Identification of Viruses and Viroids by Next-Generation Sequencing and Homology-Dependent and Homology-Independent Algorithms

Qingfa Wu,[1] Shou-Wei Ding,[2] Yongjiang Zhang,[3] and Shuifang Zhu[3]

[1]School of Life Sciences, University of Science and Technology of China, Hefei, Anhui, 230026 China; email: wuqf@ustc.edu.cn

[2]Department of Plant Pathology and Microbiology, University of California, Riverside, California 92521; email: dingsw@ucr.edu

[3]Institute of Plant Quarantine, Chinese Academy of Inspection and Quarantine, Chaoyang District, Beijing, 100123 China; email: zhusf@caiq.gov.cn

Annu. Rev. Phytopathol. 2015. 53:425–44

First published online as a Review in Advance on May 29, 2015

The *Annual Review of Phytopathology* is online at phyto.annualreviews.org

This article's doi: 10.1146/annurev-phyto-080614-120030

Keywords

pathogen discovery, viral metagenomics, vdSAR, PFOR

Abstract

A fast, accurate, and full indexing of viruses and viroids in a sample for the inspection and quarantine services and disease management is desirable but was unrealistic until recently. This article reviews the rapid and exciting recent progress in the use of next-generation sequencing (NGS) technologies for the identification of viruses and viroids in plants. A total of four viroids/viroid-like RNAs and 49 new plant RNA and DNA viruses from 18 known or unassigned virus families have been identified from plants since 2009. A comparison of enrichment strategies reveals that full indexing of RNA and DNA viruses as well as viroids in a plant sample at single-nucleotide resolution is made possible by one NGS run of total small RNAs, followed by data mining with homology-dependent and homology-independent computational algorithms. Major challenges in the application of NGS technologies to pathogen discovery are discussed.

INTRODUCTION

Viruses are ubiquitous pathogens found in all types of cellular organisms. Typically, viral particles or virions range from 30 to 450 nm in size and contain a small genome packaged by capsid proteins with or without a lipid envelope. Whereas cellular organisms store genetic information exclusively in DNA, both DNA and RNA can serve as the genomes of viruses. In addition to viruses, the circular noncoding RNA viroids also cause diseases in plants. Because of their small size and inability to propagate outside living host cells, viruses and viroids remain difficult to detect and identify compared to cellular pathogens. Enzyme-linked immunosorbent assay (ELISA), polymerase chain reaction (PCR), and nucleic acid hybridization techniques (including microarray) developed in the last several decades collectively provide rapid and inexpensive diagnoses for the known viruses and viroids and are widely used in agriculture and medicine (20, 64, 70, 83). However, because these assays depend on the reagents (antibodies, primers, or probes) developed from the characterized viruses and viroids, they are ineffective when the disease is caused by a new pathogen or a mixture of pathogens that share little or no sequence similarity with those described previously.

Major advances in DNA sequencing technology over the last decade have led to the development of new approaches for the identification and detection of viruses and viroids. These new approaches, frequently referred to as metagenomics (64), sequence the total nucleic acid content in disease samples by next-generation sequencing (NGS) technologies for the subsequent identification of pathogens by bioinformatics tools. Unlike existing methods, the metagenomics approach does not require prior knowledge of the pathogens and can potentially identify both the known and new viruses and viroids in a disease sample. For example, use of NGS technologies has played a key role in the identification of Israeli acute paralysis virus associated with colony collapse disorder in honeybees (16) and of a polyomavirus in human Merkel cell carcinoma (30). Rapid development in NGS technologies has dramatically reduced the cost and time for pathogen identification by metagenomics approaches, leading to a recent explosion in metagenomics studies on viruses and viroids in plants (46, 61, 84). After a short introduction of the main NGS platforms currently available, this article reviews recent applications of the homology-dependent and homology-independent metagenomics approaches to the identification and detection of viruses and viroids in plants.

NEXT-GENERATION SEQUENCING PLATFORMS

The meaning of the term high-throughput sequencing (HTS) has changed with the advancement of sequencing technologies. The Sanger technique had been the standard sequencing method since it was first described in 1977. For almost a decade, DNA sequencing was a labor-intensive process that involved determining only a few hundred nucleotides at a time (47). In a typical Sanger sequencing reaction, short DNA fragments made from supersonic or enzyme digestion have to be cloned into plasmids for amplification in *E. coli* before enzymatic dideoxy-sequencing. The sequencing protocol includes the use of radioisotopes and harsh chemicals. Development of capillary array electrophoresis and appropriate detection systems in the early 1990s (44) eventually led to the production of 96-channel capillary HTS sequencers (47). As an example of the modern Sanger capillary sequencing systems, the 3730xl DNA Analyzer from Applied Biosystems, Inc. (ABI) yields 96 or 384 sequences of approximately 600–1,000 nucleotides (nt) in length per run and a maximum of approximately 1.5 Mb of DNA sequence per day (**Table 1**).

New "massively parallel" DNA sequencing platforms have become available since 2005 to sequence populations of amplified template DNA molecules with a typical "wash-and-scan" technique (47). Because these new sequencing platforms outperform Sanger sequencing technologies

Table 1 Next-generation sequencing platforms

Platform	Principle[a]	Throughput/run	Read length	Time/run	Single-pass error rate
ABI 3730xl	Sanger Sequencing	64 kb	400–900 bp	1 h	0.1–1%
GS FLX+	SBS	700 Mb	700 bp	23 h	1%
HiSeq 2500	SBS	5–1,000 Gb	2 × 125 bp	1–6 days	0.1%
MiSeq	SBS	0.3–15 Gb	2 × 300 bp	5–55 h	0.1%
Ion Torrent	SBS	1.2–2 Gb	400 bp	7.3 h	1%
Ion Proton	SBS	10 Gb	200 bp	2–4 h	1%
SOLiD 5500xl	SBH	95 Gb	2 × 60 bp	6 days	5%
PacBio RSII	SMRT	500–1,000 Mb	10–15 kb	0.5–5 h	13%

[a]SBS, sequencing by synthesis; SBH, sequencing by hybridization; SMRT, single molecule, real time.

by a factor of 100–1,000 in daily throughput, they were termed next-generation sequencing (NGS) instead of HTS (47). These NGS approaches (**Table 1**) have different underlying biochemistries and differ in template preparation, sequencing and imaging, and data analysis (63). For sequencing by the 454 platform (e.g., GS FLX+; see **Table 1**), the DNA template is first amplified by emulsion PCR in which a single DNA template is amplified to thousands of copies in each droplet of an oil solution. The amplified DNA templates are incubated with a DNA polymerase, single-strand DNA binding proteins, and the ATP sulfurylases and luciferases; the light emitted during the incorporation is captured for all wells in parallel using a high-resolution charge-coupled device (CCD) camera. After capture of the light intensity, the remaining unincorporated nucleotides are washed away before the next nucleotide is added. The 454 technology is referred to as pyrosequencing because the sequencer generates visible light from inorganic pyrophosphate of deoxynucleotides (dNTPs) released during DNA synthesis by sulfurylase and luciferase. The average number of reads produced per run by the current GS FLX+ sequencer is more than 1 million reads of approximately 700 bases with a run time of approximately 23 h (47, 60).

The strategy of template preparation and sequencing of Illumina instruments is quite different from that of 454 pyrosequencing. Bridge amplification of each DNA template produces a cluster of thousands of original sequences in very close proximity to each other on the flow cell. A typical slide generates hundreds of millions of spatially separated clusters per run. The Illumina platform uses four dNTPs with different reversible dye terminators in the sequencing reaction so that the incorporation reaction is stopped after each base and the incorporated base is easily read out with fluorescent dyes. The fluorescently labeled terminator is imaged and then cleaved from the nascent end of DNA for the next cycle. The Illumina platform supports both single-end and mate-paired-end sequencing. Illumina HiSeq 2500 allows a powerful combination of 2 × 125 base pair (bp) read lengths and up to 1 Tetrabase per run within 6 days, whereas MiSeq generates up to 15 Gb of data in only 55 h (**Table 1**) (59, 63).

In the SOLiD platform of ABI the sequence extension reaction is carried out by ligases, rather than polymerases as in the 454 and Illumina platforms. The single-stranded copies of the DNA library molecules are first hybridized with a sequencing primer before the addition of a mixture of 8-mer probes carrying four distinct fluorescent labels to compete for ligation to the sequencing primer. After the fluorescence determined by the two 3′-most nucleotides of the probe is read out, three bases including the dye are cleaved from the 5′ end of the probe to leave a free 5′ phosphate for further ligation. After multiple cycles of ligations (typically up to 10 cycles), the synthesized strands are melted, and the ligation product is washed away before a new sequencing primer (shifted by one nucleotide) is annealed. The same process is repeated for the remaining three

primers, facilitating the readout of the dinucleotide encoding for each start position in the DNA sequence. Using specific fluorescent label encoding, the dye readouts (i.e., colors) are converted to a DNA sequence. For parallelization, the sequencing process uses beads covered with multiple copies of the sequence to be determined. These beads are created in a fashion similar to that described above for the 454 platform. The SOLiD 5500xl system currently generates 95 Gb of data per run during 6 days with a typical read length combination of 2 × 60 bp (47, 92).

