

# Applications of Next Generation High Throughput Sequencing Technologies in Characterization, Discovery and Molecular Interaction of Plant Viruses

K. Prabha · V. K. Baranwal · R. K. Jain

Received: 6 October 2012 / Accepted: 20 April 2013 / Published online: 11 May 2013  
© Indian Virological Society 2013

**Abstract** Present era of molecular biology is witnessing revolutionary developments in sequencing technology. This advancement has considerably influenced plant virology in the field of diagnostics and host virus interaction. Next generation high-throughput sequencing technology has made it possible to directly detect, identify and discover novel viruses in several plants in an unbiased manner without antibodies or prior knowledge of the virus sequences. Entire viral genome could be sequenced from symptomatic or asymptomatic plants through next generation sequencing of total nucleic acids including small RNAs. It provides census of viral population in a particular ecosystem or cropping system. Viral genome variability, evolution within the host and virus defence mechanism in plants can also be easily understood by massive parallel sequencing. In this article, we provide an overview of the applications of next generation sequencing technology in characterization, discovery and molecular interaction of plant viruses.

**Keywords** Deep sequencing · Plant virome · Metagenomics

## Introduction

Viruses comprise almost half (47 %) of the reported plant emerging infectious diseases [4]. The detection and

identification of existing, emerging and new viruses have been relied on a large range of techniques viz., biological assay, electron microscopy, nucleic acid based techniques like polymerase chain reaction (PCR) and its variants, serology based techniques such as enzyme linked immunosorbent assay (ELISA) and more recently microarrays [7, 26]. Biological assays and electron microscopy are the oldest techniques employed as a “broad screen” for plant virus detection in situations of identification of new viruses [24]. Biological testing is dependent on the availability of suitable high quality indicator and propagative host plants. It also requires appropriate environmental conditions and in some cases symptoms appearance on indicator hosts may take several months. Moreover, the interpretation of test results is subjective. Electron microscopy provides direct evidence for the presence of virus and depends upon stains used and concentration of the virus in the plant saps however, it is not confirmative. Virus diagnostics became easily available because of two major breakthroughs during last four decades. The first one was serological assay in form of ELISA which used monoclonal or polyclonal antibodies for detection of viruses [10]. The other one was the use of in vitro amplification of DNA commonly called as PCR [8]. Both these techniques were modified for broad based detection of plant viruses. Group-specific degenerate primers were designed and used for broad-based detection of several plant virus genera [22]. The degenerate primers based on less number of sequences often give false results. The “consensus decay” of potential primer sequences has been observed as the sequence information on viruses is increasing over time [44]. Another drawback in using degenerate primer sets is the inability of the resultant PCR protocol to detect all, or at least only a fraction of the viruses that are known or unknown in the targeted virus group. Also, with poorly designed primers, non-specific

---

K. Prabha · V. K. Baranwal (✉) · R. K. Jain (✉)  
Plant Virology Unit, Division of Plant Pathology, Indian  
Agricultural Research Institute, New Delhi 110012, India  
e-mail: vbaranwal2001@yahoo.com

R. K. Jain  
e-mail: rakeshjain56@yahoo.co.in

amplification of plant host nucleic acid may occur resulting in false positive reactions. Cocktail polyclonal antibodies were used for broad based detection of relative strains or different species of virus but further characterization is difficult. Thus these techniques suffer from several significant drawbacks, especially when used in diagnostics of ‘unknown’ agents, e.g. either a virus infecting a new host or a new uncharacterized virus or when the virus is significantly varying from the other species of the same genus as all these routinely used techniques require previous knowledge about the virus, either its sequence information or antibody. Given these drawbacks, novel approaches such as next generation sequencing (NGS) technologies have been used for improved detection and characterization of new or unusual viruses affecting diverse plants. Next generation sequencing has also been used to dissect the mystery behind diseases of unknown viral etiology as well as host interaction of plant viruses. This review provides an insight into the progress made in applications of NGS in characterization, discovery and molecular interaction of plant viruses Table 1.

### Next Generation Sequencing (NGS) Technologies

During earlier days decipheration of genetic information encoded in DNA was achieved through capillary electrophoresis (CE)-based Sanger sequencing. The NGS technology based on the similar principle of CE, involves rapid sequencing of large stretches of DNA base pairs spanning entire genomes, with the latest instruments capable of producing hundreds of giga bases of data in a single sequencing run. It overcomes the inherent draw backs of Sanger sequencing in throughput, scalability, speed, and resolution. The principles and methodology of NGS has been reviewed by many authors and recently Capobianchi and associates [9] has discussed this in details along with application of NGS in clinical virology. High throughput NGS techniques are powerful tools as they are the backbone of metagenomics based strategy for identification of unknown disease associated viruses and discovery of novel viruses. It has made a huge impact in many areas of microbial evolution, mechanisms of pathogenesis and phylogeography [36]. Next Generation Sequencing or second and third -generation sequencing enables the direct detection, identification and discovery of viruses in an unbiased manner without requiring antibodies or prior knowledge of the pathogen macromolecular sequence [5, 12, 32]. The 454 FLX pyrosequencing platform (<http://www.454.com/>), developed by 454 life sciences was the first next-generation high-throughput sequencing technology. The genome analyzer (<http://www.illumina.com>) of Illumina, was released in early 2007 developed by Solexa GA. In recent years many more

