplification prior to sequencing whereas the latter utilizes individual DNA molecules as template; therefore third-generation sequencing is also referred to as single-molecule sequencing (Rhoads and Au 2015; Wang et al. 2015b; Heather and Chain 2016).

Various HTS technologies are available commercially with additional platforms at the precommercialization stage. Several reviews have been published that describe how these different HTS platforms perform the sequencing and the amount and reliability of data that each generates (Rothberg and Leamon 2008; Metzker 2010; Glenn 2011; Mardis 2013, 2008; Buermans and den Dunnen 2014; Feng et al. 2015; Reuter et al. 2015; Rhoads and Au 2015; Goodwin et al. 2016; Heather and Chain, 2016; Levy and Myers 2016; Lu et al. 2016; Mardis 2017) and therefore this aspect will not be discussed here.

HTS platforms share three common steps: (i) DNA fragmentation to create the library; (ii) addition of synthetic DNA adapters to individual fragments; and (iii) sequencing of each fragment. Notably, when RNA is used as a starting material, e.g., transcriptome analysis (RNASeq experiments), the fragmented RNA is first reverse transcribed to create a cDNA library. Additionally, RNA selection is performed prior to library construction. Two commonly used selection methods are (i) ribosomal RNA depletion and (ii) enrichment for polyadenylated RNAs. In the case of virus detection, ribosomal RNA depletion is the most commonly employed selection method as it allows for detection of virtually all viruses present in the sample whereas polyadenylated RNA selection is particularly suited if the target virus contains of a poly-A tail in the 3'-terminal end of its genome.

In general, different HTS platforms can be categorized according to (i) the method of detection of nucleotide sequence, (ii) the proximate source of the nucleotide, and (iii) the sequencing chemistry employed (Levy and Myers 2016). In the first category, platforms that use optical detection to identify the base incorporated include Illumina, PacBio-Illumina, SOLiD (fluorescence detection) and 454 (pyrosequencing), whereas nonoptical detection such as Ion Torrent records pH changes during polymerization. In the second category, each sequence read from Illumina, Ion Torrent, Roche 454, and SOLiD platforms is derived from a clonally amplified DNA, whereas corresponding sequences from PacBio-Illumina and Oxford Nanopore originate from a single molecule. Finally, in the third category, Illumina, Ion Torrent, PacBio-Illumina, and Roche 454 platforms all utilize polymerase to drive sequencing by synthesis reaction; SOLiD employs sequencing by

ligation whereas Oxford Nanopore directly determines DNA sequence as it passes through a nanopore. A brief description of currently available technologies is presented below and a comparison of selected features between platforms is summarized to provide relevant information regarding platforms to use to address various research questions.

Second-generation sequencing platforms. The first HTS platform is called Roche 454. Originally released as 454 in 2005 and later acquired by Roche in 2007, this platform captures a template molecule in a bead that is further loaded on a well of a picotiter plate for amplification using emulsion PCR and finally sequenced using pyrosequencing (Rothberg and Leamon 2008). The GS FLX+ Titanium is the last manufactured instrument of this platform and is capable of producing over 600 million base pairs with read lengths of up to 1 kbp. Roche discontinued this platform in 2013 with concomitant termination of support in the middle of 2016.

Illumina. This second-generation sequencer was released in 2005 by Solexa (now Illumina). It is based on sequencing by synthesis using fluorescently labeled dye-terminators; the process involves clonal amplification (known as bridge amplification) of adaptor-ligated DNA fragments on a glass slide surface or flow cell (Bentley et al. 2008). Using a strategy known as cyclic reversible termination, bases are identified one at a time through a cycle of base incorporation, washing, imaging and cleavage. The Illumina platform consists of a variety of instruments and is the most widely used technology as it provides the highest throughput, has the lowest error rate and is the most cost effective among currently available HTS platforms. The recently released NovaSeq 6000 system can deliver 20 billion reads per run with a maximum paired read length of 150 bp.

SOLiD. Originally released in 2007 by Applied Biosystems, which became Life Technologies after merging with Invitrogen in 2008 and now Thermo Fisher Scientific after its acquisition in 2014), this system utilizes a sequence by ligation method using DNA ligase. Repeated cycles of ligation of fluorescent labeled probes, washing of nonligated probes, and imaging (Valouev et al. 2008) ultimately determines the nucleotide sequence with each base read twice. The technology provides the second highest throughput after Illumina but only accommodates 75 bp as the longest read length (100 bp for paired end read) and without any updated models since the release of 5500xl W instrument in 2013, which is capable of delivering up to 6 billion reads per run.

Ion Torrent. This technology bears similarities to the Roche 454 pyrosequencing platform in that amplification of an adaptor-ligated fragment is performed on a bead using emulsion PCR. In addition, the sequence by synthesis reaction is done on each bead-containing microwell to determine the sequence. Unlike pyrosequencing, base determination is made by measuring pH changes, resulting from the release of hydrogen ions during base incorporation (Rothberg et al. 2011). These pH changes are converted into a voltage signal, the amplitude of which is proportional to the number of bases added sequentially in each cycle. The Ion PGM 314, 316, 318, and S5 instruments can generate 400-bp read length. Currently, the Proton 1, the latest instrument, produces the highest throughput (up to 80 million reads) but still is lower than that output of Illumina and SOLiD systems.

Third-generation sequencing platforms. PacBio-Illumina is the only platform that offers both second- and third-generation sequencing capability. Originally released in late 2010, PacBio (now part of Illumina) is the most widely used technology of its class. In addition to its ability to sequence single molecules, it uses hairpin adaptors to form a closed ssDNA template, SMRTbell (from the acronym SMRT, which stands for single molecule, real-time sequencing) (Rhoads and Au 2015). Two key sequencing features of the technology worth noting are (i) the SMRTbell template is placed in a zeptoliter-sized chamber, the zero-mode waveguide (ZMW) with an attached single polymerase molecule at the bottom of the

chamber and (ii) the addition of fluorescent-labeled nucleotides at the phosphate group is detected in real time. The latest instrument Sequel, can deliver up to 370,000 reads. This technology generates the highest read length (at least 20 kb) but also has high error rate of base incorporation.

Oxford Nanopore. Another single molecule sequencing technology is the nanopore platform, with Oxford Nanopore leading its development and commercialization. Although its commercialization was made after PacBio technology, the use of nanopore technology for sequencing is already three decades old (Deamer et al. 2016). In the Oxford Nanopore, characteristic changes in current induced are measured as the bases pass through a biologic nanopore, anchored on to it by a molecular motor protein (Lu et al. 2016). Currently, the PromethION device offers the highest throughput with about 144,000 nanopore channels compared with the three other instruments available in this platform. Overall, this technology still has lower throughput and higher error rate than its PacBio counterpart and significantly less throughput than the Illumina platform, yet a platform like MinION can provide cost effective, real-time results if no large datasets are needed.

Precommercial platforms. Two second-generation sequencing platforms are in their precommercial stage (Levy and Myers, 2016), GENIUS (from Genapsys: <http://www.genapsys.com>) and GeneReader (from Qiagen: <https://www.qiagen.com/us/shop/sequencing/qiagen-generader-platform>). The GENIUS platform, which uses a semiconductor chip similar to the Ion Torrent, promises to deliver low-cost instruments while able to sequence human genome-sized DNA (3×10^9 bp) for under \$1,000.

Four precommercial platforms are single molecule sequencing technologies and will be coming from Roche (<https://sequencing.roche.com/en/technology-research/technology/nanopore-sequencing.html>), Base 4 (<http://www.base4.co.uk/>), Quantum Biosystems (<https://www.quantumbiosystems.com/>), and SeqLL (<http://seqll.com/>). SeqLL will deliver an improved version of Helicos platform (Braslavsky et al. 2003). The remaining three utilize novel methods, greatly improving accuracy and reducing cost. Roche's single molecule sequencing will be a nanopore platform based on Genia's technology (Fuller et al. 2016), Quantum Biosystems will deliver a novel method based on gating-nanopore method, using tunneling current measurement, and Base4 will employ microdroplet sequencing (where each nucleotide, released after pyrophosphorylation, is captured in a microdroplet and optically detected).