New sequencing platforms with improved performance continue to be developed and released (**Table 1**). Approaches that sequence single large DNA molecules without the need to halt between read steps are sometimes referred as third-generation sequencing (88). However, Ion Torrent and Proton are between the second and third generation, since they do not completely fulfill the features assigned to either category (9). Advancement of sequencing technology provides unprecedented opportunity for pathogen discovery in plants by unbiased viral metagenomics.

HOMOLOGY-DEPENDENT PATHOGEN IDENTIFICATION BY NEXT-GENERATION SEQUENCING TECHNOLOGIES

Viral metagenomics is defined as the culture-independent study of the collective set of viral genomes in samples from the environment, as well as from individual plants and animals (74). Encouraged by early success in the discovery of animal viruses (16, 30, 68, 71), at least 66 studies on the metagenomics of plant viruses and viroids have been published since 2009 (**Table 2**). Typically (**Figure 1**), identification of viruses and viroids by the metagenomics approach includes sample preparation, DNA sequencing by an NGS platform, and bioinformatics analysis (114), which are discussed below.

Sample Preparation: Strategies to Enrich Virus Sequences

Six new plant viruses have been identified by directly sequencing the total RNA from host plants (1, 3, 67, 95, 103, 111) (**Table 2**). To maximize the detection of viruses, however, several strategies have been developed for enrichment of the deep-sequencing reads specific to viruses and/or viroids. Because of the overwhelming abundance of host rRNAs, rRNA depletion from total RNA preparations will increase the proportion of virus and viroid reads relative to those of host and nonhost origin (3). Similarly, rRNA depletion is also achieved by sequencing only the polyadenylated transcripts (poly-A RNAs), and enrichment of poly-A RNAs for deep sequencing led to the discovery of seven new RNA viruses and two DNA badnaviruses (34, 107, 113). However, this enrichment strategy will also delete reads from many families of RNA viruses that do not synthesize poly-A RNAs, including, for example, *Tobacco mosaic virus*, *Cucumber mosaic virus*, and *Turnip yellow mosaic virus*. Thus, enrichment for the following molecules has been more common in metagenomics studies of plant viruses.

Double-stranded RNAs. Because double-stranded RNA (dsRNA) is synthesized by RNA viruses and viroids as replicative intermediates (**Figure 1**) and plants do not normally produce dsRNA, sequencing total dsRNA dramatically increases the proportion of reads specific to viruses and viroids (2–4, 13, 15, 76). For example, a study compared deep sequencing of total RNA and dsRNA from the same plant samples and found that virus reads increased from 2% to 53% after dsRNA enrichment (3). Use of CF11 cellulose spin-column or chromatography for the purification of viral dsRNA is a well-established procedure in plant virology (26, 84, 85). Recently, an improved dsRNA purification method based on anion exchange chromatography utilizing convective interaction media (CIM) monolithic columns has been developed, which can be used to obtain highly pure

Table 2 Detection of plant viruses and viroids by next-generation sequencing technologies

Enrichment	NGS platform	Assembly tools	Viruses			Viroids		Host species	Reference
			Known	New		Known	New		
				RNA	DNA				
Total RNA	454	CLC	3	1		3		Grapevine	3
	454	CLC	1	1				Tomato/*Gomphrena globosa*	1
	454	CLC		1				Cassava	67
	454	CLC	1					Canna	66
	Illumina	Geneious	1					*Hardenbergia comptoniana*	104
	Illumina	Velvet/Geneious		1				*H. comptoniana*	103
	454	CLC		1				Lettuce	95
	454	CLC	1					Iris	43
	Illumina	Geneious/CLC		1				Yellow tailflower	111
	Illumina	Geneious	1					*Capsicum annuum*	99
Poly-A	Illumina	Geneious	1					*Passiflora caerula*	105
	Illumina	Geneious/CLC		1				Orchid	113
	Illumina	Geneious/CLC	1					Iris	44
	Illumina	Geneious/CLC	5					*Narcissus* and *Hippeastrum*	106
	Illumina	Geneious/CLC	4	4				Orchid	107
	Illumina	Edena/Velvet/ SOAPdenovo/ CAP3	11	2	2			Sweet potato	34
	Illumina	Velvet, Oases	1					*Prunus domestica*	80
rRNA depletion	Illumina		3			4		Grapevine	75
dsRNA	454	BLAST	3	1		3		Grapevine	3
	Illumina	Velvet	4					Grapevine	15
	454	Newbler	1					Grapevine	4
	Illumina	Velvet/ CodonCode		1				Red raspberry	76
	454	CLC	1					Cherry	13
	Illumina	Velvet		1				Grapevine	6
	454	Newbler	2					Maize	2v
	Illumina	Geneious/CLC		1				Orchid	110
	Illumina	Velvet			1			Grapevine	5
	Illumina	Velvet		2				Japanese persimmon	40
	Illumina	Velvet					PVd2	*Diospyros virginiana*	41
	Illumina	CLC	1					Citrus	115
	Illumina	Geneious/CLC	10					Garlic	112

(*Continued*)

Table 2 (*Continued*)

Enrichment	NGS platform	Assembly tools	Viruses			Viroids		Host species	Reference
			Known	New		Known	New		
				RNA	DNA				
Virus-like particles	454	Newbler	1					Sweet potato	81
	SOLiD	Velvet		1				Eggplant	28
	454	Newbler	1					*P. domestica*	93
	454	Newbler/CLC	1					Potato	79
	SOLiD	Velvet		1				Pepper	27
	454	CLC	1		1			Sugarcane	12
	454	Newbler			3			Black pepper	36
Small RNAs	Illumina	Velvet/VCAKE	2		3			Sweet potato	49
	454	Velvet	2					Cocksfoot grass	72
	Illumina	Blast	12					Grapevine	73
	Illumina	SSAKE	27		1			Grapevine	117
	Illumina	Velvet	1					Sweet potato	17
	Illumina	Velvet	2					Tomato	35
	Illumina	Velvet	1					Squash	35
	Illumina	Velvet	6					Sweet potato	45
	Illumina	Velvet	1	1		1		Tomato	54
	Illumina	Velvet	3	1		2		Grapevine	32
	Illumina	Velvet		1		2		Citrus	56
	Illumina	Velvet			1			Citrus	57
	Illumina	Velvet		1				Yam bean	31
	Illumina	Velvet	10	1				*Arctium tomentosum*	10
	Illumina	Velvet	1					Soybean	65
	Illumina	Velvet		1				Sweet potato	19
	Illumina	Velvet	1	1				Potato	48
	Illumina	Velvet		1				Tomato	53
	Illumina	Velvet/Oases		1				Citrus	86
	Illumina	Velvet/Oases		1				Citrus	97
	Illumina	CLC	1		1			Sugarcane	12
	Illumina	CodonCode	1					Papaya	118
	Illumina	Geneious/Velvet/CLC/Geneious	3	1				Tomato	59
	Illumina	Velvet	19					Sweet potato	62
	Illumina	Velvet	1					Potato	50
	Illumina	Velvet			1	2		Grapevine	91
	Illumina	Velvet	1	3				Cassava	14
	Illumina	Velvet			1			Pagoda	98
	Illumina	Velvet/CAP3	3	1				Rose	39
	Illumina	PFOR				2	1*	Grapevine	102
	Illumina	PFOR2				5	2*	Apple/grapevine	119