improved platforms have been constructed and released one after the other, like Heliscope by Helicos (<http://www.helicosbio.com/>), Ion Torrent PGM by Life Technologies (<http://www.iontorrent.com/>) and a real-time sequencing platform by Pacific Biosciences (<http://www.pacificbiosciences.com/>). The real time sequencing platform as well as other novel sequencing platforms are referred as “third-generation” because they sequence processively single large DNA molecules without the need to halt between read steps. 454 pyrosequencing, Illumina GA and SOLiD methods sequence populations of amplified template-DNA molecules with a typical “wash-and-scan” technique and represent second generation system [34]. Ion Torrent PGM and Heliscope sequencing technologies are in between “second-” and “third-generation” technologies, since they possess many of the second and some of the third generation technologies [5].

### Applications of NGS Technologies in Plant Virology

Next-generation sequencers have hastened up the advancement in hunting for viruses and their diagnostics by metagenomic analysis and deep sequencing. De novo sequencing of viruses using deep sequencing is a new technique that has successfully identified known and unknown viruses from long or short reads [9]. Different approaches are employed in next generation deep sequencing and metagenomics for the detection and identification of plant viruses. One such method is the isolation of total nucleic acid from infected host and massive sequencing. Adams et al. [1] analyzed total RNA from tomatoes infected with the *Pepino mosaic virus* (PepMV). Another strategy is the enriching of viral RNA by increasing the amount of dsRNA and sequencing of this dsRNA fraction using NGS [11]. In a virus-infected plant, small interfering RNAs (siRNAs) derived from the viral genome form a considerable proportion of the small RNA population. Reassembling significant portions of the virus sequence from overlapping siRNA sequences to identify the virus is another successful method employed through massive sequencing of siRNAs [20]. Vector-enabled metagenomics using NGS is another practical strategy to study the diversity of viruses whose transmission is through vectors [28]. Applications of NGS technologies in phyto-virology have been discussed.

#### Discovery of Novel Viruses and Resolving Unknown Etiology of Diseases

High-throughput parallel sequencing of small RNAs was used for the first time to identify novel plant viruses and