Applications in plant virus discovery and detection. The use of HTS in plant virology was first reported in 2006 (Wren et al. 2006) and 2008 (Melcher et al. 2008; Muthukumar et al. 2008) in efforts to unravel diversity of viruses in noncultivated plant ecosystems. In 2009, four groups used HTS on different agricultural crops (Adams et al. 2009; Al Rwahnih et al. 2009; Donaire et al. 2009; Kreuze et al. 2009). Three groups identified new viruses using technologies for detection of known viruses (Kreuze et al. 2009), for investigating the etiology of a disease (Al Rwahnih et al. 2009) or both (Adams et al. 2009). Interestingly, whereas the two etiological studies identified novel viruses [grapevine Syrah virus (GSyV-1) from a declining 'Syrah' wine grape (Al Rwahnih et al. 2009) and gayfeather mild mottle virus (GMMV) from a flowering plant *Liatris spicata* (Adams et al. 2009)], the conclusions drawn differ markedly. The lack of other viruses identified from the HTS reads indicated that GMMV was closely associated with disease (and perhaps the causal agent), whereas the presence of mixed infections of diverse viruses in the declining 'Syrah' vine precluded the association of GSyV-1 with the disease. The work of Donaire et al. (2009) did not result in discovery of a novel virus, but showed that small RNA (sRNA) profiles generated by HTS from diverse groups of single-stranded RNA and DNA viruses can also be used for detection.

Different types of nucleic acid templates have been used in HTS-based plant virus discovery and detection studies. This includes

sRNA, total RNA, double-stranded RNA (dsRNA) or preparations enriched for virus-like particles (VLPs) (Roossinck et al. 2015; Wu et al. 2015). Recently, a comparison between two RNA templates to detect an ssRNA virus, citrus tristeza virus (CTV), in grapefruit indicated that ribo-depleted total RNA was superior to sRNA in achieving higher read coverage of the virus genome (Visser et al. 2016a). On the contrary, sRNA yielded higher virus recovery reads compared with ribo-depleted total RNA in the case of ssDNA viruses and viroids (Pecman et al. 2017). Additionally, if the target viruses for detection are DNA viruses that have circular (ssDNA viruses within *Geminiviridae* or *Nanoviridae*) or pseudocircular (dsDNA such as members of the family *Caulimoviridae*) genomes, a specific enrichment involves the use of rolling circle amplification (Idris et al. 2014; Rosario et al. 2013). Regardless of the template, there is a growing number of plant viruses discovered through HTS in agricultural crops (Barba et al. 2014; Roossinck et al. 2015; Wu et al. 2015). In a span of 12 years (2006 to 2018) since its first application in plant virology, there are hundreds of plant viruses characterized by HTS (partial list can be found in Roossinck et al. 2015 and Wu et al. 2015). The agricultural importance of the vast majority of these plant viruses is not known but some are closely associated with disease (Table 1). With the widespread use of HTS in plant virus research, this trend is likely to continue until the virome of most, if not all, agricultural crops is documented. A situation now exists where given the massive amount of viruses identified, it is critical to study their biology and their role in disease, primarily because of its implications on trade; a topic that will be discussed later in this review.

Whereas subsequent studies on many new viruses have not been pursued after their initial characterization, a few examples highlight successful elucidation of disease etiology and other related studies after discovery by HTS. Two diseases exemplify the role of HTS in determining etiology and are worth elaborating here. Grapevine red blotch disease was first observed in 2008 in a *Vitis vinifera* 'Cabernet Sauvignon' vineyard in California and 'Cabernet franc' vineyard in New York (Sudarshana et al. 2015) with symptoms resembling that of grapevine leafroll disease. The complete genome sequence of a virus associated with the disease, initially named grapevine red blotch-associated virus (GRBaV, now grapevine red blotch virus, GRBV), was determined simultaneously by HTS and rolling circle amplification in 2012 (Al Rwahnih et al. 2012; Krenz et al. 2012) and subsequently reported elsewhere using HTS (Poojari et al. 2013; Sequin et al. 2014b). Prevalence and population structure across the U.S. grape growing regions were determined (Krenz et al. 2014) and an infectious clone of the virus induced typical disease symptoms (Cieniewicz et al. 2017; Fuchs et al. 2015; Yepes et al. 2018). Another example is rose rosette, a disease that had long been presumed to be caused by a virus. Although this disease was described in the 1940s, the genome of the virus (named rose rosette virus [RRV]) was sequenced by HTS, along with evidence of its strong association with the disease (Laney et al. 2011) and consequent demonstration of Koch's postulate, thereby showing RRV as the causal agent of the disease (Di Bello et al. 2015).

Studies on genetic diversity of viruses are important in understanding virus evolution (Roossinck 2017). The use of HTS in virus population studies has been largely limited to determining the mutational landscapes of a virus, specifically, identifying single nucleotide polymorphisms (SNPs) across the virus genome (Huang et al. 2015; Katsiani et al. 2017; Kinoti et al. 2017; Kutnjak et al. 2015; Simmons et al. 2012). With the expected improvement of single molecule sequencing platform of HTS, the potential for HTS to accurately identify whole genome variants within the population is promising. In contrast, genetic diversity studies of plant viruses are useful in the development of robust virus detection protocols. Although there is no doubt about the effectiveness of HTS for virus detection, as has been demonstrated in various studies (Adams et al. 2009; Donaire et al. 2009; Eichmeier et al. 2016; Kreuze et al.

2009), the technology is still expensive for routine use, however it could be very effectively used in quarantine facilities (Al Rwahnih et al. 2015; Rott et al. 2017). Nevertheless, data produced by HTS could be used to either validate the broad-spectrum capability of current detection methods (Di Bello et al. 2018) or improve existing protocols to detect newly identified variants of the virus (Marais et al. 2014, 2015a, b).

ANALYSIS OF HTS DATA FOR VIRUS DETECTION

Prior to the advent of HTS, analysis of Sanger sequencing data for virus detection was straightforward—(i) the sequence quality is first verified by the presence of well-defined chromatogram peaks and (ii) subsequent comparison with database to reveal the identity of the sequenced material. In the case of HTS data, the large amount of sequence reads produced in a single run (at least an order of magnitude greater than a single Sanger sequencing reaction) necessitate complex data preprocessing and processing steps before meaningful information can be extrapolated. Implemented using a high performance computer that executes instructions from a series of algorithms compiled in a unified bioinformatics software package, these steps are generally referred to as workflows or pipelines and are briefly describe below.

Available pipelines. A typical HTS data set is stored in FASTQ format, which principally contains the sequence order for each read and quality score of each base in a read. The quality control is based on Phred quality scores, a system used for Sanger sequencing. A commonly used program, FASTQC, generates comprehensive quality control reports, which could be used for subsequent trimming of low quality reads. FASTQC is written in Java and therefore could be used in major operating systems (Windows, Mac, and Linux). The program (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) is open access and can be run as a stand-alone application or be incorporated into a larger pipeline.

Raw sequence data are trimmed not only based on quality scores but also to remove adapter sequences. Additionally, if data originates from a library comprised of multiple samples, demultiplexing is done based on barcode sequences unique to each PCR primer. There is available software for this purpose, adequately summarized in Blawid et al. (2017). After trimming, sequences are assembled into contigs (after removal of host sequences when available) and annotated; most commonly using homology search programs such as BLAST. Another method of annotating contigs is by “profile methods” such as PSI-BLAST and profile Hidden Markov models or HMMs (Skewes-Cox et al. 2014). Alternatively, if the objective is to detect known viruses (and its variants), the trimmed sequences are mapped to a database of virus sequences (reference-guided assembly). Various de novo assembly (<https://omictools.com/genome-assembly-category>) and read mapping softwares are available for download in the links provided.