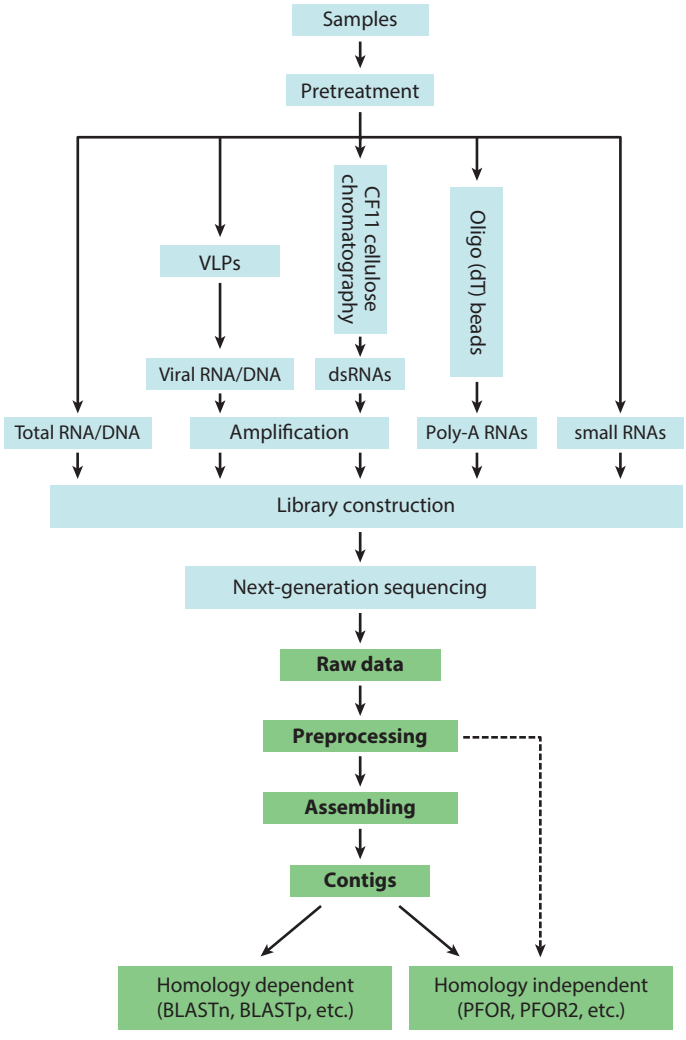

Figure 1

The key steps in the discovery of viruses and viroids by next-generation sequencing. The experimental part, including the strategies of viral pathogen enrichment, is shown in blue, whereas the informatics process is shown in green. Abbreviations: BLASTn, nucleotide-nucleotide BLAST; BLASTp, protein-protein BLAST; dsRNAs, double-stranded RNAs; dT, thymine deoxyribonucleotide; PFOR, progressive filtering of overlapping small RNAs; VLP, virus-like particle.

dsRNA preparations shorter than 1 kb (82). However, six of the seven new viruses discovered by dsRNA sequencing contain RNA genomes, possibly because plant DNA viruses do not produce sufficiently long dsRNA in their life cycle. We note that the new DNA geminivirus was identified from sequencing a total dsRNA preparation that was not treated with RNase and DNase before library construction (5).

Virus-like particles. Viral genomic RNA or DNA packaged in viral particles is protected from DNase and RNase treatments. Therefore, enrichment of VLPs (**Figure 1**) by homogenization, filtration, and ultracentrifugation has been widely used for virus discovery in ocean, environmental,

and fecal samples (20, 52, 64, 96) as well as in several studies of plant samples (12, 27, 28, 36, 79, 81, 93). VLP preparations contain contaminating mitochondria and bacteria, so VLPs are often treated with chloroform to disrupt mitochondrial and bacterial membranes before nuclease digestion and the extraction of VLP-associated nucleic acids for deep sequencing (51, 100). Unfortunately, enveloped viruses are also sensitive to chloroform treatment. Moreover, because successful virion purification of many viruses requires development of specific protocols, it is unlikely that all viruses can be captured by a single protocol for VLP enrichment. Nevertheless, deep sequencing of the total nucleic acids from VLPs extracted from plants led to the discovery of four DNA viruses and two RNA viruses (**Table 2**).

Because the total nucleic acid content from both VLP and dsRNA preparations is often very low, an extra step to amplify the extracted nucleic acids by PCR or reverse transcription PCR (RT-PCR) in a sequence-independent manner is necessary before the construction of libraries for deep sequencing (**Figure 1**) (16, 64). An improved rolling-circle amplification technique to amplify circular dsDNA viral genomes is available (89, 94).

Small RNAs. It is known that plants and invertebrates produce virus-derived small interfering RNAs (siRNAs) in response to infection by both RNA and DNA viruses (**Figure 1**) (22–24). Recent studies have also detected abundant production of viral siRNAs in mammalian cells after infection by two unrelated positive-strand RNA viruses (55, 58). Moreover, dsRNA replicative intermediates of viroids and satellite RNAs are also processed into siRNAs in plants (22–24). The first deep-sequencing study of viral siRNAs has revealed that viral siRNAs in fact overlap each other extensively despite being only 21 to 24 nt long (7). Ubiquitous production in diverse eukaryotic hosts and the overlapping property of viral siRNAs have been utilized independently by two groups to develop a novel strategy for virus discovery (**Figure 2**) by enriching and sequencing small RNAs (49, 101). In this approach, known as virus discovery by deep sequencing and assembly of total small RNAs (vdSAR), small RNAs are enriched from diseased cells/tissues for deep sequencing by NGS platforms and assembled into large sequence contigs/fragments that are then used for virus discovery as those sequences are obtained from dsRNAs and VLPs.

The original protocols to construct a library of small RNAs for sequencing took up to two weeks (7, 8). However, current protocols no longer require gel purification of small RNAs either before or after ligation with 5′ and 3′ adaptors and may be completed within one day (55). Because the sequence-independent amplification required for dsRNA and VLP preparations is not necessary, sample preparation for small RNA sequencing is less technically demanding and time-consuming than are dsRNA and VLP purification protocols (**Figure 1**).

There are several reasons why more investigators chose to enrich and sequence small RNAs for the identification of plant viruses and viroids over other enrichment strategies (**Table 2**). First, replication of RNA and DNA viruses and replication of subviral agents such as viroids and satellite RNAs in plants all induce extremely abundant accumulation of the pathogen-specific siRNAs, which represent up to 30% of total small RNAs sequenced from diseased plants (22–24). Because both the amount of sequencing and data complexity are greatly reduced, sequencing of multiple samples marked by barcodes in a single lane still provides sufficient depth for pathogen discovery. Thus, vdSAR is cost-effective. Second, unlike dsRNA and VLP preparations, all replicating viruses and viroids in a diseased plant can be detected in principle from the deep sequencing of a single library of small RNAs. This feature is particularly attractive for many disease diagnostic and quarantine applications. Third, viral and subviral siRNAs are the products of an active host immune response to infection and exhibit specific patterns of size distribution in distinct host species due to their biogenesis by specific Dicer proteins (22–24). Therefore, the size distribution pattern of small RNAs may reveal if the identified viral and subviral pathogen actively replicates in plants or

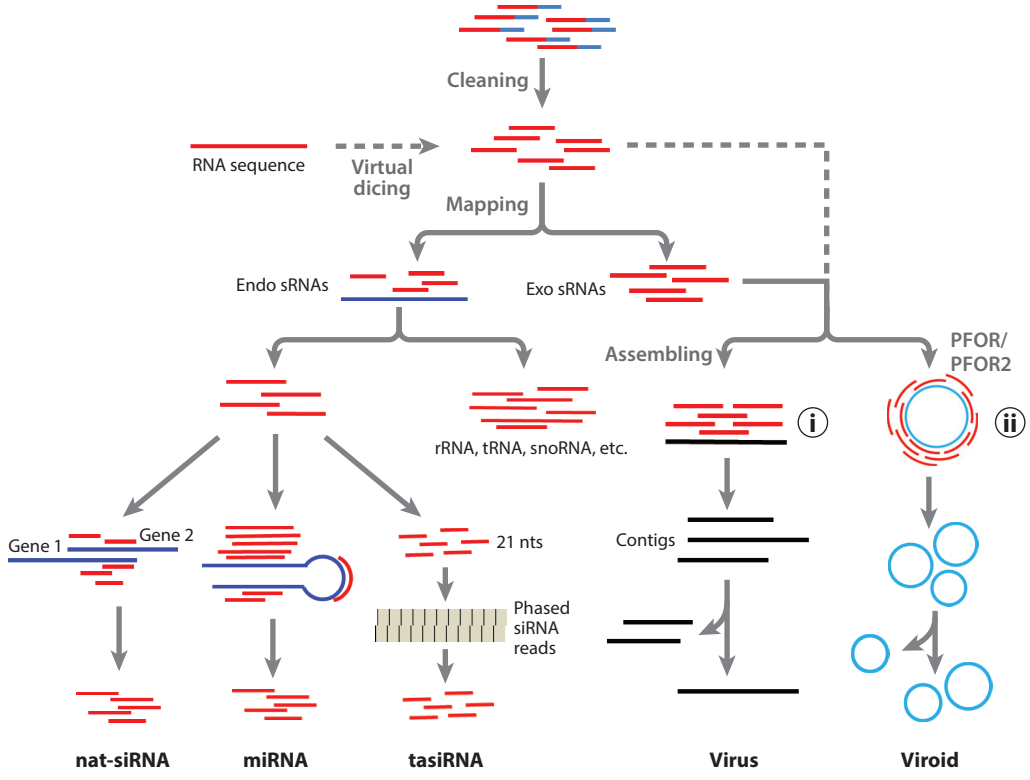

Figure 2

The classification principles of endogenous and exogenous small RNAs (sRNAs) sequenced from plants in the discovery of (*i*) viruses by deep sequencing and assembly of total small RNAs (vdSAR) and (*ii*) viroids by progressive filtering of overlapping small RNAs (PFOR). Abbreviations: nat-siRNAs, natural antisense siRNAs; nt, nucleotides; tasiRNAs, trans-acting siRNAs; miRNAs, microRNAs; SLS, splitting longer reads into shorter fragments.

in the associated fungal or insect species (119). Finally, data mining of the same libraries of small RNAs may lead to the discovery of novel pathogens that exhibit no sequence similarity detectable by currently available informatics tools (**Figures 1** and **2**; see also below).