**Table 1** Applications of next generation sequencing technologies in plant virus research

| Species                               | Strategy   | Sequencing platform                | Virus studied   | Application   | Reference  |  |  |
|---------------------------------------|--|------------------------------------|---|---|------------|--|--|
| <b>Plants system</b>                  |  |                                    |   |   |            |  |  |
| <i>Ipomoea batatas</i> (sweet potato) | Small RNA deep sequencing  | Illumina                           | <i>Sweet potato feathery mottle virus (Potyvirus)</i> ; <i>Sweet potato chlorotic stunt virus (Crinivirus)</i> ; two novel DNA viruses ( <i>Badnavirus</i> & <i>Mastrevirus</i> )   | Discovery of novel virus; complete genome sequencing and RNAi | [20]       |  |  |
| <i>Solanum lycopersicum</i> (Tomato)  | Small RNA deep sequencing  | Illumina                           | <i>Potato spindle tuber viroid (Pospiviroid)</i> ; <i>Pepino mosaic virus (Potexvirus)</i> and a novel <i>Potyvirus</i> provisionally named ‘‘Tomato necrotic stunt virus’’   | Discovery of novel virus and complete genome sequencing       | [21]       |  |  |
|                                       | Next-generation sequencing coupled with metagenomic analysis of cDNA | 454 FLX                            | <i>Pepino mosaic virus (Potexvirus)</i>   | Complete genome sequencing                                    | [1]        |  |  |
|                                       | Small RNA deep sequencing  | Illumina                           | Two strains of <i>tomato spotted wilt virus (Tospovirus)</i> ; an unidentified <i>Tospovirus</i> ; a squash-infecting DNA virus   | Quasi species analysis  | [17]       |  |  |
| <i>Vitis vinifera</i> (Grapevine)     | Sequencing of dsRNAs   | 454 FLX                            | <i>Tomato yellow leaf curl virus (Begomovirus)</i>  | Virus–host interaction  | [14]       |  |  |
|                                       |  | 454 FLX                            | <i>Penicillium chrysogenum virus</i> , <i>Cryphonectria nitschkei chrysovirus 1</i> , <i>Helminthosporium victoriae 145S virus</i> ( <i>Chrysovirus</i> ); <i>Cryphonectria hypovirus 3 (Hypovirus)</i> ; <i>Ophiostoma mitovirus</i> , <i>Cryphonectria parasitica mitovirus 1-NB631</i> , <i>Thielaviopsis mitovirus 1</i> ; <i>Botrytis cinerea</i> debilitation-related virus ( <i>Mitovirus</i> ); <i>Grennmetiella abietina</i> RNA virus 1, <i>Fusarium poae virus 1</i> , <i>Aspergillus ochraceus</i> <i>Virus1 (Partivirus)</i> ; <i>Pepper cryptic virus 1 (Alphacryptovirus)</i> ; <i>Saccharomyces cerevisiae virus L-A 1 (L1) (Totivirus)</i> , <i>Botryotinia fuckeliana virus 1</i> , <i>Helminthosporium victoriae virus 190S</i> , Black raspberry virus F, <i>Sphaeropsis sapinea</i> RNA virus 1, <i>Magnaporthe oryzae virus 1 (Victorivirus)</i> ; <i>Botrytis virus F (Mycoflexivirus)</i> ; <i>Botrytis virus x (Botrevirus)</i> and <i>Fusarium graminearum dsRNA virus</i> ; <i>Amasya cherry disease associated mycovirus</i> , <i>Diaporthe ambigua</i> RNA virus, <i>Curvularia thermal tolerance virus</i> (unassigned) | Discovery of novel virus and the ‘‘virome’’ sequencing        | [2, 3, 11] |  |  |
|                                       |  | Illumina                           | <i>Grapevine rupestris stem pitting-associated virus</i> , <i>Hop stunt viroid (Hostuviroid)</i> and <i>Grapevine yellow speckle viroid 1 (Apscaviroid)</i> , <i>Grapevine rupestris vein feathering virus</i> and <i>Grapevine Syrah virus 1 (Marafivirus)</i>   | Discovery of novel virus and the ‘‘virome’’ sequencing        | [16]       |  |  |
|                                       |  | 454 FLX                            | Four <i>Potato virus Y</i> variants ( <i>Potyvirus</i> )  | Quasi species analysis  | [15]       |  |  |
|                                       |  | Illumina                           | <i>Hop stunt viroid (Hostuviroid)</i>   | Virus-host interaction  | [23]       |  |  |
|                                       |  | Illumina                           | <i>Rice stripe virus (Tenivirus)</i>  | Virus-host interaction  | [42]       |  |  |
|                                       |  | 454 FLX                            | <i>Melon necrotic spot virus (Carmovirus)</i> ; <i>Watermelon mosaic virus (Potyvirus)</i>  | Virus-host interaction  | [14]       |  |  |
|                                       |  | Illumina                           | Cotton leafroll dwarf virus ( <i>Polerovirus</i> )  | Virus-host interaction  | [35]       |  |  |
|                                       |  | <i>Capsicum annuum</i> L. (pepper) | Total RNA deep sequencing   | 454 FLX   |            |  |  |
|                                       |  | <i>Cucumis sativus</i> (cucumber)  | Small RNA deep sequencing   | Illumina  |            |  |  |
| <i>Oryza japonica</i> (rice)          | Small RNA deep sequencing  | Illumina                           |   |   |            |  |  |
| <i>Cucumis melo</i> (muskmelon)       | Small RNA deep sequencing  | 454 FLX                            |   |   |            |  |  |
| <i>Gossypium hirsutum</i> (cotton)    | Virus derived dsRNA deep sequencing                                  | Illumina                           |   |   |            |  |  |

Table 1 continued

| Species   | Strategy  | Sequencing platform                        | Virus studied   | Application  | Reference  |
|---|---|--|---|--|--|
| <i>Arabidopsis thaliana</i>   | Small RNA deep sequencing                                 | Illumina                                   | <i>Tobacco mosaic virus (Tobamovirus)</i><br><i>Cauliflower mosaic virus (Caulimovirus)</i><br>Oilseed rape mosaic virus ( <i>Tobamovirus</i> )<br><i>Cucumber mosaic virus (Bromovirus)</i> , <i>Tobacco rattle virus (Tobravirus)</i> ; <i>Turnip mosaic virus (Potyvirus)</i><br><i>Cymbidium ringspot virus (Tombusvirus)</i><br><i>Pepper mild mottle virus (Badnavirus)</i> ; <i>Potato virus X (Potexvirus)</i><br>'Gayfeather mild mottle virus' (Unassigned) | Virus-host interaction<br>Virus-host interaction<br>Virus-host interaction<br>Virus-host interaction<br>Virus-host interaction<br>Virus-host interaction<br>Discovery of novel virus | [30]<br>[6]<br>[18]<br>[14]<br>[14, 41]<br>[14]<br>[1] |
| <i>Nicotiana benthamiana</i>  | Small RNA deep sequencing                                 | 454 FLX                                    | <i>Cereal yellow dwarf virus (Polerovirus)</i>  | Discovery of novel virus   | [29]   |
| <i>Gomphrena Globosa</i> (Globe Amaranth or Bachelor Button)<br><i>Dactylis glomerata</i> (cocksfoot grass) | Metagenomic analysis of cDNA<br>Small RNA deep sequencing | 454 FLX and Illumina<br>454 FLX<br>454-FLX | <i>Hardenbergia mosaic virus</i> , an isolate of <i>Passionfruit woodiness virus (Polyvirus)</i> , and an undescribed virus with identity to members of the family <i>Betaflexiviridae</i>  | Discovery of novel virus and the "virome" sequencing   | [40]   |
| <i>Hardenbergia comptoniana</i> (Lilac Vine) and <i>Passiflora carulea</i> (Bluecrown passionflower)        | Polyadenylated RNA deep sequencing                        | Illumina                                   |   |  |  |
| <b>Insect vectors</b><br>Whiteflies ( <i>Bemisia tabaci</i> )   | Vector enabled metagenomics                               | –  | <i>Cucurbit leaf crumple virus (Begomovirus)</i>  | Discovery of novel virus and the "virome" sequencing   | [28]   |