Pipelines developed to detect and discover viruses consist of software packages organized to perform stepwise data processing, with the option to trim based on quality, adapter sequences or both. These pipelines include Visitor (Antoniewski 2011), VirusFinder (Wang et al. 2013) and VirusFinder 2 with VERSE algorithm (<https://bioinfo.uth.edu/VirusFinder/>), VirusSeq (Chen et al. 2013), VirusHunter (Zhao et al. 2013), viRome (Watson et al. 2013), VirFind (Ho and Tzanetakis 2014), SearchSmallRNA (de Andrade and Vaslin 2014), MISIS (Seguin et al. 2014a), MISIS-2 (Seguin et al. 2016), VIP (Li et al. 2016), VirusTAP (Yamashita et al. 2016), Truffle (Visser et al. 2016b), VirusSeeker (Zhao et al. 2017), VSD toolkit (Barrero et al. 2017), VirusDetect (Zheng et al. 2017), and Virtool (Rott et al. 2017). Commercial software packages developed for the analysis of SNPs, transcriptomics, de novo assembly etc. can also be used for virus detection and discovery. The most popular include CLC Genomics Workbench, Geneious and DNASTAR's Lasergene Genomics Suite. Two main drawbacks of these packages are their cost and the lack of full automation;

TABLE 1
High throughput sequencing (HTS)-mediated identification of viruses strongly associated with diseases of agricultural crops

Disease	Virus(es) identified	HTS platform	HTS template	Reference
Apple rubbery wood	Apple rubbery wood virus 1 and 2	Illumina	dsRNA	Rott et al. 2018
Blackberry yellow vein complex	Blackberry leaf mottle-associated virus	Illumina	dsRNA	Hassan et al. 2017
Carrot internal necrosis	Carrot yellow leaf virus	Illumina	Total RNA	Adams et al. 2014
Citrus chlorotic dwarf	Citrus chlorotic dwarf-associated virus	Illumina	Small RNA	Loconsole et al. 2012b
Citrus leprosis	Citrus leprosis virus cytoplasmic type 2	Illumina	Small RNA	Roy et al. 2013
Citrus vein enation	Citrus vein enation virus	Illumina	Small RNA	Vives et al. 2013
Citrus yellow vein clearing	Citrus yellow vein clearing virus	Illumina	Small RNA	Loconsole et al. 2012b
Grapevine red blotch	Grapevine redblotch virus	Illumina	dsRNA, total RNA	Al Rwahnih et al. 2012; Poojari et al. 2013
Grapevine vein clearing	Grapevine vein clearing virus	Illumina	Small RNA	Zhang et al. 2011
<i>Liatris spicata</i> (symptomatic)	Gayfeather mild mosaic virus	Roche 454	dsRNA	Adams et al. 2009
Nectarine stem pitting	Nectarine stem pitting-associated virus, nectarine virus M	Illumina	dsRNA, total RNA	Bag et al. 2015; Villamor et al. 2016
Pigeon pea sterility mosaic	Pigeon pea sterility mosaic virus	Illumina	dsRNA	Elbeaino et al. 2014a
<i>Pleione</i> sp. (symptomatic)	Pleione flower breaking virus (PLFBV)	Illumina	Total RNA	Kominek et al. 2019
Raspberry leaf blotch	Raspberry leaf blotch virus	Roche 454	dsRNA	McGavin et al. 2012
Rose rosette	Rose rosette virus	Illumina	dsRNA	Laney et al. 2011
Shirofugen stunt	Little cherry virus 1	Roche 454	dsRNA	Candresse et al. 2013

notably, the CLC Genomics Workbench allows for the creation of a workflow by which results for each step can be channeled to the next analysis. However, results of homology search via BLAST still need to be sorted out by the user.

Pipelines such as viRome (Watson et al. 2013), SearchSmallRNA (de Andrade and Vaslin 2014), MISIS (Seguin et al. 2014a), and MISIS-2 (Seguin et al. 2016) can reconstruct viral genomes from small RNA HTS data as well as identify siRNA hot spots in the virus genome. These pipelines are essentially mapping softwares that have graphical user interface (GUI) but are limited to detection of only known viruses. Another approach for virus detection is through the use of e-probe diagnostic nucleic acid assay (EDNA), which has been applied to several pathosystems (Jooste et al. 2017; Stobbe et al. 2013, 2014; Visser et al. 2016b). The main advantage of this method is its lower computational resource requirement than their de novo assembly-based detection counterparts allowing for faster detection of known viruses. Conversely, some pipelines such as Visitor (Antoniewski 2011), VirusHunter (Zhao et al. 2013), VirusFinder 2 with VERSE algorithm (<https://bioinfo.uth.edu/VirusFinder/>), VIP (Li et al. 2016), VirusSeeker (Zhao et al. 2017), and Virtool (<https://www.virtool.ca/>) have the capacity to detect new viruses but do not have a GUI and need skilled personnel adept with the command line interface of the Linux operating system. Two other pipelines designed to detect known human and animal viruses include a Linux-based pipeline known as VirusSeq (Chen et al. 2013), and a software that has GUI capability known as

VirusTAP (Yamashita et al. 2016). On the other hand, three web-based pipelines with GUI that have been designed and tested for plant virus detection and discovery include VirFind (Ho and Tzanetakis 2014) (<http://virfind.org/j/>), VirusDetect (Zheng et al. 2017) (<http://bioinfo.bti.cornell.edu/tool/VirusDetect/>), and VSD toolkit (Barrero et al. 2017). VirusDetect and VSD toolkit work specifically for virus detection and discovery from small RNA sequencing data. Additionally, the functionality of the VSD toolkit as a web-based pipeline requires the use of an open source internet-based analytical environment called Yabi (Hunter et al. 2012) for the implementation of its virus detection workflows. Finally, VirFind uses a homology search of de novo assembled contigs for both the detection of known and novel viruses (Ho and Tzanetakis 2014).

A question of which pipelines to use depends on intended goal in virus detection (Table 2). While pipelines that mapped small RNA reads such as viRome, SearchSmallRNA, MISIS, and MISIS-2 can be used for virus detection, they are primarily designed to visualize the siRNA hotspots within virus genome. Arguably, the e-probe approach is ideal for detection of known viruses largely because of the speed at which analysis is completed (Stobbe et al. 2013, 2014). This approach could be valuable to nursery certification programs and is already notably used in targeted virus detection on specific crops such as grapevines (Visser et al. 2016b) and citrus (Jooste et al. 2017). On the other hand, pipelines that are capable of detecting new viruses have greatest utility in etiological studies of virus-like diseases and virus detection in quarantine

TABLE 2
Available pipelines for plant virus detection and discovery

Pipeline	Detection of known viruses	New virus discovery	Remarks	Reference	Website
Linux based					
VIP	Yes	Yes	Could be used for plant virus detection	Li et al. 2016	https://github.com/keylabivdc/VIP
VirusHunter	Yes	Yes	Could be used for plant virus detection	Zhao et al. 2013	https://www.ibridgenetwork.org/wustl/virushunter
VirusSeeker	Yes	Yes	Could be used for plant virus detection	Zhao et al. 2017	https://wupathlabs.wustl.edu/virusseeker/
VirusFinder 2 with VERSE algorithm	Yes	Yes	Could be used for plant virus detection	Wang et al. 2015a	https://bioinfo.uth.edu/VirusFinder/
Virtool	Yes	Yes	Tested for plant virus detection	Rott et al. 2017	https://www.virtool.ca/
Truffle	Yes	No	Tested for plant virus detection; e-probe based detection; has GUI	Visser et al. 2016b	https://sourceforge.net/projects/truffle/
Web-based GUI					
VirFind	Yes	Yes	Tested for plant virus detection	Ho and Tzanetakis 2014	http://virfind.org/j/
VirusDetect	Yes	Yes	Tested for plant virus detection	Zheng et al. 2017	http://virusdetect.feilab.net/cgi-bin/virusdetect/index.cgi
VSD toolkit	Yes	Yes	Tested for plant virus detection; needs open source internet-based analytical environment called Yabi (Hunter et al. 2012)	Barrero et al. 2017	Source code for Yabi: https://github.com/muccg/yabi
VirusTAP	Yes	No	Could be used for plant virus detection	Yamashita et al. 2016	https://gph.niid.go.jp/cgi-bin/virustap/index.cgi

facilities, particularly clean plant programs. In this regard, there are currently more pipelines available for virologists who are adept in using the command line interface of the Linux operating system. Most notable in this group is the Virtool pipeline, which has been tested for plant viruses (Rott et al. 2017). Additionally, the combination of read mapping approach for detection of known viruses and a profile search (based on HMM) of annotating de novo assembled contigs for new virus discovery allows completion of analysis by Virtool relatively faster than its counterparts. On the other hand, plant virologists who do not have thorough training in bioinformatics could benefit from pipelines that have GUI such as VirFind (Ho and Tzanetakis 2014), VirusDetect (Zheng et al. 2017), and VSD (Barrero et al. 2017). Whereas the VirusDetect and VSD pipelines are limited to analysis of small RNA data, VirFind accepts data generated from other platforms, and with its recent update of switching from Velvet (Zerbino and Birney 2008) to both SPADes (Bankevich et al. 2012) and Trinity (Grabherr et al. 2011) programs to generate de novo assembled contigs, VirFind has improved to accommodate virus detection from small RNA data (T. Ho and I. Tzanetakis, *unpublished data*).