Bioinformatics Analysis

Prior to sequence assembly and pathogen identification, the raw data generated by NGS platforms must be preprocessed to remove adaptors and low-quality sequences (**Figure 1**). The quality control is dependent on the sequencing technology used. Standard parameters and thresholds are usually provided by the manufacturer. It is now a standardized process on "older" technologies like 454/Roche pyrosequencing or Illumina sequencing by synthesis. For multiplex sequencing of mixed libraries in a single lane, an extra step of demultiplexing using barcodes built in the PCR primers is necessary before sequence assembly. When the genome sequence of the host plant is available, in silico subtraction of host-specific sequences (**Figure 2**) before assembly will speed up the downstream bioinformatics analysis (54, 64). It should be pointed out that adaptor removal and computational filtering of host sequences can lead to other artifacts (43, 84).

The assembly of the preprocessed reads can be executed by several mainstream algorithms that are publicly available, including Velvet, Oases, and VCAKE (42, 90, 116). The parameters used

in sequence assembly are similar to those for genome assembly and are defined by the algorithm employed. Subsequently, the assembled contig sequences are queried by homology search tools against previously documented sequences stored either in a local database or in public databases such as GenBank (64). A common approach is to compare the assembled sequences with the nonredundant nucleotide database of GenBank using a BLAST package. BLASTn and BLASTx are the two frequently used programs for nucleotide and amino acid comparison, respectively. New and known viruses are readily identified when contigs show high similarity (>90% similarity and 85% coverage) with a known virus (78, 101). When a contig shows distant homology with a known virus, especially only at the protein level, the contig often represents a new virus that can be taxonomically assigned only at the level of the virus family (78, 101).

A bottleneck in the viral metagenomics approach is the de novo annotation of thousands of the assembled contigs. The computing time required for the annotation is likely to increase as sequence databases continue their exponential growth. The use of dedicated databases or of subsections of GenBank could overcome this problem; however, it will not be possible to identify bacterial, archaeal, and nonhost eukaryotic sequences, leading to an overestimation of the fraction of unknowns (64). To meet the demand of the huge number of sequencing data from NGS, computer programs that are orders of magnitude faster than BLAST have been developed, such as USEARCH and HHbits (29, 77).

Although software packages used for viral metagenomics are freely accessible, data analysis often requires a trained bioinformaticist, a requirement that hinders the routine application of NGS in the identification of viruses and viroids. Recently, an easy-to-use graphical interface, SearchSmallRNA, became available to reconstruct viral genomes with high reliability using small RNA data from NGS. SearchSmallRNA bypasses the need of line command and basic bioinformatics knowledge (18). Commercial packages such as CLC Genomics Workbench and Geneious or open platforms such as Galaxy also provide user-friendly interfaces and simplify the use and parameterization of these tools. These efforts will facilitate identification of viruses and viroids by NGS technologies.

Homology-Dependent Discovery of Viruses and Viroids in Plants

Diverse eudicot and monocot plant species have been selected for the identification of viruses and viroids by metagenomics approaches (**Figure 3**). These efforts have resulted in the discovery of at least 49 new RNA and DNA viruses in total. Among these new viruses, 36 can be classified into the existing genera of 16 different families, nine may become the founding members of new genera in five different families, and the remaining four viruses cannot be assigned to a known virus family (**Table 3**). The comparison shown in **Table 2** shows that enrichment of small RNAs for deep sequencing is the most effective strategy in terms of both the number and diversity of the new viruses identified. Both RNA and DNA viruses are readily detected by enriching VLPs for sequencing. However, viroids are not found in VLPs as expected. Detection of the known viroids is possible by sequencing small RNAs and total RNAs with or without rRNA depletion (**Table 2**). A new viroid was identified by total dsRNA sequencing (41). In contrast, new viroids have not been sequenced from many plant samples with the available bioinformatics tools. We note that a majority of investigators chose Illumina as the NGS platform for viral metagenomics studies in plants.

HOMOLOGY-INDEPENDENT IDENTIFICATION OF VIROIDS

Although prior knowledge of the pathogens is not necessary to initiate the diagnostic procedure by the metagenomics approach, a pathogen is identified only when it exhibits nucleotide or amino acid

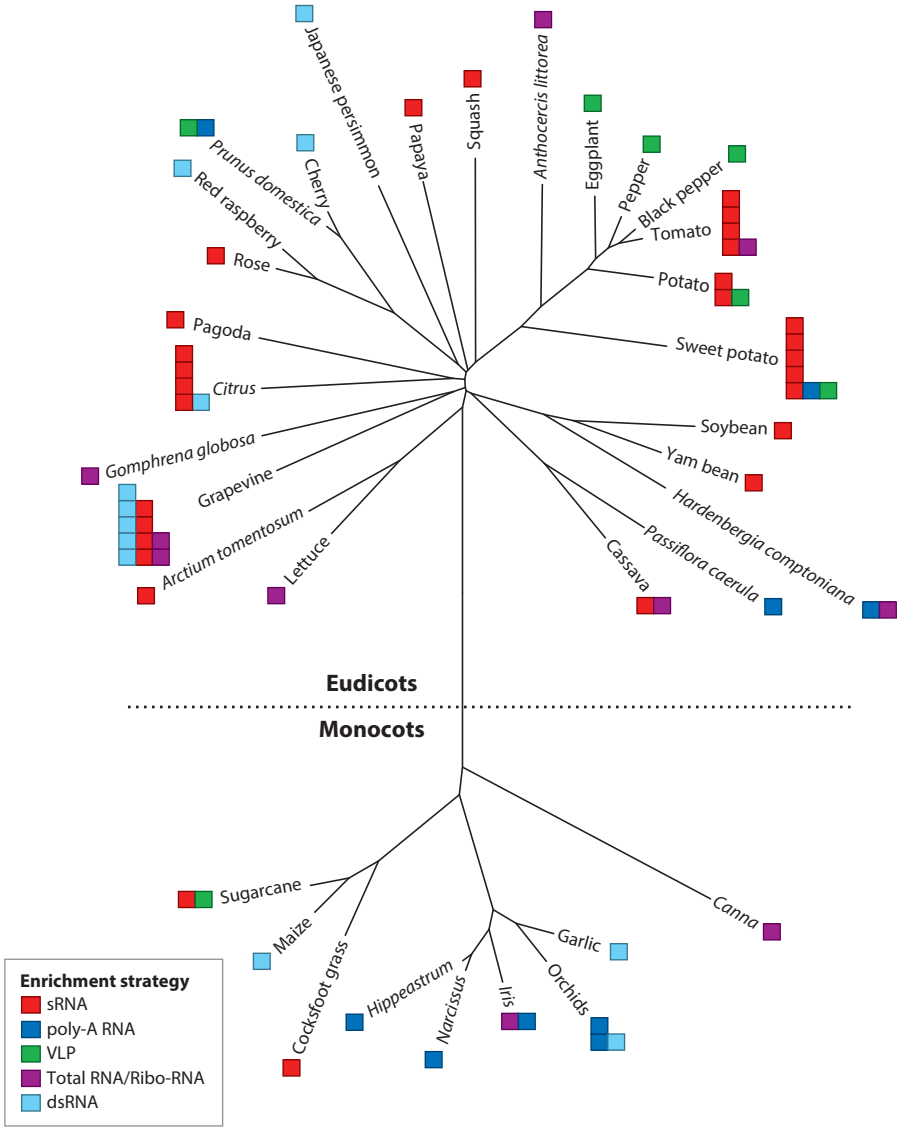

Figure 3

The taxonomy of host plant species selected for the identification of viruses and viroids by next-generation sequencing. Also shown are the strategies (*frequencies are marked by colored boxes*) used to enrich virus and/or viroid sequences in each host species. The clustering of host species on a tree may not be strictly consistent with the phylogenetic relationship of these species. Abbreviations: dsRNA, double-stranded RNA; poly-A RNA, polyadenylated transcript; ribo-RNA, ribosomal RNA-depleted RNA; sRNA, small RNA; VLPs, virus-like particles.

sequence homology with a known pathogen in a database. Therefore, the metagenomics approach as described above cannot discover novel pathogens that show little or no sequence homology with known pathogens. In a hypothetical scenario, for example, BLAST searches would not be able to identify NGS reads of geminiviruses as virus-specific in 1977 (33, 38) because geminiviral genomes show no detectable sequence similarity with other viruses (37).

