Available online at www.agrifarming.org

DOI: http://dx.doi.org/10.18782/2582-7146.344

ISSN: 2582 – 7146

Curr. Rese. Agri. Far. (2026) *7*(4), 1-14

Current Research in Agriculture and Farming

Review Article

Peer-Reviewed, Refereed, Open Access Journal

Next Generation Sequencing (NGS): A Breakthrough Technique for Plant Virus Detection

Swadhin Kumar Swain[1*], Dhanashree Dhanaji Shinde[2], Rahul Soni[3], Vijay Krishna[4], Pranghita Das[5] and Rishav Kumar[6]

[1]Research Scholar, Department of Nematology, College of Agriculture, OUAT, Bhubaneswar, Odisha

[2]Highly Skilled Researcher, Department of Biotechnology and Bioinformatics,
ICAR- Directorate of Onion and Garlic Research, Pune, Maharashtra

[3]Research Scholar, Department of Botany, University of Allahabad, Uttar Pradesh

[4]B.Sc. Scholar, ICAR- IARI, Jharkhand

[5]Research Scholar, Department of Genetics and Plant Breeding,
Uttar Banga Krishi Vishwavidyalaya, Pundibari, Coochbehar

[6]M.Sc., College of Agriculture, Central Agricultural University, Iroisemba, Imphal – 795004, Manipur

*Corresponding Author E-mail: rkswadhinkumar@gmail.com

Received: 2.06.2026 | Revised: 28.07.2026 | Accepted: 13.08.2026

ABSTRACT

Next Generation Sequencing (NGS) has emerged as a powerful and high-throughput technology for the rapid and comprehensive detection of plant viruses. Unlike conventional diagnostic methods that require prior knowledge of viral pathogens, NGS enables the identification of both known and novel viruses directly from infected plant tissues. By generating millions of sequence reads, NGS facilitates viral genome reconstruction, characterization of mixed infections, and detection of divergent or emerging viral strains. Its application has significantly improved plant virus surveillance, epidemiological studies, and disease management. Integration of NGS with advanced bioinformatics tools further enhances sequence analysis, pathogen identification, and phylogenetic investigation. Despite challenges related to cost, data processing, and standardization, NGS represents a transformative approach for early and accurate plant virus diagnosis and offers substantial opportunities for sustainable crop protection and biosecurity.

Keywords: *Next Generation Sequencing, Plant Viruses, Virus Detection, Metagenomics, Bioinformatics*

INTRODUCTION

Plant viruses are among the major constraints in agriculture, causing severe losses in crop yield, quality, and market value worldwide (Rubio et al., 2020; & Hull, 2014). Unlike fungi or bacteria, plant viruses cannot be cultured easily in artificial media, and many infections remain latent or symptomless until damage becomes significant (Hull, 2014; & Mehetre et al., 2021).

Cite this article: Swain, S. K., Shinde, D. D., Soni, R., Krishna, V., Das, P., & Kumar, R. (2026). Next Generation Sequencing (NGS): A Breakthrough Technique for Plant Virus Detection , *Curr. Rese. Agri. Far. 7*(4), 1-14. doi: http://dx.doi.org/10.18782/2582-7146.344

Conventional diagnostic methods such as ELISA, PCR, RT-PCR, and electron microscopy have been widely used for virus detection; however, these methods often require prior knowledge of the pathogen and are usually limited to detecting one or a few targeted viruses at a time (Mehetre et al., 2021; & Wang et al., 2022). This creates challenges when dealing with mixed infections, emerging viruses, or completely unknown pathogens (Adams & Fox, 2016; & Massart et al., 2017).

The emergence of Next-Generation Sequencing (NGS) has transformed plant virology by enabling rapid, high-throughput, and unbiased detection of viral nucleic acids directly from infected plant tissues (Massart et al., 2017; & Adams & Fox, 2016). NGS refers to a group of advanced sequencing technologies capable of simultaneously reading millions of DNA or RNA fragments in parallel, generating massive amounts of sequence data in a single run (Metzker, 2010; & Goodwin et al., 2016). Because of this massively parallel nature, NGS is also known as high-throughput sequencing (HTS) or deep sequencing (Goodwin et al., 2016).