IMPACT ON CERTIFICATION AND PLANT MOVEMENT

Movement of plant material has long been recognized as a potential pathway for long distance spread of plant pathogens and therefore has led to the development of plant quarantine and certification programs. There are numerous examples of rapid spread of pathogens of agricultural crops via plant material, the most famous of which is the *Phytophthora infestans* introduction to Europe that led to the Irish Potato Famine in the mid-1800s (Schumann 1991). Plum pox virus spread across Europe during the 20th century (Cambra et al. 2006) and was introduced into North America in the 1990s, with over 53 million U.S. dollars spent to eradicate this virus from a small area in Pennsylvania (Welliver 2012). In the 21st century Huanglongbing or citrus greening caused by '*Candidatus Liberibacter asiaticus*' has decimated the citrus industry in Florida and other areas around the world resulting in significant decreases in production (Singerman and Useche 2016) and *Xylella fastidiosa* is severely impacting olive production in Italy after its introduction in 2013 (Abbott 2017).

Quarantine and certification programs have become much more sophisticated over the past 50 years because of the introduction of laboratory tests such as ELISA and PCR with its variants, allowing for sensitive pathogen detection (Rowhani et al. 2005). However, despite these advances, biological indexing is still considered the gold standard for detection of systemic pathogens (viruses, viroids, phytoplasmas, and systemic bacteria) since the laboratory tests can be very specific and unable to identify diverse isolates and strains. Although these tests are sensitive, they require prior knowledge of the target pathogens and therefore are not effective at detecting unknown pathogens. For this reason bioassays are still required to meet quarantine and certification standards especially for woody perennial crops; a process that can take up to 4 years for some crops. The lag time reduces the amount of material that can be processed; affecting the efficiency and sustainability of agricultural industries. A similar lag time exists in certification programs where bioassays, done by grafting onto woody indicators, delay new cultivars from entering into the nursery system and ultimately to the growers.

Even when bioassays are not needed, many laboratory tests are often required to assay for known viruses and viroids in woody crops as there are more than 70 and 80 virus and virus-like agents that infect grapevines (Dolja et al. 2017) and berry crops (Martin and Tzanetakis 2015), respectively. Other woody perennial crops such as citrus, tree fruits, and ornamentals also have a plethora of these pathogens. Therefore, the costs associated with performing this number of assays are high compared with the current cost of HTS. Additionally, these laboratory assays can give false negative

results since they are often designed against one or a few isolates of a virus and can fail to detect diverse variants (Poudel and Tzanetakis 2013; Poudel et al. 2012).

With the use of HTS, quarantine and certification programs are poised for improvements in terms of time and cost. There have been comparisons of HTS with traditional methods including bioassays, ELISA, and various PCR formats. HTS has proven as good as or better than bioassays in detecting viruses and viroids in woody plants (Al Rwahnih et al. 2015; Rott et al. 2017).

There are key questions to consider when using HTS: How many sequences are required to call a plant infected? How much of a virus genome needs to be identified before a plant is considered positive? What is the best template for cDNA libraries? What is the best sequencing platform? Are some bioinformatics pipelines better than others? Before HTS can be used in quarantine and certification programs, there needs to be documentation on its sensitivity, reproducibility and the range of crops that it is applicable to. A challenging task is the development of standard operating procedures (SOPs) to gain confidence in the data presented by different laboratories. As an example, we will present that of virus nucleic acid enrichment approaches (small RNA, double-stranded RNA, total RNA minus ribosomal RNA, purified virus) before library construction. Each of them potentially has specific biases to particular viruses. Within each enrichment approach, there are also different protocols that can produce very different libraries. Negative-strand RNA viruses generally have lower titer in plants compared with positive-strand RNA viruses, which made a dsRNA enrichment procedure unsuccessful for the detection of those viruses (Ho and Tzanetakis 2014). Commercial library prep kits are available and many prep protocols are published by different labs. To achieve optimal results, it may be that each sample should be processed using more than one nucleic acid purification, and each HTS dataset analyzed with multiple pipelines. SOPs need to be reviewed and approved by multiple quarantine agencies if the technology is to be accepted as a method to meet international standards. How does the discovery of new viruses, not identifiable by conventional methods impact plant movement? Will there be a requirement to have some information on the biology of a virus before it is added to quarantine lists? In our opinion this should be a requirement.

While it will take time before HTS is implemented as a standard diagnostic method, it has already positively impacted the certification and registration processes, particularly in the case of perennial crops. The discovery of novel viruses as a result of using HTS has resulted in the improvement of testing procedures for virus and virus-like agents in clean plant foundation programs; this is mostly reflected by the integration of additional PCR/RT-PCR tests for the new viruses.

More significantly, HTS has facilitated the acceptance of a "provisional release" category of plant selections in clean plant programs in the United States. "Provisional release" of a plant selection is based on negative HTS results (i.e., no virus or viroid-like agent detected), which consequently, allows for the limited propagation of an HTS-negative selection in designated areas approved by the U.S. Department of Agriculture Animal and Plant Health Inspection Service (USDA-APHIS), prior to its official release pending completion of all bioassays and laboratory tests. This 'provisional release propagation' allows for the buildup of material for commercial production ahead of its official release and can significantly reduce the time before plants are available to growers. If there is a positive in the conventional techniques then plants are destroyed at the nursery's expense.

FUTURE PERSPECTIVES

Virus detection and discovery using HTS is improving following the advancement of high performance computing (HPC) hardware and bioinformatics. Until recently, there were few labs able to

perform the steps described in this chapter; those having access to dedicated HPC facilities and bioinformaticians capable of writing computer scripts. Currently, there is ample published literature on virus detection and discovery pipelines. Yet most of them are a collection of bioinformatics programs linked together and there are only a few public hardware services to run the pipelines. Recent automated bioinformatics services for virus discovery with a public web interface such as VirFind (<http://virfind.org/j>) and VirusDetect (<http://virusdetect.feilab.net>) aim to resolve those issues.

Although virus discovery using HTS is a powerful tool, there is need for verification using alternative methods. One example are badnaviruses, dsDNA viruses that can be found in two forms: the first one is when the virus integrates into the plant genome and does not cause disease; and the other, episomal, when the virus actively replicates in the cell and may cause disease. Current bioinformatics pipelines can discover a virus-like sequence but cannot distinguish the two forms; not allowing for a clear view of the significance of the findings and therefore additional steps are needed to identify the virus form (Shahid et al. 2017). Likewise, while not known to integrate into the plant genome, evidence of the presence of a circular genome via rolling circle amplification is also needed for ssDNA viruses. Another concern is low sequencing coverage and depth. A hit is often easy to call if the sequencing depth and coverage is high (i.e., assembled virus contig encompass near full genome of the virus) but there is no current standard for low virus read depth and coverage. In this case, a confirmatory test is necessary. On the other hand, HTS can produce unverifiable results. Virus-like sequences in HTS datasets may not be confirmed by other detection methods and it remains unknown whether these sequences were assembly artifacts or the virus was in low enough titer to be undetectable by other methods.

Undoubtedly, the future of HTS lies in the third-generation sequencing platforms (PacBio-Illumina and Oxford Nanopore). This technology does not only hold promise in whole virus variant identification but could also provide better resolution for plant virus detection than their second-generation counterparts. In theory, the issue of low sequence coverage and depth could be mitigated by using third-generation platforms. Because of its inherent characteristic to sequence single molecule, virus sequences detected in this platform would cover a large portion, if not the entirety, of the virus genome. While the major drawback of third-generation sequencing platforms is its high error rates, a method to circumvent this issue is through hybrid sequencing (Au et al. 2012; Koren et al. 2012; Salmela and Rivals 2014) which includes a supplemental use of second-generation sequencing platform to correct for error rates. Additionally, the recent acquisition of PacBio by Illumina could result in the rapid improvement of the error rate of this platform. Oxford Nanopore is also currently able to directly sequence RNA, which theoretically decreases its error rate.