Table 3 New plant viruses and viroids discovered by next-generation sequencing

Virus	Family (genus)	Infective?	Reference
RNA viruses			
Gayfeather mild mottle virus (GMMV)	*Bromoviridae* (*Cucumovirus*)	Yes	1
Grapevine Syrah virus-1 (GSyV-1)	*Tymoviridae* (*Marafivirus*)	NA	3
Cassava brown streak virus (CBSV)	*Potyviridae* (*Ipomovirus*)	NA	67
Hardenbergia virus (HarVA)	*Betaflexiviridae*	NA	103
Lettuce necrotic leaf curl virus (LNLCV)	*Secoviridae* (*Torradovirus*)	Yes	95
Yellow tailflower mild mottle virus (YTMMV)	*Virgaviridae* (*Tobamovirus*)	Yes	111
Caladenia virus A (CalVA)	*Potyviridae* (*Poacevirus*)	NA	113
Donkey orchid virus A (DOVA) Diuris virus A (DiVA) Diuris virus B (DiVB) Diuris pendunculata cryptic virus (DPCV)	*Potyviridae* (*Potyvirus*) *Betaflexiviridae* (*Capillovirus*) *Betaflexiviridae* (*Capillovirus*) *Partitiviridae*	NA	107
Ipomoea batatas Rhabdovirus N-like sequences (IbRNLS) Cymbidium mosaic virus (CymMV)	*Rhabdoviridae* *Alphaflexiviridae* (*Potexvirus*)	NA	34
Raspberry latent virus (RpLV)	*Reoviridae* (*Reovirus*)	Yes	76
Grapevine virus F (GVF)	*Betaflexiviridae* (*Vitivirus*)	Yes	6
Donkey orchid symptomless virus (DOSV)	Unassigned	Yes	110
Persimmon virus A (PeVA) Persimmon latent virus (PeLV)	*Rhabdoviridae* (*Cytorhabdovirus*) Unassigned	Yes	40
Eggplant mild leaf mottle virus (EMLMV)	*Potyviridae*	Yes	28
Pepper yellow leaf curl virus (PYLCV)	*Luteoviridae* (*Polerovirus*)	Yes	27
Tomato necrotic stunt virus (ToNSV)	*Potyviridae* (*Potyvirus*)	Yes	54
Grapevine pinot gris virus (GPGV)	*Betaflexiviridae* (*Trichovirus*)	NA	32
Citrus yellow vein clearing virus Y1 (CYVCV-Y1)	*Alphaflexiviridae* (*Mandarivirus*)	Yes	56
Yam bean mosaic virus (YBMV)	*Potyviridae* (*Potyvirus*)	Yes	31
Woolly burdock yellow vein virus (WBYVV)	Unassigned	NA	10
Sweet potato C6 virus (SPC6V)	*Betaflexiviridae* (*Carlavirus*)	Yes	19
Andean potato mild mosaic virus (APMMV)	*Tymoviridae* (*Tymovirus*)	Yes	48
Tomato mottle mosaic virus (ToMMV)	*Virgaviridae* (*Tobamovirus*)	Yes	53
Citrus leprosis virus cytoplasmic type 2 (CiLV-C2)	*Cilevirus*	Yes	86
Citrus vein enation virus (CVEV)	*Luteoviridae* (*Enamovirus*)	Yes	97
Tomato matilda virus (TMaV)	*Iflaviridae* (*Tomavirus*)	Yes	59
Cassava polero-like virus (CsPLV) Cassava new alphaflexivirus (CsNAV) Cassava torrado-like virus (CsTLV)	*Luteoviridae* (*Polerovirus*) *Alphaflexiviridae* (*Potexvirus*) *Secoviridae* (*Torradovirus*)	Yes	14
Rose leaf rosette-associated virus (RLRaV)	*Closteroviridae* (*Closterovirus*)	NA	39
DNA viruses			
Sweet potato badnavirus C1 (SPBV-C1) Sweet potato badnavirus C2 (SPBV-C2)	*Caulimoviridae* (*Badnavirus*) *Caulimoviridae* (*Badnavirus*)	NA	34
Grapevine red blotch-associated virus (GRBaV)	*Geminiviridae*	Yes	5
Sugarcane white streak virus (SWSV)	*Geminiviridae* (*Mastrevirus*)	NA	12
Piper yellow mottle virus (PYMoV) Piper DNA virus 1 (PDV-1) Piper DNA virus 2 (PDV-2)	*Caulimoviridae* (*Badnavirus*) *Caulimoviridae* *Caulimoviridae*	Yes NA NA	36

(*Continued*)

Table 3 (*Continued*)

Virus	Family (genus)	Infective?	Reference
Two badnaviruses One mastrevirus	*Caulimoviridae* (*Badnavirus*) *Geminiviridae* (*Mastrevirus*)	NA	49
Grapevine vein clearing virus (GVCV)	*Caulimoviridae* (*Madnavirus*)	Yes	117
Citrus chlorotic dwarf-associated virus (CCDaV)	*Geminiviridae*	Yes	57
Grapevine geminivirus (GVGV)	*Geminiviridae*	NA	91
Pagoda yellow mosaic associated virus (PYMAV)	*Caulimoviridae* (*Badnavirus*)	NA	98
Viroids			
Persimmon viroid 2 (PVd2)	*Apscaviroid*	Yes	41
Grapevine hammerhead viroid-like RNA (GHVd RNA)	*Avsunviroidae*	NA	102
Apple hammerhead viroid-like RNA (AHVd-like RNA)	*Avsunviroidae*	NA	119
Grapevine latent viroid (GLVd)	*Pospiviroidae*	Yes	119

Abbreviation: NA, not available.

A novel homology-independent metagenomics approach has been recently developed for the discovery of viroids (102). This approach (**Figure 2**) involves analyzing the sequences of the total small RNAs of the infected plants obtained by NGS with a novel computational algorithm, progressive filtering of overlapping small RNAs (PFOR). Viroid infection triggers production of viroid-derived overlapping siRNAs that cover the entire genome at high densities (11, 21, 69). However, these viroid siRNAs cannot be assembled into complete viroid genomes by conventional genome assembly algorithms such as Velvet because of the high heterogeneity of viroid populations (102). PFOR retains viroid siRNAs for genome assembly by progressively eliminating small RNAs that do not overlap and those that overlap but cannot be assembled into a direct repeat RNA, which is synthesized during the rolling-circle replication of the viroid RNA. Viroids from both of the known viroid families are readily identified and their full-length sequences assembled by PFOR from small RNAs sequenced from infected plants. The recent algorithm update significantly enhances the performance by adopting parallel programming into the filtering step, which takes up more than 90% of the running time (119).

A new viroid, Grapevine latent viroid (GLVd), was identified by the homology-independent metagenomics approach (119). GLVd was proposed as a new species of *Apscaviroid* because it contains the typical structural elements found in this group and independently infects grapevine seedlings (119). PFOR/PFOR2 also enabled the discovery of two putative viroids designated grapevine hammerhead viroid-like RNA (GHVd RNA) and apple hammerhead viroid-like RNA (AHVd RNA), neither of which exhibits sequence similarity with any of the known molecules detectable by the available algorithms. Assembly of AHVd and GHVd RNA repeats by PFOR from siRNAs, with their 21-nt/22-nt ratio characteristic of the viral siRNAs targeting plant RNA viruses known to replicate in the cytoplasm, indicates that both circular RNAs replicate via the rolling-circle mechanism. Both AHVd and GHVd RNAs also encode a biologically active hammerhead ribozyme in each polarity that is structurally conserved in viroids from the *Avsunviroidae*. Moreover, AHVd RNA was not specifically associated with any of the viruses found in apple plants by deep sequencing. Although independent infectivity has yet to be demonstrated, the available lines of evidence suggest that AHVd and GHVd RNAs may represent novel viroids.

A simple computational program known as SLS (for splitting longer reads into shorter fragments) was recently developed as part of PFOR2 to discover biologically active circular RNAs via

the deep sequencing of long RNAs instead of small RNAs (119). The program (**Figure 2**) cuts the sequenced long RNAs into 21-nt virtual small RNAs of 20-nt overlap with their 5′ and 3′ neighboring small RNAs before PFOR2 analysis to retain only those that could be assembled into direct repeat RNAs. SLS-PFOR2 was validated by successful assembly of the full-length PSTVd genome from the total RNAs deep sequenced from the infected potato plants after rRNA depletion or RNase R treatment to enrich circular RNAs (119). The development of SLS-PFOR2 allows the discovery of viroids that infect a host species but do not trigger Dicer-dependent biogenesis of viroid siRNAs, which may lead to the identification of novel viroids.