sequence of an entire viral genome from diseased, as well as symptomless sweet potato plants was obtained by Kreuze et al. [20]. Contigs were assembled from using small sequence assembly software (velvet) and RNA, ssDNA and dsDNA reverse transcribing viruses were identified. A novel *Cucumovirus* was identified from deep sequencing of long cDNA reads [1]. A sample of *Gomphrena globosa* infected through mechanical inoculation with an unknown pathogen originally isolated from the flowering plant *Liatrix spicata*. RNA from this *G. globosa* sample was sequenced using one-quarter of a plate from a GSFLX Genome Sequencer (Advanced Genomics Facility, Liverpool University, Liverpool, UK) and it contained a new *Cucumovirus*, for which the name ‘*Gayfeather mild mottle virus*’ was suggested. Analysis of small RNA sequences from leaves of wild *Dactylis glomerata* (cocksfoot grass) using deep sequencing (454 Life Sciences, Roche Diagnostics) revealed unrecognised *Cereal yellow dwarf virus*, genus *Luteovirus*, family Luteoviridae) infections in a wild population of cocksfoot grass [29].

Analysis of small RNA populations in two cultivars of Pinot gris (PG) grapevines showing field symptoms of chlorotic mottling and leaf deformations in the Trentino region (Italy) also led to the discovery of new species of virus. Illumina high throughput sequencing showed an unrecorded virus along with other known viruses. This virus had a genome organization identical to that of *Grapevine berry inner necrosis virus* (GINV), a *Trichovirus* reported only from Japan, but molecular differences with GINV were wide enough to warrant classification of the virus in question as a new species, for which the provisional name of *Grapevine Pinot gris virus* is proposed [16]. Ng et al. [28] used vector enabled metagenomics to describe the DNA viral community from whiteflies (*Bemisia tabaci*) collected from two important agricultural regions in Florida, USA and successfully characterized the active and abundant viruses that produce disease symptoms in crops, dominated by *Cucurbit leaf crumple virus*, a *Begomovirus* known to infect watermelon plants at this sampling site as well as the less abundant viruses infecting adjacent native vegetation. Small RNA deep sequencing using NGS technology followed by de novo sequence assembly of virus infection suspected tomato samples could identify a novel *Potyvirus* with less than 60 % overall genome nucleotide sequence identity to other known viruses and its full genome sequence was obtained [21].

#### Ecogenomics of Viruses and Characterization of Viral Population or “Virome”

Next generation sequencing enabled massive parallel sequencing or metagenomics allows the census of virus

populations in a particular ecosystem or cropping system and the viral population is known as “virome”. Roossinck and associates [33] conducted a biodiversity inventory of viruses through metagenomics of 10–20 g of plant materials collected from Tall Grass Prairie Preserve in north-eastern Oklahoma, an area of relatively low plant diversity; and the *Area de Conservacio'n Guanacaste* in northwestern Costa Rica, an area of extremely high plant diversity and observed the infection of viruses of the following families; Bromoviridae, Caulimoviridae, Chrysovriidae, Closteroviridae, Endornaviridae, Luteoviridae, Narnaviridae, Partitiviridae, Potyviridae, Totiviridae, Tymoviridae and some of the unclassified viruses. In their metagenomic study both polyadenylated and non-polyadenylated viruses including DNA viruses could be sequenced. On the other hand multiple polyadenylated RNA viruses could be detected in pooled cultivated and wild plant samples using massively parallel sequencing of polyadenylated RNA species from RNA extracted from 120 leaf specimens from 17 plant species and twelve of the viruses identified belonged to previously described genera *Potyvirus*, *Nepovirus*, *Allexivirus*, and *Carlavirus* and four were unknown and are proposed as members of the genera *Potyvirus*, *Sadwavirus*, and *Trichovirus* [38]. Polyadenylated RNA from a plant of the Western Australian endemic legume *Hardenbergia comptoniana* and from a plant of the naturalised weed species *Passiflora cearulea* was sequenced using Illumina technology. De novo assembly revealed the complete genome sequences of two distinct isolates of *Hardenbergia mosaic virus* (HarMV), an isolate of *Passionfruit woodiness virus*, and an undescribed virus with identity to members of the family Betaflexiviridae [40].