At present, the backbone of virus discovery pipelines are core bioinformatics tools that were written for generic purposes such as sequence mapping (e.g., bowtie [Langmead and Salzberg 2012]), sequence de novo assembly (e.g., Velvet, Trinity, SPAdes [Grabherr et al. 2011; Nurk et al. 2013; Zerbino and Birney 2008]), and NCBI BLAST (Altschul et al. 1997). There is no special core tool written specifically for virus discovery purposes. There is also a need to have a centralized virus sequence database for higher virus discovery resolution. Currently, when constructing a virus sequence database for comparison purpose, the programmer extracts viral sequences from GenBank; nt for nucleotides and nr for amino acids. There is also a much smaller refseq database from GenBank containing only reference virus sequences that can be used for virus detection in order to conserve computer resources. Many virus sequences remain unpublished until manuscripts are published. Curation of a centralized virus sequence database could ameliorate the delay in sequence release that affect sequence analysis quality.

As technology evolves so do the opportunities for better assessment of plant health status. This was the case of ELISA, a technique that revolutionized plant virus detection in the 1970s and PCR and its variants in the 1990s. The same is true for HTS. It has transformed the way we approach testing for pathogens and particularly viruses. The additional advantage that HTS provides compared with bioindicators and all previously used technology is the capacity to test for all known viruses but also identify agents that were unknown, in a single assay.

Historically, the Achilles heel of HTS for detection and discovery of plant viruses has been the lack of bioinformatics tools and the amount of virus sequences in databases. The former could be addressed with the new generation of biological scientists, graduating with high quality bioinformatics skills in the near future. In addition, biocomputing capacity continues to expand and as a consequence, so is our ability to explore the plant virome. As databases expand and algorithms become more robust, a greater ability will be gained to identify viruses that show minimal similarity to those known today. The computational power expansion will allow for memory-intense comparisons based on the tertiary structure of proteins. Protein structures being more conserved than primary sequences will allow for the identification of viruses that have no obvious protein orthology to known viruses, yet their protein structures will point to their presence in plants. New innovative algorithms for discovering viruses should be created to search not only for sequence homology, but also conserved virus domains (Marchler-Bauer et al. 2017) or virus protein folding motifs (Sobhy 2016).

As has been the case in the past, new technologies have to pass rigorous tests and prove equal or better than the current standards before approval for routine use. With globalization, movement of plant material has become a pressing issue and often a point of dispute between countries. HTS could potentially resolve these issues given its capacity to detect and discover. There have been studies (Al Rwahnih et al. 2015; Rott et al. 2017) that have passed peer-review and point to the superiority of the technology when compared with other methodologies. The next step is to integrate HTS in certification programs around the world and accelerate the availability of virus-tested propagation material to producers (Gergerich et al. 2015; Martin and Tzanetakis 2015). As we move forward, states, provinces and countries have to pass rules and legislation that allows for the use of HTS technology in quarantine schemes. This will take time as there is a need to educate regulators on the merits of the technology but the consensus is that HTS is the technology of the future for certification programs around the globe (Barrero et al. 2017; Martin et al. 2016).

There is an obvious need to modernize and harmonize rules and procedures among states and countries to reflect the changes that HTS brings into certification pipelines. This is of particular importance given that globalization has allowed for the expansion and presence of agricultural companies to several countries and continents. If those steps are not taken, HTS may have the opposite results of those predicted. Instead of allowing for the timely dispersal of high quality, virus-tested material, it will cause border closure. The new viruses that have and will be identified cause mayhem for plant quarantine and certification without appropriate vetting based on evidence-driven science and not on fear of a new virus that may have no effect on crops or the ecosystem.

LITERATURE CITED

- Abbott, A. 2017. Italy rebuked for failure to prevent olive-tree tragedy. *Nature* 546:193-194.
- Adams, I. P., Glover, R. H., Monger, W. A., Mumford, R., Jackeviciene, E., Navalini-skiene, M., Samuitiene, M., and Boonham, N. 2009. Next-generation sequencing and metagenomic analysis: A universal diagnostic tool in plant virology. *Mol. Plant Pathol.* 10:537-545.
- Adams, I. P., Skelton, A., Macarthur, R., Hodges, T., Hinds, H., Flint, L., Nath, P. D., Boonham, N., and Fox, A. 2014. Carrot yellow leaf virus is associated with carrot internal necrosis. *PLoS One* 9:e109125.