In plant virus detection, NGS offers a non-targeted diagnostic approach, the prior information about the virus are not essential (Massart et al., 2017; & Mehetre et al., 2021). This is especially valuable because many plant viruses exhibit extensive genetic diversity, rapid mutation rates, and frequent recombination events (Hull, 2014; & García-Arenal et al., 2001). Next-generation sequencing (NGS) approaches can provide us with a snapshot of the virome present in the propagule as they are effective not only in detecting the known viral pathogens and their variants, if any, but also in unravelling unknown one(s) (Jo et al., 2018). Among the various NGS based approaches, sRNA (sRNAome) and mRNA (mRNAome) sequencing are commonly used to reveal the virome of a given sample (Jones et al., 2017; & Massart et al., 2019). Recently, a few studies attempted to reveal the virome of different crops like grapevine, apple, and pepper from publically available mRNAome data (Jo et al., 2017). The ability to recover full genome sequences is particularly important for taxonomic classification, evolutionary studies, and development of downstream diagnostic assays (Adams et al., 2018; & Massart et al., 2017).

Both sRNA and mRNA pools can effectively capture single as well as double-stranded RNA viruses and some DNA viruses (Seguin et al., 2014; Roossinck et al., 2015; & Jo et al., 2017). However, the relatively lower representation of viral RNA in the background of total plant RNA limits the use of mRNAome compared to sRNAome for viral detection (Beuve et al., 2018; & Maliogka et al., 2018). As mRNA-based methods can give longer contigs, they are more useful for variant detection, especially when significant genetic diversity exists in some of the viruses (Xiao et al., 2019). Thus, it would be worthwhile to study the virome using both these methods for robust identification of the entire virome of a plant species. In 2009 is widely recognized as the starting point for NGS in plant virology (Adams et al., 2009).

Next Generation Sequencing (NGS)

Plant virus identification has been transformed by the evolution of sequencing technologies, moving from specific, targeted methods to unbiased, high-throughput approaches (Metzker, 2010; & Goodwin et al., 2016). The transition from First-Generation (Sanger) to Second Generation (Next Generation Sequencing/NGS) has allowed researchers to move beyond detecting only known threats to uncovering the entire "virome" of a host plant (Adams et al., 2009; & Roossinck et al., 2015).

First-Generation Sequencing (Sanger Sequencing)

First-generation sequencing is synonymous with the Sanger dideoxy method, developed in

1977 (Sanger et al., 1977; & Metzker, 2010). This technique utilizes chain-terminating inhibitors (ddNTPs) to produce DNA fragments of varying lengths, which are then separated by size via capillary electrophoresis (Sanger et al., 1977; & Heather & Chain, 2016). In plant virology, Sanger sequencing is primarily a "targeted" approach; it requires prior knowledge of the virus to design specific primers for PCR amplification (Adams et al., 2009; & Kchouk et al., 2017). Consequently, it is excellent for confirming the identity of a known virus or verifying the results of high-throughput screens but is ineffective for discovering novel or highly divergent viruses that do not match existing primers (Kchouk et al., 2017; & Goodwin et al., 2016).

The primary strength of Sanger sequencing lies in its high read accuracy (up to 99.99%) and relatively long individual read lengths (700–1,000 bp) (Heather & Chain, 2016; & Goodwin et al., 2016). However, it is limited by extremely low throughput, as it sequences only one DNA fragment per reaction (Metzker, 2010; & Kchouk et al., 2017). For complex plant samples containing multiple viral strains or low-titer infections, Sanger sequencing often fails to capture the full genetic diversity (Adams et al., 2009; & Roossinck et al., 2015). It remains a staple for routine diagnostics of well-characterized pathogens where speed and low cost for a single sample are prioritized over depth of information (Al-Kaethy, 2025).

Second-Generation Sequencing

Second-generation sequencing, or Next-Generation Sequencing (NGS), represents a paradigm shift toward massively parallel sequencing (Metzker, 2010; & Goodwin et al., 2016). Platforms such as Illumina and Ion Torrent can sequence millions of small DNA fragments simultaneously (Goodwin et al., 2016; & Kchouk et al., 2017). This allows for an "unbiased" or "shotgun" approach, where all nucleic acids (host and viral) are sequenced without the need for virus-specific primers (Adams et al., 2009; & Massart et al., 2017). This capability has led to the discovery of hundreds of novel plant viruses that were previously undetectable by traditional methods (Villamor et al., 2019).