Deep sequencing approach has been used to characterize plant virus infection in grapevine from Italy, South Africa and USA, 2, 3, 11, 16. A number of grapevine-infecting viruses were discovered from long reads of dsRNA isolated from an individual plant, or from pooled tissue from a vineyard [2, 11]. Two novel DNA viruses belonging to the genus *Badnavirus* and the genus *Mastrevirus* were discovered. Deep sequencing analysis of RNAs from grapevines showing syrah decline symptoms using Life Sciences 454 high-throughput sequencing showed multiple viruses as well as viroid infection and a novel virus *Grapevine syrah virus 1* (GSyV-1) was discovered from grapevines in USA [2]. Next generation sequencing of dsRNAs of pooled samples from 44 grapevine plants, detected a number of grapevine-infecting viruses as well as putative fungal viruses [11]. Characterization of the virome in single grapevines by 454 high-throughput sequencing of double stranded RNA recovered from the vine stem was dominated by mycoviruses. The analysis revealed a substantial set of sequences similar to those of fungal viruses which represented half of all known mycoviral families including

the Chrysoviridae, Hypoviridae, Narnaviridae, Partitiviridae, and Totiviridae [3]. An analysis of small RNA populations from PG grapevines in the Trentino region (Italy) was carried out by Illumina high throughput sequencing. The study disclosed the virus and viroids contents of the two vines that was composed by *Grapevine rupestris stem pitting-associated virus* (GRSPaV), two viroids *Hop stunt viroid* (HSVd) and *Grapevine yellow speckle viroid 1* (GYSVd1), the Marafiviruses *Grapevine rupestris vein feathering virus* (GRVfV) and *GSyV-1*, and a hitherto unrecorded virus [16].

### Complete Genome Sequencing

Meta-analysis of existing small interfering RNAs (siRNA) libraries derived from drosophila and mosquito cell cultures helped the reconstruction of five previously undescribed RNA viruses, showing that in virus-infected insect cells, virus-homologous siRNAs are sufficiently abundant and overlap enough to allow their assembly into contiguous segments of sufficient length to be useful for diagnostic purposes [37]. This approach was extended to plant viruses in intact plants by Kreuze et al. [20] and successfully identified and reconstructed a *Potyvirus* and *Closterovirus* from viral small RNAs that had been experimentally introduced into sweet potato plants. Next-generation sequencing coupled with metagenomic analysis was used to produce large quantities of cDNA sequence in a model system of tomato infected with PepMV and 20.1 % of the total sequences were PepMV [1]. The extracted PepMV sequences were aligned against a reference genome and minimum of 200-fold sequencing coverage was achieved for the whole virus genome [1]. The complete genome sequence for an isolate of the Ugandan and Tanzanian strain types of *Cassava brown streak virus* from symptomatic cassava have been determined using the novel approach of non-directed NGS using GS-FLX Genome Sequencer (Advanced Genomics Facility, Liverpool University, UK) [25]. The complete genome sequence of a *Passion fruit woodiness virus* isolate from Australia was determined using deep sequencing on Illumina sequencing platform and the accuracy of the genome sequence information was 42-1691-fold sequence coverage and viral RNA accounted for 7.38 % of total polyadenylated RNA from the host plant [39]. Four suspected virus-infected tomato samples collected in the U.S. and Mexico were deep sequenced and after in silico subtraction of the tomato small RNAs, the remaining virus-like siRNA pools were assembled with or without reference virus or viroid genomes. A complete genome was obtained for *Potato spindle tuber viroid* (PSTVd) using siRNA alone and a near complete virus genome (98 %) was obtained for PepMV [21].

### Investigation on Viral Quasispecies

RNA viruses are notorious for the plasticity of their genomes due to the error prone nature of RNA dependent RNA polymerase and there occurs a wide variety of mutants that are present in virus populations, which are often referred to as quasispecies. Information on these sequence variants is critically relevant to know the viral evolution and spread, virulence, evasion of the host immune response etc. Hagen and associates [17] used de novo sequencing of virus-derived siRNAs to reconstruct and sequence two strains of *Tomato spotted wilt virus* that differ considerably in their infectivity characteristics, an unidentified *Tospovirus* infecting *S. lycopersicum*, and a squash-infecting DNA virus. Small RNA sequence processing, assembly, and virus and viroid genome identification was done from virus infection suspected tomato samples and a common mixed infection of two strains of PepMV (EU and US), which shared 82 % of genome nucleotide sequence identity, also could be differentially assembled into their respective genomes [21]. Using high-throughput sequencing, Fabre et al. [15] analyzed the population dynamics of four *Potato virus Y* (PVY) variants differing at most by two substitutions involved in pathogenicity properties by developing mathematical model that can accurately describe both selection and genetic drift processes shaping the evolutionary dynamics of viruses within their hosts.

### Study of Antiviral Defence Mechanism

Several recent studies by profiling of virus-specific small interfering RNAs (vsRNAs) using NGS platforms implicated the role of plant-encoded RNA dependent RNA polymerases (RDR) in vsRNA biogenesis and vsRNA-mediated antiviral defence [31]. The viral small RNA profile was analyzed in wild-type plants as well as mutants by applying small RNA deep sequencing technology using a crucifer-infecting strain of *Tobacco mosaic virus* (TMV-Cg) and *thaliana* as a model system. The study revealed an important role for host RDRs in viral siRNA biogenesis and for viral siRNA-mediated virus-host interactions in viral pathogenicity and host specificity [30]. Profiling of *Cymbidium ringspot virus* (CymRSV) derived short RNAs (vsRNAs) from the first systemically infected leaves of *Nicotiana benthamiana* using two different high-throughput sequencing platforms suggest that CymRSV siRNAs are produced from the structured positive strand rather than from perfect double stranded RNA or by RNA dependent RNA polymerase [41]. Donaire and associates [14] profiled the vsRNAs produced by eight other viruses, including *Cucumber mosaic virus*, *Potato virus X*, *Melon necrotic spot virus* (MNSV), *Tobacco rattle virus* (TRV), *Pepper mild mottle virus* (PMMoV), *Watermelon mosaic virus*,

*Turnip mosaic virus*, and *Tomato yellow leaf curl virus*, a DNA virus. They found that similar to CymRSV, TRV and PMMoV were associated with vsRNAs that were biased towards (+) strand origin, although the difference was less pronounced than for CymRSV.