- Al Rwahnih, M., Daubert, S., Golino, D., Islas, C., and Rowhani, A. 2015. Comparison of next-generation sequencing versus biological indexing for the optimal detection of viral pathogens of grapevines. *Phytopathology* 105: 758-763.
- Al Rwahnih, M., Daubert, S., Golino, D., and 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., Dave, A., Anderson, M., Uyemoto, J. K., and Sudarshana, M. R. 2012. Association of a circular DNA virus in grapevines affected by red blotch disease in California. In: *Proc. 17th Congr. Int. Counc. Study of Virus and Virus-Like Diseases of the Grapevine (ICVG)*, Davis, CA.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389-3402.
- de Andrade, R. R., and Vaslin, M. F. 2014. SearchSmallRNA: A graphical interface tool for the assemblage of viral genomes using small RNA libraries data. *Virol. J.* 11:45.
- Antoniewski, C. 2011. Visitor, an informatic pipeline for analysis of viral siRNA sequencing datasets. *Methods Mol. Biol.* 721:123-142.
- Au, K. F., Underwood, J. G., Lee, L., and Wong, W. H. 2012. Improving PacBio long read accuracy by short read alignment. *PLoS One* 7:e46679.
- Bag, S., Al Rwahnih, M., Li, A., Gonzalez, A., Rowhani, A., Uyemoto, J. K., and Sudarshana, M. R. 2015. Detection of a new luteovirus in imported nectarine trees: A case study to propose adoption of metagenomics in post-entry quarantine. *Phytopathology* 105:840-846.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Pribelski, A. D., Pyskhin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., and Pevzner, P. A. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19:455-477.
- Barba, M., Czosnek, H., and Hadidi, A. 2014. Historical perspective, development, and applications of next-generation sequencing in plant virology. *Viruses* 6:106-136.
- Barrero, R. A., Napier, K. R., Cunningham, J., Liefing, L., Keenan, S., Frampton, R. A., Szabo, T., Bulman, S., Hunter, A., Ward, L., Whattam, M., and Bellgard, M. I. 2017. An internet-based bioinformatics toolkit for plant biosecurity diagnosis and surveillance of viruses and viroids. *BMC Bioinformatics* 18:26.
- Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., and Bignell, H. R. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456:53-59.
- Blawid, R., Silva, J. M. F., and Nagata, T. 2017. Discovering and sequencing new plant viral genomes by next-generation sequencing: description of a practical pipeline. *Ann. Appl. Biol.* 170:301-314.
- Braslavsky, I., Hebert, B., Kartalov, E., and Quake, S. R. 2003. Sequence information can be obtained from single DNA molecules. *PNAS* 100: 3960-3964.
- Buermans, H. P. J., and den Dunnen, J. T. 2014. Next generation sequencing technology: Advances and applications. *Biochim. Biophys. Acta* 1842: 1932-1941.
- Cambra, M., Capote, N., Myrta, A., and Llacer, G. 2006. Plum pox virus and the estimated costs associated with sharka disease. *OEPP/EPPO Bull.* 36: 202-204.
- Candresse, T., Marais, A., Faure, C., and Gentit, P. 2013. Association of *Little cherry virus 1* (LChV1) with the shirofugen stunt disease and characterization of the genome of a divergent LChV1 isolate. *Phytopathology* 103: 293-298.
- Chen, Y., Yao, H., Thompson, E. J., Tannir, N. M., Weinstein, J. N., and Su, X. 2013. VirusSeq: Software to identify viruses and their integration sites using next-generation sequencing of human cancer tissue. *Bioinformatics* 29:266-267.
- Cieniewicz, E., Perry, K., and Fuchs, M. 2017. Grapevine red blotch: Molecular biology of the virus and management of the disease. Pages 303-314 in: *Grapevine Viruses: Molecular Biology, Diagnostics and Management*. B. Meng, G. Martelli, D. Golino, and M. Fuchs, eds. Springer, Cham.
- Deamer, D., Akeson, M., and Branton, D. 2016. Three decades of nanopore sequencing. *Nat. Biotechnol.* 34:518-524.
- Di Bello, P. L., Ho, T., and Tzanetakis, I. E. 2015. The evolution of emaraviruses is becoming more complex: Seven segments identified in the causal agent of Rose rosette disease. *Virus Res.* 210:241-244.
- Di Bello, P. L., Thekke-Veetil, T., Druciarek, T., and Tzanetakis, I. E. 2018. Transmission attributes and resistance to rose rosette virus. *Plant Pathol.* 67: 499-504.
- Dolja, V. V., Meng, B., and Martelli, G. P. 2017. Evolutionary aspects of grapevine virology. Pages 659-688 in: *Grapevine Viruses: Molecular Biology, Diagnostics and Management*. B. Meng, G. Martelli, D. Golino, and M. Fuchs, eds. Springer, Cham.
- Donaire, L., Wang, Y., González-Ibeas, D., Mayer, K. F., Aranda, M. A., and Llave, C. 2009. Deep sequencing of plant viral small RNAs reveals effective and widespread targeting of viral genomes. *Virology* 392: 203-214.
- Eichmeier, A., Komínková, M., Komínek, P., and Baránek, M. 2016. Comprehensive virus detection using next generation sequencing in grapevine vascular tissues of plants obtained from the wine regions of Bohemia and Moravia (Czech Republic). *PLoS One* 11:e0167966.
- Elbeaino, T., Digiario, M., Uppala, M., and Sudini, H. 2014a. Deep sequencing of pigeonpea sterility mosaic virus discloses five RNA segments related to emaraviruses. *Virus Res.* 188:27-31.
- Elbeaino, T., Giampetruzzi, A., De Stradis, A., and Digiario, M. 2014b. Deep-sequencing analysis of an apricot tree with vein clearing symptoms reveals the presence of a novel betaflexivirus. *Virus Res.* 181:1-5.
- Feng, Y., Zhang, Y., Ying, C., Wang, D., and Du, C. 2015. Nanopore-based fourth-generation DNA sequencing technology. *Genomics Proteomics Bioinform.* 13:4-16.
- Fuchs, M. F., Krenz, B., Yepes, L. M., Thompson, J., McLane, H., and Perry, K. L. 2015. Is Grapevine red blotch-associated virus the causal agent of red blotch disease? Pages 72-73 in: *Proceedings of the 18th Congress of the International Council for the Study of Virus and Virus-like Diseases of the Grapevine*, Ankara, Turkey.
- Fuller, C. W., Kumar, S., Porel, M., Chien, M., Bibillo, A., Stranges, B., Dowart, M., Tao, C., Li, Z., and Guo, W. 2016. Real-time single-molecule electronic DNA sequencing by synthesis using polymer-tagged nucleotides on a nanopore array. *PNAS* 113:5233-5238.
- Gergerich, R., Welliver, R. A., Osterbauer, N., Kamenidou, S., Martin, R., Golino, D. A., Eastwell, K. C., Fuchs, M., Vidalakis, G., and Tzanetakis, I. E. 2015. Safeguarding fruit crops in the age of agricultural globalization. *Plant Dis.* 99:176-187.
- Giampetruzzi, A., Roumi, V., Roberto, R., Malossini, U., Yoshikawa, N., La Notte, P., Terlizzi, F., Credi, R., and Saldarelli, P. 2012. A new grapevine virus discovered by deep sequencing of virus and viroid-derived small RNAs in cv. Pinot gris. *Virus Res.* 163:262-268.
- Glenn, T. 2011. Field guide to next-generation DNA sequencers. *Mol. Ecol. Resour.* 11:759-769.
- Goodwin, S., McPherson, J. D., and McCombie, R. 2016. Coming of age: Ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* 17: 333-351.
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., and Regev, A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29:644-652.
- Hassan, M., Di Bello, P., Keller, K. E., Martin, R. R., Sabanadzovic, S., and Tzanetakis, I. E. 2017. A new, widespread emaravirus discovered in blackberry. *Virus Res.* 235:1-5.
- Heather, J. M., and Chain, B. 2016. The sequence of sequencers: The history of sequencing DNA. *Genomics* 107:1-8.
- Ho, T., and Tzanetakis, I. E. 2014. Development of a virus detection and discovery pipeline using next generation sequencing. *Virology* 471-473: 54-60.
- Huang, L., Li, Z., Wu, J., Xu, Y., Yang, X., Fan, L., Fang, R., and Zhou, X. 2015. Analysis of genetic variation and diversity of Rice stripe virus populations through high-throughput sequencing. *Front. Plant Sci.* 6:176.
- Hunter, A., Macgregor, A. B., Szabo, T. O., Wellington, C. A., and Bellgard, M. I. 2012. Yabi: An online research environment for grid, high performance and cloud computing. *Source Code Biol. Med.* 7:1.
- Idris, A., Al-Saleh, M., Piatek, M. J., Al-Shahwan, I., Ali, S., and Brown, J. K. 2014. Viral metagenomics: Analysis of bego-moviruses by Illumina high-throughput sequencing. *Viruses* 6:1219-1236.
- Jooste, T. L., Visser, M., Cook, G., Burger, J. T., and Maree, H. J. 2017. In silico probe-based detection of citrus viruses in NGS Data. *Phytopathology* 107:988-993.
- Katsiani, A., Lamour, K., and Tzanetakis, I. E. 2017. Unveiling the population structure of rose rosette virus. (Abstr.) *Phytopathology* 107(suppl.):S37.
- Kinoti, W. M., Constable, F. E., Nancarrow, N., Plummer, K. M., and Rodoni, B. 2017. Analysis of intra-host genetic diversity of *Prunus necrotic ringspot virus* (PNRSV) using amplicon next generation sequencing. *PLoS One* 12: e0179284.
- Kominek, P., Massart, S., Pham, K., van Leeuwen, P., and Kominkova, M. 2019. Characterisation of a novel virus infecting orchids of the genus *Pleione*. *Virus Res.* 261:56-59.
- Koren, S., Schatz, M. C., Walenz, B. P., Martin, J., Howard, J. T., Ganapathy, G., Wang, Z., Rasko, D. A., McCombie, W. R., Jarvis, E. D., and Phillippy, A. M.