CONCLUDING REMARKS

Determination of the total nucleic acid content in a biological sample by NGS technologies provides a powerful tool for plant pathologists in the diagnosis and identification of viruses and viroids. NGS technologies will also likely transform the inspection and quarantine services required to provide a fast, accurate, and full indexing of viruses and viroids in plant samples. In contrast to the current techniques such as ELISA, PCR, and microarray, the metagenomics approaches do not require prior knowledge of the pathogens. Successful detection of diverse viruses and viroids by the metagenomics approaches has been described in plants since 2009. Available data show that among the strategies developed to enrich NGS reads specific to viruses and/or viroids, targeting the total small RNAs for deep sequencing has been the most effective to identify RNA and DNA viruses as well as viroids (25). Analyses of plant samples by NGS and homology-dependent computational algorithms have identified two new viroids and 49 new viruses from 20 known or assigned virus families in the last six years. Notably, development of the algorithm PFOR allows the discovery of viroids independent of not only the tedious purification of the naked circular RNAs but also any sequence similarity to a known viroid. Despite widespread distribution in potato, horticultural species, and fruit trees, viroids have so far not been found in many major monocot or small fruit crops, or in animals. The recent algorithm update SLS-POFR2 further expands the range of host species for viroid discovery to include those that may not produce overlapping siRNAs targeting the replicating circular RNAs.

Many challenges remain in the application of NGS technologies to pathogen discovery. First, there is a critical need for the development of new computational algorithms capable of discovering novel viruses from NGS datasets in a homology-independent manner, as illustrated by PFOR in viroid discovery. These algorithms may also help identify the origins of the large volume of unknown sequences present in any deep-sequencing project that are not homologous to any entry in GenBank. Second, development of user-friendly software interfaces such as SearchSmallRNA that are accessible to the public and require little informatics training will facilitate more widespread applications of NGS technologies to the diagnosis and identification of viruses and viroids. Third, Koch's postulates have not been fulfilled for many viruses and viroids discovered by the metagenomics approaches. However, because production of virus- and viroid-specific siRNAs reveals induction of the host immune response to an active infection, identification of viruses and viroids by small RNA sequencing also provides evidence for their replication in the host.

SUMMARY POINTS

1. The power of next-generation sequencing (NGS) technologies to allow rapid determination of the total nucleic acid content in a biological sample has transformed the diagnosis and identification of viruses and viroids.

2. All viruses and viroids identical or similar to those described previously can be identified in a plant sample by a single NGS run and homology-dependent algorithms.
3. A total of 49 new viruses and one new viroid have been identified by NGS and homology-dependent algorithms since 2009. Twelve of these viruses may become the founding members of new virus genera or families.
4. Common strategies to enrich viruses and/or viroids for deep sequencing include the purification of dsRNAs, virus-like particles, or small RNAs. However, identification of RNA and DNA viruses as well as viroids in a single NGS run is possible only by sequencing total small RNAs.
5. The recent development of the homology-independent algorithm PFOR/PFOR2 makes it possible to discover novel viroids by deep sequencing of small RNAs and total RNAs either deleted of rRNA or enriched for circular RNAs. It is likely that PFOR will facilitate discovery of previously uncharacterized viroids in diverse plant and animal species in both agriculture and the environment.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors' research projects described in this review were supported by grants to Q.W. from the National Basic Research Program of China (No. 2014CB138405) and the Chinese National Natural Science Foundation (No. 31272011) and to S.W.D. from the US–Israel Binational Agricultural Research and Development Fund (BARD-IS-4513-2), the US–Israel Binational Science Foundation (BSF-2011302), and the US Department of Agriculture Research Service (6659-22000-025).

LITERATURE CITED

1. Adams IP, Glover RH, Monger WA, Mumford R, Jackeviciene E, et al. 2009. Next-generation sequencing and metagenomic analysis: a universal diagnostic tool in plant virology. *Mol. Plant Pathol.* 10:537–45
2. Adams IP, Miano DW, Kinyua ZM, Wangai A, Kimani E, et al. 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 Pathol.* 62:741–49
3. 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
4. Al Rwahnih M, Daubert S, Urbez-Torres JR, Cordero F, Rowhani A. 2011. Deep sequencing evidence from single grapevine plants reveals a virome dominated by mycoviruses. *Arch. Virol.* 156:397–403
5. Al Rwahnih M, Dave A, Anderson MM, Rowhani A, Uyemoto JK, Sudarshana MR. 2013. Association of a DNA virus with grapevines affected by red blotch disease in California. *Phytopathology* 103:1069–76
6. Al Rwahnih M, Sudarshana MR, Uyemoto JK, Rowhani A. 2012. Complete genome sequence of a novel vitivirus isolated from grapevine. *J. Virol.* 86:9545
7. Aliyari R, Wu Q, Li HW, Wang XH, Li F, et al. 2008. Mechanism of induction and suppression of antiviral immunity directed by virus-derived small RNAs in Drosophila. *Cell Host Microbe* 4:387–97

8. Ambros V, Lee RC. 2004. Identification of microRNAs and other tiny noncoding RNAs by cDNA cloning. *Methods Mol. Biol.* 265:131–58
9. Barzon L, Lavezzo E, Militello V, Toppo S, Palu G. 2011. Applications of next-generation sequencing technologies to diagnostic virology. *Int. J. Mol. Sci.* 12:7861–84
10. Bi YQ, Tugume AK, Valkonen JPT. 2012. Small-RNA deep sequencing reveals *Arctium tomentosum* as a natural host of *Alstroemeria virus X* and a new putative Emaravirus. *PLOS ONE* 7:e427587
11. Bolduc F, Hoareau C, St-Pierre P, Perreault JP. 2010. In-depth sequencing of the siRNAs associated with peach latent mosaic viroid infection. *BMC Mol. Biol.* 11:16
12. Candresse T, Filloux D, Muhire B, Julian C, Galzi S, et al. 2014. Appearances can be deceptive: revealing a hidden viral infection with deep sequencing in a plant quarantine context. *PLOS ONE* 9:e102945
13. Candresse T, Marais A, Faure C, 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* 103:293–8
14. Carvajal-Yepes M, Olaya C, Lozano I, Cuervo M, Castano M, Cuellar WJ. 2014. Unraveling complex viral infections in cassava (*Manihot esculenta* Crantz) from Colombia. *Virus Res.* 186:76–86
15. Coetzee B, Freeborough MJ, Maree HJ, Celton JM, Rees DJ, Burger JT. 2010. Deep sequencing analysis of viruses infecting grapevines: virome of a vineyard. *Virology* 400:157–63
16. Cox-Foster DL, Conlan S, Holmes EC, Palacios G, Evans JD, et al. 2007. A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* 318:283–87
17. Cuellar WJ, Cruzado RK, Fuentes S, Untiveros M, Soto M, Kreuze JF. 2011. Sequence characterization of a Peruvian isolate of *Sweet potato chlorotic stunt virus*: further variability and a model for p22 acquisition. *Virus Res.* 157:111–15
18. de Andrade RRS, Vaslin MFS. 2014. SearchSmallRNA: a graphical interface tool for the assemblage of viral genomes using small RNA libraries data. *Virol. J.* 11:45
19. De Souza J, Fuentes S, Savenkov EI, Cuellar W, Kreuze JF. 2013. The complete nucleotide sequence of sweet potato C6 virus: a carlavirus lacking a cysteine-rich protein. *Arch. Virol.* 158:1393–96
20. Delwart EL. 2007. Viral metagenomics. *Rev. Med. Virol.* 17:115–31
21. Di Serio F, Gisel A, Navarro B, Delgado S, Martinez de Alba AE, et al. 2009. Deep sequencing of the small RNAs derived from two symptomatic variants of a chloroplastic viroid: implications for their genesis and for pathogenesis. *PLOS ONE* 4:e7539
22. Ding S-W, Lu R. 2011. Virus-derived siRNAs and piRNAs in immunity and pathogenesis. *Curr. Opin. Virol.* 1:533–44
23. Ding SW. 2010. RNA-based antiviral immunity. *Nat. Rev. Immunol.* 10:632–44
24. Ding SW, Lu R. 2011. Virus-derived siRNAs and piRNAs in immunity and pathogenesis. *Curr. Opin. Virol.* 1:533–44
25. Ding SW, Wang Y, Cao M, Ramachandran V, Du P, Wu Q. 2012. Development of next-generation technologies for the diagnosis and identification of citrus viruses and viroids. *Citrograph* Nov./Dec.:32–35
26. Dodds JA, Morris TJ, Jordan RL. 1984. Plant viral double-stranded RNA. *Annu. Rev. Phytopathol.* 22:151–68
27. Dombrovsky A, Glanz E, Lachman O, Sela N, Doron-Faigenboim A, Antignus Y. 2013. The complete genomic sequence of *Pepper yellow leaf curl virus* (PYLCV) and its implications for our understanding of evolution dynamics in the genus Polerovirus. *PLOS ONE* 8:e70722
28. Dombrovsky A, Sapkota R, Lachman O, Antignus Y. 2012. *Eggplant mild leaf mottle virus* (EMLMV), a new putative member of the genus Ipomovirus that harbors an HC-Pro gene. *Virus Genes* 44:329–37
29. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–61
30. Feng H, Shuda M, Chang Y, Moore PS. 2008. Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science* 319:1096–100
31. Fuentes S, Heider B, Tasso RC, Romero E, Zum Felde T, Kreuze JF. 2012. Complete genome sequence of a potyvirus infecting yam beans (*Pachyrhizus* spp.) in Peru. *Arch. Virol.* 157:773–76
32. Giampetruzzi A, Roumi V, Roberto R, Malossini U, Yoshikawa N, et al. 2012. A new grapevine virus discovered by deep sequencing of virus- and viroid-derived small RNAs in Cv *Pinot gris*. *Virus Res.* 163:262–8