High-throughput pyrosequencing and the profiling of the HSVd small RNAs (vd-sRNAs) recovered from the leaves and phloem of infected cucumber (*Cucumis sativus*) plants suggest that viroid-derived double-stranded RNA functions as one of the main precursors of vd-sRNAs and the existence of a selective trafficking of vd-sRNAs to the phloem tissue of infected cucumber plants [23]. Deep sequencing of the small RNAs derived from two symptomatic variants of a chloroplastic viroid and further analysis support the concept that viroids are elicitors and targets of the RNA silencing machinery of their hosts [13]. To gain insights on the genesis and possible role of vd-sRNAs in plant-viroid interaction, sRNAs isolated from *Vitis vinifera* infected by HSVd and GYSVd1 were sequenced by the high-throughput platform Illumina. Findings showed different Dicer-like enzymes target viroid RNAs, preferentially accessing to the same viroid domains which suggest that viroid RNAs may interact with host enzymes involved in the RNA-directed DNA methylation pathway, indicating more complex scenarios than previously thought for both vd-sRNAs genesis and possible interference with host gene expression [27]. Yan et al. [42] characterized siRNAs derived from *Rice stripe virus* (RSV), a member of the genus *Tenuivirus* through deep sequencing of small RNAs from infected rice leaves and showed that siRNAs were derived almost equally from virion and complementary RNA strands and were mostly 20–22 nucleotides long. High-resolution small RNA map for a monopartite begomovirus and its associated betasatellite was generated using Illumina-based deep sequencing and the results suggest that viral transcript might act as RDR substrates resulting in dsRNA and secondary siRNA production. In addition, the betasatellite affected the amount of vsRNAs detected in *S. lycopersicum* and *N. benthamiana* plants [43].

Small RNA (sRNA) deep-sequencing combined with transcriptome profiling determined the global impact of *Tobamovirus* infection on *Arabidopsis* sRNAs and their mRNA targets. Infection of *Arabidopsis* plants with *Oilseed rape mosaic virus* (*Tobamovirus*) causes a global size-specific enrichment of micro RNAs (miRNAs) and other phased siRNAs. The observed patterns of sRNA enrichment suggest that in addition to a role of the viral silencing suppressor, the stabilization of sRNAs might also occur through association with unknown host effector complexes induced upon infection [18]. Small RNA analysis by Illumina sequencing from a non-coding region of *Cauliflower mosaic virus* (CaMV) in *Arabidopsis* indicated the

mechanism of CaMV small RNAs as decoy RNAs to evade silencing machinery from viral promoter and coding regions [6]. Silva et al. [35] attempted the characterization of a population of vsRNAs from cotton plants infected with *Cotton leafroll dwarf virus*, a member of the genus *Polemovirus*, family Luteoviridae and results suggest that virus-derived double-stranded RNA functions as one of the main precursors of vsRNAs.

## Concluding Remarks

Next generation sequencing based identification of viral genome depends heavily on the softwares used for comparative analysis, sequence informations and reference genomes. However such sequence based identification may be erroneous due to incomplete and inaccurate data accumulation in the reference databases [19]. Improved algorithms of sequence analysis can reduce such errors. Next generation sequencing is a beneficial tool to detect and diagnose virus and virus like and other pathogens present in a sample and has been successfully demonstrated particularly when prior knowledge of the causal agent of the disease is not known. It will unfold many mysteries of virus-host interaction. The etiology of several nationally important diseases such as urdbean leaf crinkle, pigeonpea sterility, mango malformation, tapping panel dryness of rubber, coconut root wilt can be established using high throughput sequencing technology. It will also be helpful in studying the spectrum of pathogens in economically important crops. Viral derived small RNA sequencing is emerging as a promising strategy for exploration of viruses and this can be a valuable part of the virologist's tools for diagnosis and discovery of new viruses in plants and other organisms. This technology can be useful for everyday diagnostics of plants, animals and human beings, once it becomes cheaper and widely available.