2012. Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nat. Biotechnol.* 30:693-700.
- Krenz, B., Thompson, J., Fuchs, M. F., and Perry, K. L. 2012. Complete genome sequence of a new circular DNA virus from grapevine. *J. Virol.* 86: 7715.
- Krenz, B., Thompson, J. R., McLane, H. L., Fuchs, M., and Perry, K. L. 2014. *Grapevine red blotch-associated virus* is widespread in the United States. *Phytopathology* 104:1232-1240.
- Kreuze, J. F., Perez, A., Untiveros, M., Quispe, D., Fuentes, S., Barker, I., and 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.
- Kutnjak, D., Rupar, M., Gutierrez-Aguirre, I., Curk, T., Kreuze, J. F., and Ravnikar, M. 2015. Deep sequencing of virus-derived small interfering RNAs and RNA from viral particles shows highly similar mutational landscapes of a plant virus population. *J. Virol.* 89:4760-4769.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., and Zody, M. C. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860-921.
- Laney, A., Keller, K. E., Martin, R. R., and Tzanetakis, I. E. 2011. A discovery 70 years in the making: characterization of the Rose rosette virus. *J. Gen. Virol.* 92:1727-1732.
- Langmead, B., and Salzberg, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9:357-359.
- Levy, S. E., and Myers, R. M. 2016. Advancements in next-generation sequencing. *Annu. Rev. Genomics Hum. Genet.* 17:95-115.
- Li, Y., Wang, H., Nie, K., Zhang, C., Zhang, Y., Wang, J., Niu, P., and Ma, X. 2016. VIP: An integrated pipeline for metagenomics of virus identification and discovery. *Sci. Rep.* 6:1-10.
- Loconsole, G., Onelge, N., Potere, O., Giampetruzzi, A., Bozan, O., Satar, S., De Stradis, A., Savino, V., Yokomi, R. K., and Saponari, M. 2012a. Identification and characterization of *Citrus yellow vein clearing virus*, a putative new member of the genus *Mandavivirus*. *Phytopathology* 102: 1168-1175.
- Loconsole, G., Saldarelli, P., Doddapaneni, H., Savino, V., Martelli, G. P., and Saponari, M. 2012b. Identification of a single-stranded DNA virus associated with citrus chlorotic dwarf disease, a new member in the family Geminiviridae. *Virology* 432:162-172.
- Lu, H., Giordano, F., and Ning, Z. 2016. Oxford nanopore MinION sequencing and genome assembly. *Genomics Proteomics Bioinformatics* 14:265-279.
- Marais, A., Faure, C., Couture, C., Bergey, B., Gentit, P., and Candresse, T. 2014. Characterization by deep sequencing of divergent *Plum bark necrosis stem pitting-associated virus* (PBNSPA) isolates and development of a broad-spectrum PBNSPA detection assay. *Phytopathology* 104:660-666.
- Marais, A., Faure, C., Mustafayev, E., Barone, M., Alioto, D., and Candresse, T. 2015a. Characterization by deep sequencing of *Prunus virus T*, a novel tepovirus infecting *Prunus* species. *Phytopathology* 105:135-140.
- Marais, A., Faure, C., Mustafayev, E., and Candresse, T. 2015b. Characterization of new isolates of apricot vein clearing-associated virus and of a new *Prunus*-infecting virus: Evidence for recombination as a driving force in Betaflexiviridae evolution. *PLoS One* 10:e0129469.
- Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C., Geer, L. Y., and Bryant, S. H. 2017. CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* D200-D203.
- Mardis, E. R. 2008. Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.* 9:387-402.
- Mardis, E. R. 2013. Next-generation DNA sequencing platforms. *Annu. Rev. Anal. Chem. (Palo Alto, Calif.)* 6:287-303.
- Mardis, E. R. 2017. DNA sequencing technologies: 2006–2016. *Nat. Protoc.* 12:213-218.
- Marsh, M., Tu, O., Dolnik, V., Roach, D., Solomon, N., Bechtol, K., Smietana, P., Wang, L., Li, X., Cartwright, P., Marks, A., Barker, D., Harris, D., and Bashkin, J. 1997. High-throughput DNA sequencing on a capillary array electrophoresis system. *J. Capillary Electrophor.* 4:83-89.
- Martin, R. R., Constable, F., and Tzanetakis, I. E. 2016. Quarantine regulations and the impact of modern detection methods. *Annu. Rev. Phytopathol.* 54: 189-205.
- Martin, R. R., and Tzanetakis, I. E. 2015. Control of virus diseases in berry crops. *Adv. Virus Res.* 91:271-309.
- McGavin, W. J., Mitchell, C., Cock, P. J. A., Wright, K. M., and MacFarlane, S. A. 2012. *Raspberry leaf blotch virus*, a putative new member of the genus *Emaravirus*, encodes a novel genomic RNA. *J. Gen. Virol.* 93: 430-437.
- Melcher, U., Muthukumar, V., Wiley, G. B., Min, B. E., Palmer, M. W., Verchot-Lubicz, J., Nelson, R. S., Roe, B. A., Ali, A., Thapa, V., and Pierce, M. L. 2008. Evidence for novel viruses by analysis of nucleic acids in virus-like particle fractions from *Ambrosia psilostachya*. *J. Virol. Methods* 152:49-55.
- Metzker, M. L. 2010. Sequencing technologies—the next generation. *Nat. Rev. Genet.* 11:31-46.
- Muthukumar, V., Melcher, U., Pierce, M. L., Wiley, G. B., Roe, B. A., Palmer, M. W., Thapa, V., Ali, A., and Ding, T. 2008. Non-cultivated plants of the Tallgrass Prairie Preserve of northeastern Oklahoma frequently contain virus-like sequences in particulate fractions. *Virus Res.* 149:169-173.
- Nurk, S., Bankevich, A., Antipov, D., Gurevich, A., Korobeynikov, A., Lapidus, A., Prjibelsky, A., Pyshkin, A., Sirotkin, A., Sirotkin, Y., Stepanauskas, R., McLean, J., Lasken, R., Clingenpeel, S. R., Woyke, T., Tesler, G., Alekseyev, M. A., and Pevzner, P. A. 2013. Assembling genomes and mini-metagenomes from highly chimeric reads. Pages 158-170 in: *Research in Computational Molecular Biology: 17th Annual International Conference. RECOMB 2013 Proceedings*. M. Deng, R. Jiang, F. Sun, and X. Zhang, eds. Springer Berlin, Heidelberg, Berlin.
- Pecman, A., Kutnjak, D., Gutierrez-Aguirre, I., Adams, I., Fox, A., Boonham, A., and Ravnikar, M. 2017. Next generation sequencing for detection and discovery of plant viruses and viroids: Comparison of two approaches. *Front. Microbiol.* 8:1-10.
- Poojari, S., Alabi, O. J., Fofanov, V. Y., and Naidu, R. A. 2013. A leafhopper-transmissible DNA virus with novel evolutionary lineage in the family Geminiviridae implicated in grapevine redleaf disease by next generation sequencing. *PLoS One* 8:e64194.
- Poudel, B., Sabanadzovic, S., Bujarski, J., and Tzanetakis, I. E. 2012. Population structure of *Blackberry yellow vein associated virus*, an emerging crinivirus. *Virus Res.* 169:272-275.
- Poudel, B., and Tzanetakis, I. E. 2013. Population structure of blackberry chlorotic ringspot virus. *Arch. Virol.* 158:667-672.
- Reuter, J. A., Spacek, D. V., and Snyder, M. P. 2015. High-throughput sequencing technologies. *Mol. Cell* 58:586-597.
- Rhoads, A., and Au, K. F. 2015. PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics* 13:278-289.
- Robinson, T. S., Scherm, H., Brannen, P. M., Allen, R., and Deom, C. M. 2016. *Blueberry necrotic ring blotch virus* in southern highbush blueberry: Insights into in planta and in-field movement. *Plant Dis.* 100:1575-1579.
- Roossinck, M. J. 2017. Deep sequencing for discovery and evolutionary analysis of plant viruses. *Virus Res.* 239:82-86.
- Roossinck, M. J., Martin, D. P., and Roumagnac, P. 2015. Plant virus meta-genomics: Advances in virus discovery. *Phytopathology* 105:716-727.
- Rosario, K., Padilla-Rodriguez, M., Kraberger, S., Stainton, D., Martin, D. P., Breitbart, M., and Varsani, A. 2013. Discovery of a novel mastrevirus and alpha satellite-like circular DNA in dragonflies (Ephemeroptera) from Puerto Rico. *Virus Res.* 171:231-237.
- Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., Leamon, J. H., Johnson, K., Milgrew, M. J., and Edwards, M. 2011. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475:348-352.
- Rothberg, J. M., and Leamon, J. H. 2008. The development and impact of 454 sequencing. *Nat. Biotechnol.* 26:1117-1124.
- Rott, M. E., Kesanakurti, P., Berwarth, C., Rast, H., Boyes, I., Phelan, J., and Jelkmann, W. 2018. Discovery of negative-sense RNA viruses in trees infected with apple rubbery wood disease by next-generation sequencing. *Plant Dis.* 102:1254-1263.
- Rott, M., Xiang, Y., Boyes, I., Belton, M., Saeed, H., Kesanakurti, P., Hayes, S., Lawrence, T., Birch, C., Bhagwat, B., and Rasi, H. 2017. Application of next generation sequencing for diagnostic testing of fruit tree viruses and viroids. *Plant Dis.* 101:1489-1499.
- Rowhani, A., Uyemoto, J. K., Golino, D. A., and Martelli, G. P. 2005. Pathogen testing and certification of *Vitis* and *Prunus* species. *Annu. Rev. Phytopathol.* 43:6.1-6.18.
- Roy, A., Choudhary, N., Guillermo, L. M., Shao, J., Govindarajulu, A., Achor, D., Wei, G., Picton, D. D., Levy, L., Nakhla, M. K., Hartung, J. S., and Brlansky, R. H. 2013. A novel virus of the genus *Cilevirus* causing symptoms similar to citrus leprosis. *Phytopathology* 103:488-500.
- Salmela, L., and Rivals, E. 2014. LoRDEC: Accurate and efficient long read error correction. *Bioinformatics* 30:3506-3514.
- Sanger, F., Nicklen, S., and Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463-5467.
- Schumann, G. L. 1991. *Plant Diseases: Their Biology and Social Impact*. American Phytopathological Society, St. Paul, MN.
- Seguin, J., Otten, P., Baerlocher, L., Farinelli, L., and Pooggin, M. M. 2014a. MISIS: A bioinformatics tool to view and analyze maps of small RNAs derived from viruses and genomic loci generating multiple small RNAs. *J. Virol. Methods* 195:120-122.
- Seguin, J., Otten, P., Baerlocher, L., Farinelli, L., and Pooggin, M. M. 2016. MISIS-2: A bioinformatics tool for in-depth analysis of small RNAs and