The most widely used second-generation platform is Illumina, which uses "sequencing-by-synthesis" (Bentley et al., 2008; & Goodwin et al., 2016). DNA fragments are attached to a flow cell, amplified into clusters, and sequenced using reversible fluorescent terminators (Bentley et al., 2008). While it produces shorter reads (typically 150–300 bp) compared to Sanger, the sheer volume of data allows bioinformatics tools to assemble these "short reads" into full viral genomes (Metzker, 2010; & Goodwin et al., 2016). Another platform, Ion Torrent, detects hydrogen ions released during DNA polymerization, offering a faster workflow but facing challenges with repetitive sequences (Rothberg et al., 2011; & Sharma et al., 2024).

NGS is the gold standard for metagenomics and virome analysis (Massart et al., 2017; & Roossinck et al., 2015). By sequencing total RNA or small RNAs (sRNAs) produced by the plant's RNA interference (RNAi) defense system, researchers can identify every virus present in a single sample (Kreuze et al., 2009; & Jones et al., 2017). This is crucial for studying "complex diseases" caused by multiple co-infecting viruses. Additionally, NGS is instrumental in studying viral evolution and quasispecies, as the deep coverage can detect rare genetic variants that make up only 1% of the viral population (Roossinck et al., 2015; Sanjuán & Domingo-Calap, 2016; & Hamza & Abd-Ulridha, 2026).

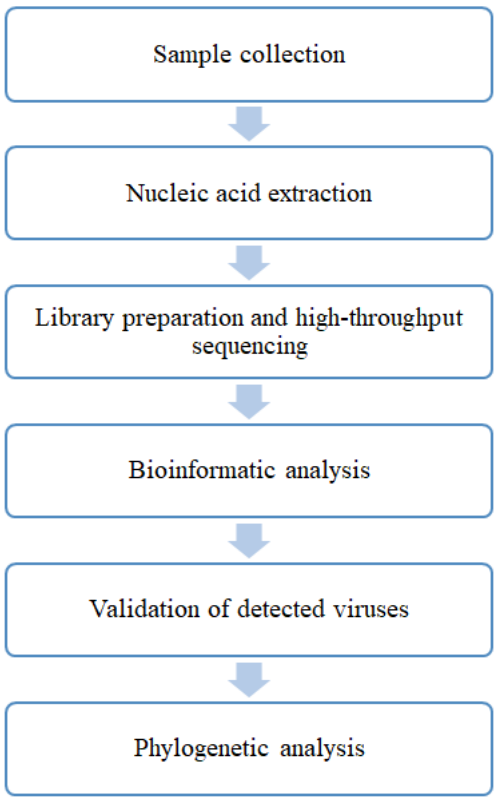

Fig. 1: Procedure of NGS

Table1. Current major methods of next-generation sequencing technologies

Sequencing Platform	**Amplification Method**	**Sequencing Chemistry**	**Read Length (bp)**	**Sequencing Speed/h**	**Maximum Output per Run**	**Accuracy (%)**
454 (Roche)	Emulsion PCR	Pyrosequencing	400–700	13 Mbp	700 Mbp	99.9
Illumina (Illumina)	Bridge PCR	Reversible terminators	100–300	25 Mbp	600 Gbp	99.9
SOLiD (Life Technologies)	Emulsion PCR	Ligation	75–85	21–28 Mbp	80–360 Gbp	99.9
PacBio (Pacific Biosciences)	No amplification (Single molecule real-time, SMRT)	Fluorescently labeled nucleotides	4,000–5,000	50–115 Mbp	200 Mb–1 Gbp	95
Helicos (Helicos Biosciences)	No amplification (Single molecule)	Reversible terminators	25–55	83 Mbp	35 Gbp	97
Ion Torrent (Life Technologies)	Emulsion PCR	Detection of released H^+	100–400	25 Mb–16 Gbp	100 Mb–64 Gbp	99
Nanopore (Oxford Technologies)	No amplification (Single molecule)	Nanopore sequencing	Very long reads (up to 50 kbp)	150 Mbp	Tens of Gbp	96

Illumina Sequencing Workflow

Sample Collection

The first step in using Illumina for plant virus detection is sample collection. Although sequencing happens later inside the machine, the quality of the final results depends heavily on how the plant samples are selected, handled, preserved, and prepared before nucleic acid extraction (Massart et al., 2017; & Kumar et al., 2020). Poor sampling can lead to low viral RNA concentration, contamination, or degradation, which may cause false negatives or incomplete viral genome recovery (Roossinck et al., 2015; & Jones et al., 2017). Therefore, sample collection is considered the foundation of successful NGS-based plant virus diagnosis (Massart et al., 2017).