## References

1. Adams IP, Glover RH, Monger WA, Mumford R, Jackeviciene E, Navalinskiene M, Samuitiene M, Boonham N. Next-generation sequencing and metagenomic analysis: a universal diagnostic tool in plant virology. *Mol Plant Pathol*. 2009;10(4):537–45.
2. Al Rwahnih M, Daubert S, Golino D, Rowhani A. Deep sequencing analysis of RNAs from a grapevine showing Syrah decline symptoms reveals a multiple virus infection that includes a novel virus. *Virology*. 2009;387(2):395–401.
3. Al Rwahnih M, Daubert S, Úrbez-Torres JR, Cordero F, Rowhani A. Deep sequencing evidence from single grapevine plants reveals a virome dominated by mycoviruses. *Arch Virol*. 2011;156:397–403.
4. Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P. Emerging infectious diseases of plants: pathogen

- pollution, climate change and agrotechnology drivers. *Trends Ecol Evol*. 2004;10:535–44.
5. Barzon L, Lavezzo E, Militello V, Toppo S, Palù G. Applications of next-generation sequencing technologies to diagnostic virology. *Int J Mol Sci*. 2011;12:7861–84.
  6. Blevins T, Rajeswaran R, Aregger A, Borah BK, Schepetilnikov M, Baerlocher L, Hohn T, Pooggin MM. Massive production of small RNAs from a non-coding region of cauliflower mosaic virus in plant defense and viral counter-defense. *Nucleic Acids Res*. 2011;39:5003–14.
  7. Boonham N, Glover R, Tomlinson J, Mumford R. Exploiting generic platform technologies for the detection and identification of plant pathogens. *Eur J Plant Pathol*. 2008;121:355–63.
  8. Candresse T, Cambra M, Dallot S, Lanneau M, Asensio M, Gorris MT, Revers F, Macquaire G, Olmos A, Boscia D, Quiot JB, Dunez J. Comparison of monoclonal antibodies and PCR assays for the typing of isolates belonging to the D and M serotypes of plum pox virus. *Phytopathology*. 1998;88:198–204.
  9. Capobianchi MR, Giombini E, Rozera G. Next-generation sequencing technology in clinical virology. *Clin Microbiol Infect*. 2013;19:15–22.
  10. Clark MF, Adams AN. Characteristics of the microplate method of enzyme-linked immunosorbent assay for the detection of plant viruses. *J Gen Virol*. 1977;34:475–83.
  11. Coetzee B, Freeborough MJ, Maree HJ, Celton JM, Rees DJ, Burger JT. Deep sequencing analysis of viruses infecting grapevines: virome of a vineyard. *Virology*. 2010;400(2):157–63.
  12. Cox-Foster DL, Conlan S, Holmes EC, Palacios G, Evans JD. A metagenomic survey of microbes in honey bee colony collapse disorder. *Science*. 2007;318:283–7.
  13. Di Serio F, Gisel A, Navarro B, Delgado S, Martínez De Alba AE. Deep sequencing of the small RNAs derived from two symptomatic variants of a chloroplastic viroid: implications for their genesis and for pathogenesis. *PLoS One*. 2009;4(10):e7539. doi:10.1371/journal.pone.0007539.
  14. Donaire L, Wang Y, Gonzalez-Ibeas D, Mayer KF, Aranda MA. Deep-sequencing of plant viral small RNAs reveals effective and widespread targeting of viral genomes. *Virology*. 2009;392:203–14.
  15. Fabre F, Montarry J, Coville J, Senoussi R, Simon V. Modelling the evolutionary dynamics of viruses within their hosts: a case study using high-throughput sequencing. *PLoS Pathol*. 2012;8:e1002654. doi:10.1371/journal.ppat.1002654.
  16. Giampetruzzi A, Roumia V, Robertoa R, Malossinib U, Yoshikawac N, Nottea PL, Terlizzi F, Credid R, Saldarelli P. A new grapevine virus discovered by deep sequencing of virus- and viroid-derived small RNAs in Cv Pinot gris. *Virus Res*. 2011;163:262–8.
  17. Hagen C, Frizzi A, Kao J, Jia L, Huang M, Zhang Y, Huang S. Using small RNA sequences to diagnose, sequence, and investigate the infectivity characteristics of vegetable-infecting viruses. *Arch Virol*. 2011;156:1209–16.
  18. Hu Q, Hollunder J, Niehl A, Kørner CJ, Gereige D. Specific impact of *Tobamovirus* infection on the *Arabidopsis* small RNA profile. *PLoS One*. 2011;6(5):e19549. doi:10.1371/journal.pone.0019549.
  19. Kang S, Mansfield MA, Park B, Geiser DM, Ivors KL, et al. The promise and pitfalls of sequence based identification of plant-pathogenic fungi and oomycetes. *Phytopathology*. 2010;1:732–7.
  20. Kreuze JF, Perez A, Untiveros M, Quispe D, Fuentes S. 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*. 2009;388:1–7.
  21. Li R, Gao S, Hernandez AG, Wechter WP, Fei Z. Deep sequencing of small RNAs in tomato for virus and viroid identification and strain differentiation. *PLoS One*. 2012;7(5):e37127. doi:10.1371/journal.pone.0037127.
  22. Martin RM, James D, Levesque CA. Impacts of molecular diagnostic technologies on plant disease management. *Ann Rev Phytopathol*. 2000;38:207–39.
  23. Martinez G, Donaire L, Llave C, Pallas V, Gomez G. High-throughput sequencing of hop stunt viroid-derived small RNAs from cucumber leaves and phloem. *Mol Plant Pathol*. 2010;11:347–59.
  24. Matthews REF. *Plant Virology*. 3rd ed. Academic Press 1991. ISBN 0-12-480533-1.
  25. Monger WA, Alicai T, Ndunguru J, Kinyua ZM, Potts M, Reeder RH, Miano DW, Adams IP, Boonham N, Glover RH, Smith J. The complete genome sequence of the Tanzanian strain of *Cassava brown streak virus* and comparison with the Ugandan strain sequence. *Arch Virol*. 2010;155:429–33.
  26. Mumford R, Boonham N, Tomlinson J, Barker I. Advances in molecular phytodiagnostics—new solutions for old problems. *European J Plant Pathol*. 2006;116:1–19.
  27. Navarro B, Pantaleo V, Gisel A, Moxon S, Dalmay T. Deep sequencing of viroid-derived small RNAs from grapevine provides new insights on the role of RNA silencing in plant-viroid interaction. *PLoS One*. 2009;4(11):e7686. doi:10.1371/journal.pone.0007686.
  28. Ng TFF, Duffy S, Polston JE, Bixby E, Vallad GE. Exploring the diversity of plant DNA viruses and their satellites using vector-enabled metagenomics on whiteflies. *PLoS One*. 2011;6(4):e19050. doi:10.1371/journal.pone.0019050.
  29. Pallett DW, Ho T, Cooper I, Wang H. 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*. 2010;168:223–7.
  30. Qi X, Bao FS, Xie Z. Small RNA deep sequencing reveals role for *Arabidopsis thaliana* RNA-dependent RNA polymerases in viral siRNA biogenesis. *PLoS One*. 2009;4(3):e4971. doi:10.1371/journal.pone.0004971.
  31. Qu F. Plant viruses versus RNAi: simple pathogens reveal complex insights on plant antimicrobial defense. *Wiley Interdiscip Rev RNA*. 2010;1:22–33.
  32. Quan PL, Briese T, Palacios G, Lipkin WI. Rapid sequence-based diagnosis of viral infection. *Antiviral Res*. 2008;79:1–5.
  33. Roossinck MJ, Saha P, Wiley GB, Quan J, White JD, Lai H, Chavarr'a F, Shen G, Roe BA. Ecogenomics: using massively parallel pyrosequencing to understand virus ecology. *Mol Ecol*. 2010;19(Suppl 1):81–8.
  34. Schadt EE, Turner S, Kasarskis A. A window into third-generation sequencing. *Hum Mol Genet*. 2010;19:R227–40.
  35. Silva TF, Romanel EAC, Andrade RRS, Farinelli L. Profile of small interfering RNAs from cotton plants infected with the *Polerovirus*, *cotton leafroll dwarf virus*. *BMC Mol Biol*. 2011;12:40.
  36. Studholme DJ, Glover RH, Boonham N. Application of high-throughput DNA sequencing in phytopathology. *Ann Rev Phytopathol*. 2011;49:87–105.
  37. Wu Q, Luo Y, Lu R, Lau N, Lai EC, Li WX, Ding SW. Virus discovery by deep sequencing and assembly of virus derived small silencing RNAs. *Proc Natl Acad Sci USA*. 2010;107(4):1606–11.
  38. Wylie SJ, Jones MGK. 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*. 2011;156:479–82.
  39. Wylie SJ, Jones MGK. Deep sequencing Australian native plant viruses. Oral presentation 1.6. in Abstract book—9th Australasian Plant Virology Workshop (APVW), 16–19 November 2010, Melbourne Australia p 6.
  40. Wylie SJ, Luo H, Li H, Jones MGK. Multiple polyadenylated RNA viruses detected in pooled cultivated and wild plant samples. *Arch Virol*. 2012;157:271–84.