- representation of consensus master genome in viral quasispecies. *J. Virol. Methods* 233:37-40.
- Seguin, J., Rajeshwaran, R., Malpica-Lopez, N., Martin, R. R., Kasschau, K., Dolja, V. V., Otten, P., Farinelli, L., and Pooggin, M. M. 2014b. De novo reconstruction of consensus master genomes of plant RNA and DNA viruses from siRNAs. *PLoS One* 9:e88513.
- Shahid, M. S., Aboghanem-Sabanadzovic, N., Sabanadzovic, S., and Tzanetakis, I. E. 2017. Genomic characterization and population structure of a badnavirus infecting blackberry. *Plant Dis.* 101:110-115.
- Simmons, H. E., Dunham, J. P., Stack, J. C., Dickens, B. J. A., Pagan, I., Holmes, E. C., and Stephenson, A. G. 2012. Deep sequencing reveals persistence of intra- and inter-host genetic diversity in natural and greenhouse populations of zucchini yellow mosaic virus. *J. Gen. Virol.* 93:1831-1840.
- Singerman, A., and Useche, P. 2016. Impact of citrus greening on citrus operations in Florida. University of Florida, IFAS Ext. Bull. FE983.
- Skewes-Cox, P., Sharpton, T. J., Pollard, K. S., and DeRisi, J. L. 2014. Profile hidden Markov models for the detection of viruses within metagenomic sequence data. *PLoS One* 9:e105067.
- Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connel, C. R., Heiner, C., Kent, S. B., and Hood, L. E. 1986. Fluorescence detection in automated DNA sequence analysis. *Nature* 321:674-679.
- Sobhy, H. 2016. A review of functional motifs utilized by viruses. *Proteomes* 4:3.
- Stobbe, A. H., Daniels, J., Espindola, A. S., Verma, R., Melcher, U., Ochoa-Corona, F., Garzon, C., Fletcher, J., and Schneider, W. 2013. E-probe diagnostic nucleic acid analysis (EDNA): A theoretical approach for handling of next generation sequencing data for diagnostics. *J. Microbiol. Methods* 94:356-366.
- Stobbe, A. H., Schneider, W. L., Hoyt, P. R., and Melcher, U. 2014. Screening metagenomic data for viruses using the e-probe diagnostic nucleic acid assay. *Phytopathology* 104:1125-1129.
- Sudarshana, M. R., Perry, K. L., and Fuchs, M. F. 2015. Grapevine red blotch-associated virus, an emerging threat to the grapevine industry. *Phytopathology* 105:1026-1032.
- Valouev, A., Ichikawa, J., Tonthat, T., Stuart, J., Ranade, S., Peckham, H., Zeng, K., Malek, J. A., Costa, G., McKernan, K., Sidow, A., Fire, A., and Johnson, S. M. 2008. A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence dictated positioning. *Genome Res.* 18:1051-1063.
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., and Mural, R. J. 2001. The sequence of the human genome. *Science* 291:1304-1351.
- Villamor, D. E. V., Mekuria, T. A., Pillai, S. S., and Eastwell, K. C. 2016. High-throughput sequencing identifies novel viruses in nectarine: Insights to the etiology of stem-pitting disease. *Phytopathology* 106:519-527.
- Villamor, D. E. V., Pillai, S. S., and Eastwell, K. C. 2017. High throughput sequencing reveals a novel fabavirus infecting sweet cherry. *Arch. Virol.* 162:811-816.
- Visser, M., Bester, R., Burger, J. T., and Maree, H. J. 2016a. Next-generation sequencing for virus detection: Covering all the bases. *Virol. J.* 13:85.
- Visser, M. R., Burger, J. T., and Maree, H. J. 2016b. Targeted virus detection in next-generation sequencing data using an automated e-probe based approach. *Virology* 495:122-128.
- Vives, M. C., Velazquez, K., Pina, J. A., Moreno, P., Guerri, J., and Navarro, L. 2013. Identification of a new enamovirus associated with citrus vein enation disease by deep sequencing of small RNAs. *Phytopathology* 103:1077-1086.
- Wang, Q., Jia, P., and Zhao, Z. 2013. VirusFinder: Software for efficient and accurate detection of viruses and their integration sites in host genomes through next generation sequencing data. *PLoS One* 8:e64465.
- Wang, Q., Jia, P., and Zhao, Z. 2015a. VERSE: A novel approach to detect virus integration in host genomes through reference genome customization. *Genome Med.* 7:2.
- Wang, Y., Yang, Q., and Wang, Z. 2015b. The evolution of nanopore sequencing. *Front. Genet.* 5:1-20.
- Watson, M., Schnettler, E., and Kohl, A. 2013. ViRome: An R package for the visualization and analysis of viral small RNA sequence datasets. *Bioinformatics* 29:1902-1903.
- Welliver, R. 2012. Plum pox virus case study: The eradication road is paved in gold. (Abstr.) *Phytopathology* 102(suppl.):S4.154.
- Wren, J. D., Roossinck, M. J., Nelson, R. S., Scheets, K., Palmer, M. W., and Melcher, U. 2006. Plant virus biodiversity and ecology. *PLoS Biol.* 4:e80.
- Wu, Q., Ding, S.-H., Zhang, Y., and Zhu, S. 2015. Identification of viruses and viroids by next-generation sequencing and homology-dependent and homology-independent algorithms. *Annu. Rev. Phytopathol.* 53:425-444.
- Yamashita, A., Sekizuka, T., and Kuroda, M. 2016. VirusTAP: Viral genome-targeted assembly pipeline. *Front. Microbiol.* 7:32.
- Yepes, L. M., Cieniewicz, E. J., Krenz, B., McLane, H., Thompson, J. R., Perry, K. L., and Fuchs, M. 2018. Causative role of grapevine red blotch virus in red blotch disease. *Phytopathology* 108:902-909.
- Zerbino, D. R., and Birney, E. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18:821-829.
- Zhang, Y., Singh, K., Kaur, R., and Qiu, W. 2011. Association of a novel DNA virus with the grapevine vein-clearing and vine decline syndrome. *Phytopathology* 101:1081-1090.
- Zhao, G., Krishnamurthy, S., Cai, Z., Popov, V. L., Travassos da Rosa, A. P., Guzman, H., Cao, S., Virgin, H. W., Tesh, R. B., and Wang, D. 2013. Identification of novel viruses using VirusHunter—an automated data analysis pipeline. *PLoS One* 8:e78470.
- Zhao, G., Wu, G., Lim, E. S., Droit, L., Krishnamurthy, S., Barouch, D. S., Virgin, H. W., and Wang, D. 2017. VirusSeeker, a computational pipeline for virus discovery and virome composition analysis. *Virology* 503:21-30.
- Zheng, Y., Gao, S., Padmanabhan, C., Li, R., Galvez, M., Gutierrez, D., Fuentes, S., Ling, K.-S., Kreuze, J., and Fei, Z. 2017. VirusDetect: An automated pipeline for efficient virus discovery using deep sequencing of small RNAs. *Virology* 500:130-138.