In plant virology, researchers usually collect symptomatic tissues such as leaves, stems, petioles, fruits, or flowers showing signs like mosaic, chlorosis, curling, necrosis, stunting, or deformation (Hull, 2014; & Agrios, 2005). These tissues are more likely to contain higher viral titers than healthy-looking tissues (Hull, 2014). However, in some studies, asymptomatic plants are also sampled to detect latent infections or early-stage disease (Roossinck et al., 2015). For systemic viruses, young leaves are often preferred because they actively transport nutrients and can accumulate viral particles efficiently (Hull, 2014; & Wang et al., 2022). Selecting the correct tissue type increases the probability of detecting the pathogen during sequencing.

For robust scientific analysis, biological replicates are often collected (Kumar et al., 2020; & Massart et al., 2017). This means taking samples from multiple plants of the same cultivar, treatment, or field condition. Replicates improve reliability and help confirm that observed viral sequences are truly associated with the disease and not random contamination (Jones et al., 2017). In plant virus research, replication is especially important when comparing cultivars, geographic regions, or symptom severity levels (Massart et al., 2017).

Immediately after collection, plant tissues must be preserved to prevent RNA degradation (Rio et al., 2010; & Kukurba & Montgomery, 2015). Since RNA is unstable and rapidly broken down by RNases, samples are commonly frozen in liquid nitrogen and stored at −80°C until extraction (Kukurba & Montgomery, 2015). Alternatively, silica gel drying or RNA stabilization solutions may be used for field collections where freezing is not possible (Wang et al., 2022). Maintaining RNA integrity is essential because degraded RNA reduces library quality and sequencing efficiency on Illumina platforms (Kukurba & Montgomery, 2015; & Goodwin et al., 2016). High-quality RNA leads to better cluster generation and accurate sequencing reads.

The amount of tissue collected also matters. Typically, 50–200 mg of fresh tissue is sufficient for RNA extraction in plant virus studies, though this depends on the extraction kit and downstream library preparation protocol (Kumar et al., 2020). Researchers avoid overly old, damaged, or heavily necrotic tissues because these may contain degraded nucleic acids and inhibitory compounds such as phenolics or polysaccharides (Agrios, 2005). Clean scissors, gloves, and sterile tubes are used to minimize cross-contamination between samples (Massart et al., 2017). Labeling each sample carefully is essential for traceability during sequencing and bioinformatics analysis.

For NGS-based virus detection, the goal is not to isolate the virus alone but to collect plant material containing viral nucleic acids (Adams et al., 2009; & Massart et al., 2017). Later, total RNA is extracted from this tissue. During extraction, both plant RNA and viral RNA are recovered together (Kreuze et al., 2009). The viral fraction is then enriched indirectly through methods such as rRNA depletion, poly-A selection, or small RNA isolation, depending on the study design (Zhao et al., 2015; & Kukurba & Montgomery, 2015). This makes careful sample collection even more important, because the viral RNA is usually only a small fraction of the total RNA pool (Roossinck et al., 2015; & Massart et al., 2017).

Cluster generation

After sample collection, RNA extraction, enrichment, and library preparation, the second major technical step in Illumina is cluster generation. This step is essential because the sequencer cannot accurately detect signals from a single DNA molecule alone. Instead, each DNA fragment from the prepared library must be copied many times in one small location on the flow cell to create a detectable fluorescent signal (Bentley et al., 2008; & Goodwin et al., 2016). These localized copies are called clusters. Without cluster generation, sequencing-by-synthesis would not produce strong enough signals for reliable base calling (Metzker, 2010; & Heather & Chain, 2016).

In plant virus detection, the sequencing library contains a mixture of plant-derived and virus-derived DNA fragments (usually cDNA prepared from RNA) (Adams et al., 2009; & Massart et al., 2017). Each fragment has adapters attached during library preparation. These adapters are designed to match complementary oligonucleotides fixed on the surface of the Illumina flow cell (Bentley et al., 2008). When the library is loaded into the NextSeq 500, the fragments hybridize to these oligos, anchoring themselves to the glass surface. This marks the beginning of cluster generation.

The amplification process used is commonly called bridge amplification (Bentley et al., 2008; & Metzker, 2010). First, a single-stranded DNA fragment binds to one oligo on the flow cell. Then, because the fragment is flexible, its free end bends over and hybridizes to a nearby complementary oligo, forming a bridge-like structure. DNA polymerase then synthesizes the complementary strand, creating a double-stranded bridge. The strands are then denatured, leaving two attached single strands. Repeating this cycle many times creates thousands of identical copies of the original fragment clustered together in one spot (Goodwin et al., 2016).