33. Goodman RM. 1977. Single-stranded DNA genome in a whitefly-transmitted plant virus. *Virology* 83:171–79
34. Gu YH, Tao X, Lai XJ, Wang HY, Zhang YZ. 2014. Exploring the polyadenylated RNA virome of sweet potato through high-throughput sequencing. *PLOS ONE* 9:e98884
35. Hagen C, Frizzi A, Kao J, Jia L, Huang M, et al. 2011. Using small RNA sequences to diagnose, sequence, and investigate the infectivity characteristics of vegetable-infecting viruses. *Arch. Virol.* 156:1209–16
36. Hany U, Adams IP, Glover R, Bhat AI, Boonham N. 2014. The complete genome sequence of *Piper yellow mottle virus* (PYMoV). *Arch. Virol.* 159:385–38
37. Harrison BD. 1985. Advances in geminivirus research. *Annu. Rev. Phytopathol.* 23:55–82
38. Harrison BD, Barker H, Bock KR, Guthrie EJ, Meredith G, Atkinson M. 1977. Plant viruses with circular single-stranded DNA. *Nature* 270:760–62
39. He Y, Yang Z, Hong N, Wang G, Ning G, Xu W. 2015. Deep sequencing reveals a novel closterovirus associated with wild rose leaf rosette disease. *Mol. Plant Pathol.* 16:449–58
40. Ito T, Suzaki K, Nakano M. 2013. Genetic characterization of novel putative rhabdovirus and dsRNA virus from Japanese persimmon. *J. Gen. Virol.* 94:1917–21
41. Ito T, Suzaki K, Nakano M, Sato A. 2013. Characterization of a new apscaviroid from American persimmon. *Arch. Virol.* 158:2629–31
42. Jeck WR, Reinhardt JA, Baltrus DA, Hickenbotham MT, Magrini V, et al. 2007. Extending assembly of short DNA sequences to handle error. *Bioinformatics* 23:2942–44
43. Jiang HS, Lei R, Ding SW, Zhu SF. 2014. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* 15:182
44. Kambara H, Takahashi S. 1993. Multiple-sheathflow capillary array DNA analyser. *Nature* 361:565–66
45. Kashif M, Pietila S, Artola K, Jones RAC, Tugume AK, et al. 2012. Detection of viruses in sweetpotato from Honduras and Guatemala augmented by deep-sequencing of small-RNAs. *Plant Dis.* 96:1430–37
46. Kim MS, Whon TW, Bae JW. 2013. Comparative viral metagenomics of environmental samples from Korea. *Genomics Inform.* 11:121–28
47. Kircher M, Kelso J. 2010. High-throughput DNA sequencing—concepts and limitations. *Bioessays* 32:524–36
48. Kreuze J, Koenig R, De Souza J, Vetten HJ, Muller G, et al. 2013. The complete genome sequences of a Peruvian and a Colombian isolate of Andean potato latent virus and partial sequences of further isolates suggest the existence of two distinct potato-infecting tymovirus species. *Virus Res.* 173:431–35
49. Kreuze JF, Perez A, Untiveros M, Quispe D, Fuentes S, et al. 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
50. Kutnjak D, Silvestre R, Cuellar W, Perez W, Muller G, et al. 2014. Complete genome sequences of new divergent potato virus X isolates and discrimination between strains in a mixed infection using small RNAs sequencing approach. *Virus Res.* 191:45–50
51. Lee S, Hallam SJ. 2009. Extraction of high molecular weight genomic DNA from soils and sediments. *J. Vis. Exp.* 33:1569
52. Li L, Victoria J, Kapoor A, Blinkova O, Wang C, et al. 2009. A novel picornavirus associated with gastroenteritis. *J. Virol.* 83:12002–6
53. Li R, Gao S, Fei Z, Ling KS. 2013. Complete genome sequence of a new tobamovirus naturally infecting tomatoes in Mexico. *Genome Announc.* 1:e00794-13
54. Li R, Gao S, Hernandez AG, Wechter WP, Fei Z, Ling KS. 2012. Deep sequencing of small RNAs in tomato for virus and viroid identification and strain differentiation. *PLOS ONE* 7:e37127
55. Li Y, Lu JF, Han YH, Fan XX, Ding SW. 2013. RNA interference functions as an antiviral immunity mechanism in mammals. *Science* 342:231–34
56. Loconsole G, Onelge N, Potere O, Giampetruzzi A, Bozan O, et al. 2012. Identification and characterization of *Citrus yellow vein clearing virus*, a putative new member of the genus *Mandarivirus*. *Phytopathology* 102:1168–75
57. Loconsole G, Saldarelli P, Doddapaneni H, Savino V, Martelli GP, Saponari M. 2012. Identification of a single-stranded DNA virus associated with citrus chlorotic dwarf disease, a new member in the family Geminiviridae. *Virology* 432:162–72

58. Maillard PV, Ciaudo C, Marchais A, Li Y, Jay F, et al. 2013. Antiviral RNA interference in mammalian cells. *Science* 342:235–38
59. Mardis ER. 2008. Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.* 9:387–402
60. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–80
61. Massart S, Olmos A, Jijakli H, Candresse T. 2014. Current impact and future directions of high throughput sequencing in plant virus diagnostics. *Virus Res.* 188:90–96
62. Mbanzibwa DR, Tugume AK, Chiunga E, Mark D, Tairo FD. 2014. Small RNA deep sequencing-based detection and further evidence of DNA viruses infecting sweetpotato plants in Tanzania. *Ann. Appl. Biol.* 165:329–39
63. Metzker ML. 2010. Sequencing technologies—the next generation. *Nat. Rev. Genet.* 11:31–46
64. Mokili JL, Rohwer F, Dutilh BE. 2012. Metagenomics and future perspectives in virus discovery. *Curr. Opin. Virol.* 2:63–77
65. Molina LG, da Fonseca GC, de Morais GL, de Oliveira LFV, de Carvalho JB, et al. 2012. Metatranscriptomic analysis of small RNAs present in soybean deep sequencing libraries. *Genet. Mol. Biol.* 35(Suppl. 1):292–303
66. Monger WA, Adams IP, Glover RH, Barrett B. 2010. The complete genome sequence of *Canna yellow streak virus*. *Arch. Virol.* 155:1515–18
67. Monger WA, Alicai T, Ndunguru J, Kinyua ZM, Potts M, et al. 2010. The complete genome sequence of the Tanzanian strain of *Cassava brown streak virus* and comparison with the Ugandan strain sequence. *Arch. Virol.* 155:429–33
68. Nakamura S, Yang CS, Sakon N, Ueda M, Tougan T, et al. 2009. Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. *PLOS ONE* 4:e4219
69. Navarro B, Pantaleo V, Gisel A, Moxon S, Dalmay T, et al. 2009. Deep sequencing of viroid-derived small RNAs from grapevine provides new insights on the role of RNA silencing in plant-viroid interaction. *PLOS ONE* 4:e7686
70. Nicolaisen M. 2011. An oligonucleotide-based microarray for detection of plant RNA viruses. *J. Virol. Methods* 173:137–43
71. Palacios G, Druce J, Du L, Tran T, Birch C, et al. 2008. A new arenavirus in a cluster of fatal transplant-associated diseases. *N. Engl. J. Med.* 358:991–98
72. Pallett DW, Ho T, Cooper I, Wang H. 2010. Detection of *Cereal yellow dwarf virus* using small interfering RNAs and enhanced infection rate with *Cocksfoot streak virus* in wild cocksfoot grass (*Dactylis glomerata*). *J. Virol. Methods* 168:223–27
73. Pantaleo V, Saldarelli P, Miozzi L, Giampetruzzi A, Gisel A, et al. 2010. Deep sequencing analysis of viral short RNAs from an infected Pinot Noir grapevine. *Virology* 408:49–56
74. Petrosino JF, Highlander S, Luna RA, Gibbs RA, Versalovic J. 2009. Metagenomic pyrosequencing and microbial identification. *Clin. Chem.* 55:856–66
75. Poojari S, Alabi OJ, Fofanov VY, Naidu RA. 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
76. Quito-Avila DF, Jelkmann W, Tzanetakis IE, Keller K, Martin RR. 2011. Complete sequence and genetic characterization of *Raspberry latent virus*, a novel member of the family Reoviridae. *Virus Res.* 155:397–405
77. Remmert M, Biegert A, Hauser A, Soding J. 2012. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods* 9:173–75
78. Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, et al. 2010. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* 466:334–38
79. Richards RS, Adams IP, Kreuze JF, De Souza J, Cuellar W, et al. 2014. The complete genome sequences of two isolates of potato black ringspot virus and their relationship to other isolates and nepoviruses. *Arch. Virol.* 159:811–15