41. Szittyá G, Moxon S, Pantaleo V, Toth G, Rusholme Pilcher RL. Structural and functional analysis of Viral siRNAs. *PLoS Pathogen*. 2010;6:e1000838. doi:[10.1371/journal.ppat.1000838](https://doi.org/10.1371/journal.ppat.1000838).
42. Yan F, Zhang H, Adams MJ, Yang J, Peng J, Antoniw AF, Zhou Y, Chen J. Characterization of siRNAs derived from rice stripe virus in infected rice plants by deep sequencing. *Arch Virol*. 2010;155:935–40. doi:[10.1007/s00705-010-0670-8](https://doi.org/10.1007/s00705-010-0670-8).
43. Yang X, Wang Y, Guo W, Xie Y, Xie Q. Characterization of small interfering RNAs derived from the geminivirus/betasatellite complex using deep sequencing. *PLoS One*. 2011;6(2):e16928. doi:[10.1371/journal.pone.0016928](https://doi.org/10.1371/journal.pone.0016928).
44. Zheng L, Wayper PJ, Gibbs AJ, Fourment M, Rodoni BC, Gibbs MJ. Accumulating variation at conserved sites in potyvirus genomes is driven by species discovery and affects degenerate primer design. *PLoS One*. 2008;3:e1586. doi:[10.1371/journal.pone.0001586](https://doi.org/10.1371/journal.pone.0001586).