80. Rodamilans B, Leon DS, Muhlberger L, Candresse T, Neumuller M, et al. 2014. Transcriptomic analysis of *Prunus domestica* undergoing hypersensitive response to *Plum pox virus* infection. *PLOS ONE* 9:e100477
81. Rodriguez Pardina PE, Bejerman N, Luque AV, Di Feo L. 2012. Complete nucleotide sequence of an Argentinean isolate of sweet potato virus G. *Virus Genes* 45:593–95
82. Romanovskaya A, Sarin LP, Bamford DH, Poranen MM. 2013. High-throughput purification of double-stranded RNA molecules using convective interaction media monolithic anion exchange columns. *J. Chromatogr. A* 1278:54–60
83. Roossinck MJ. 2011. The big unknown: plant virus biodiversity. *Curr. Opin. Virol.* 1:63–67
84. Roossinck MJ. 2012. Plant virus metagenomics: biodiversity and ecology. *Annu. Rev. Genet.* 46:359–69
85. Roossinck MJ, Saha P, Wiley GB, Quan J, White JD, et al. 2010. Ecogenomics: using massively parallel pyrosequencing to understand virus ecology. *Mol. Ecol.* 19(Suppl. 1):81–88
86. Roy A, Choudhary N, Guillermo LM, Shao J, Govindarajulu A, et al. 2013. A novel virus of the genus *Cilevirus* causing symptoms similar to citrus leprosis. *Phytopathology* 103:488–500
87. Saqib M, Wylie SJ, Jones MGK. 2014. Serendipitous identification of a new *Iflavirus*-like virus infecting tomato and its subsequent characterization. *Plant Pathol.* 64:519–24
88. Schadt EE, Turner S, Kasarskis A. 2010. A window into third-generation sequencing. *Hum. Mol. Genet.* 19:R227–40
89. Schowalter RM, Pastrana DV, Pumphrey KA, Moyer AL, Buck CB. 2010. Merkel cell polyomavirus and two previously unknown polyomaviruses are chronically shed from human skin. *Cell Host Microbe* 7:509–15
90. Schulz MH, Zerbino DR, Vingron M, Birney E. 2012. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28:1086–92
91. Seguin J, Rajeswaran R, Malpica-Lopez N, Martin RR, Kasschau K, et al. 2014. De novo reconstruction of consensus master genomes of plant RNA and DNA viruses from siRNAs. *PLOS ONE* 9:e88513
92. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, et al. 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 309:1728–32
93. Sheveleva A, Kudryavtseva A, Speranskaya A, Belenikin M, Melnikova N, Chirkov S. 2013. Complete genome sequence of a novel *Plum pox virus* strain W isolate determined by 454 pyrosequencing. *Virus Genes* 47:385–8
94. van der Meijden E, Janssens RW, Lauber C, Bouwes Bavinck JN, Gorbalenya AE, Feltkamp MC. 2010. Discovery of a new human polyomavirus associated with *Trichodysplasia spinulosa* in an immunocompromized patient. *PLOS Pathog.* 6:e1001024
95. Verbeek M, Dullemans AM, van Raaij HMG, Verhoeven JTJ, van der Vlugt RAA. 2014. *Lettuce necrotic leaf curl virus*, a new plant virus infecting lettuce and a proposed member of the genus *Torradovirus*. *Arch. Virol.* 159:801–5
96. Victoria JG, Kapoor A, Li L, Blinkova O, Slikas B, et al. 2009. Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. *J. Virol.* 83:4642–51
97. Vives MC, Velazquez K, Pina JA, Moreno P, Guerri J, Navarro L. 2013. Identification of a new Enamovirus associated with citrus vein enation disease by deep sequencing of small RNAs. *Phytopathology* 103:1077–86
98. Wang YL, Cheng XF, Wu XX, Wang AM, Wu XY. 2014. Characterization of complete genome and small RNA profile of pagoda yellow mosaic associated virus, a novel badnavirus in China. *Virus Res.* 188:103–8
99. Widana Gamage S, Persley DM, Higgins CM, Dietzgen RG. 2015. First complete genome sequence of a capsicum chlorosis tospovirus isolate from Australia with an unusually large S RNA intergenic region. *Arch. Virol.* 160:869–72
100. Willner D, Furlan M, Schmieder R, Grasis JA, Pride DT, et al. 2011. Metagenomic detection of phage-encoded platelet-binding factors in the human oral cavity. *Proc. Natl. Acad. Sci. USA* 108(Suppl. 1):4547–53
101. Wu Q, Luo Y, Lu R, Lau N, Lai EC, et al. 2010. Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs. *Proc. Natl. Acad. Sci. USA* 107:1606–11

102. Wu Q, Wang Y, Cao M, Pantaleo V, Burgyan J, et al. 2012. Homology-independent discovery of replicating pathogenic circular RNAs by deep sequencing and a new computational algorithm. *Proc. Natl. Acad. Sci. USA* 109:3938–43
103. Wylie S, Jones M. 2011. Hardenbergia virus A, a novel member of the family *Betaflexiviridae* from a wild legume in Southwest Australia. *Arch. Virol.* 156:1245–50
104. Wylie SJ, Jones MGK. 2011. Characterisation and quantitation of mutant and wild-type genomes of Hardenbergia mosaic virus isolates co-infecting a wild plant of *Hardenbergia comptoniana*. *Arch. Virol.* 156:1251–55
105. Wylie SJ, Jones MGK. 2011. The complete genome sequence of a *Passion fruit woodiness virus* isolate from Australia determined using deep sequencing, and its relationship to other potyviruses. *Arch. Virol.* 156:479–82
106. Wylie SJ, Jones MGK. 2012. Complete genome sequences of seven carlavirus and potyvirus isolates from *Narcissus* and *Hippeastrum* plants in Australia, and proposals to clarify their naming. *Arch. Virol.* 157:1471–80
107. Wylie SJ, Li H, Dixon KW, Richards H, Jones MGK. 2013. Exotic and indigenous viruses infect wild populations and captive collections of temperate terrestrial orchids (*Diuris* species) in Australia. *Virus Res.* 171:22–32
108. Wylie SJ, Li H, Liu J, Jones MGK. 2014. First report of *Narcissus mosaic virus* from Australia and from *Iris*. *Australas. Plant Dis. Notes* 9:1–2
109. Wylie SJ, Li H, Jones MGK. 2012. First report of an isolate of Japanese iris necrotic ring virus from Australia. *Australas. Plant Dis. Notes* 7:107–10
110. Wylie SJ, Li H, Jones MGK. 2013. Donkey orchid symptomless virus: a viral 'platypus' from Australian terrestrial orchids. *PLOS ONE* 8:e79587
111. Wylie SJ, Li H, Jones MGK. 2014. Yellow tailflower mild mottle virus: a new tobamovirus described from *Anthocercis littorea* (Solanaceae) in Western Australia. *Arch. Virol.* 159:791–95
112. Wylie SJ, Li H, Saqib M, Jones MGK. 2014. The global trade in fresh produce and the vagility of plant viruses: a case study in garlic. *PLOS ONE* 9:e105044
113. Wylie SJ, Tan AJ, Li H, Dixon KW, Jones MGK. 2012. Caladenia virus A, an unusual new member of the family Potyviridae from terrestrial orchids in Western Australia. *Arch. Virol.* 157:2447–52
114. Xie G, Yu J, Duan Z. 2013. New strategy for virus discovery: viruses identified in human feces in the last decade. *Sci. China Life Sci.* 56:688–96
115. Zablocki O, Pietersen G. 2014. Characterization of a novel citrus tristeza virus genotype within three cross-protecting source GFMS12 sub-isolates in South Africa by means of Illumina sequencing. *Arch. Virol.* 159:2133–39
116. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18:821–29
117. Zhang Y, Singh K, Kaur R, Qiu W. 2011. Association of a novel DNA virus with the grapevine vein-clearing and vine decline syndrome. *Phytopathology* 101:1081–90
118. Zhang YL, Yu NT, Huang QX, Yin GH, Guo AP, et al. 2014. Complete genome of Hainan papaya ringspot virus using small RNA deep sequencing. *Virus Genes* 48:502–8
119. Zhang ZX, Qi SS, Tang N, Zhang XX, Chen SS, et al. 2014. Discovery of replicating circular RNAs by RNA-Seq and computational algorithms. *PLOS Pathog.* 10:e1004553