This amplification is called clonal amplification because all molecules in a cluster originate from one original DNA fragment (Metzker, 2010). In simple terms, it is like making thousands of photocopies of one page and stacking them in the same place. When sequencing occurs, the machine reads the fluorescence emitted from all copies in the cluster at once. Since all copies ideally contain the same base at each cycle, the signal becomes strong and easy to interpret. This greatly improves sequencing accuracy (Heather & Chain, 2016; & Goodwin et al., 2016).

For plant virus studies, cluster generation is especially important because viral nucleic acids are often present in very low abundance compared to host plant RNA (Roossinck et al., 2015; & Massart et al., 2017). By generating clusters, even rare viral fragments become detectable during sequencing. However, optimal cluster density is critical. Too few clusters reduce sequencing output, while too many clusters cause overlapping signals and poor data quality (Illumina, 2016; & Goodwin et al., 2016). Illumina provides cluster optimization guidelines to ensure balanced loading concentrations for accurate sequencing.

In the Illumina, cluster generation occurs directly on the flow cell inside the cartridge-based system. This reduces manual handling and improves consistency (Illumina, 2016). Once clusters are formed, the system proceeds to imaging and sequencing-by-synthesis. During the first cycles, the instrument maps cluster positions this is called template generation. The coordinates of every cluster are recorded so that fluorescence can be tracked throughout all sequencing cycles (Bentley et al., 2008).

A major challenge during cluster generation is phasing and prephasing. These occur when some DNA molecules in a cluster fall behind or move ahead during synthesis cycles (Metzker, 2010; & Goodwin et al., 2016). Over time, this reduces synchronization within the cluster and weakens signal clarity. Illumina's software performs phasing correction computationally to maintain sequencing accuracy. This is one reason why Illumina short-read platforms are highly reliable for virus genome assembly and diagnostics.

In practical plant virus identification workflows, successful cluster generation ensures that viral reads are sufficiently amplified and accurately detected, enabling downstream assembly using tools such as Trinity, SPAdes, or CLC Genomics Workbench (Grabherr et al., 2011; & Bankevich et al., 2012). Thus, cluster generation is the bridge between molecular preparation and digital sequence data. It transforms invisible DNA fragments into readable signals that drive virus discovery and diagnostics.

Sequencing:

The third major step in the Illumina workflow is the actual sequencing process, where the amplified DNA clusters on the flow cell are read base by base (Bentley et al., 2008; & Metzker, 2010). This is the stage where biological molecules are converted into digital sequence data. In plant virus detection, this step is crucial because it generates the short reads that will later be analyzed to identify viral genomes, detect mixed infections, and discover novel viruses (Adams et al., 2009; & Massart et al., 2017). The NextSeq 500 uses Sequencing by Synthesis (SBS) technology, Illumina's core chemistry for high-throughput DNA sequencing (Bentley et al., 2008; & Goodwin et al., 2016).

In plant virus studies, the material entering the sequencing step is usually a cDNA library prepared from extracted RNA (viral + plant RNA) (Kreuze et al., 2009; & Jones et al., 2017). Each fragment in the library has platform-specific adapters attached. After cluster generation, each cluster contains thousands of identical copies of one original fragment. The sequencer reads all clusters simultaneously in a massively parallel manner, allowing millions of fragments to be analyzed in one run (Goodwin et al., 2016). This high-throughput nature makes NextSeq 500 highly effective for metagenomics and transcriptomics-based virus identification (Massart et al., 2017; & Villamor et al., 2019).

The sequencing chemistry relies on reversible terminator nucleotides (Bentley et al., 2008; & Metzker, 2010). During each cycle, four fluorescently labeled nucleotides (A, T, G, C) are added together. Each nucleotide has a temporary blocking group that allows only one base to be incorporated at a time. When DNA polymerase adds the correct nucleotide to the growing strand, synthesis stops temporarily because of the blocker. The instrument then captures an image of the fluorescence emitted by that nucleotide. The color identifies which base was added. After imaging, the blocking group and fluorescent dye are chemically removed, allowing the next cycle to begin. This process repeats hundreds of times (Bentley et al., 2008).

This method ensures controlled, one-base-at-a-time sequencing, which improves accuracy (Heather & Chain, 2016; & Goodwin et al., 2016). In simple terms, imagine reading a sentence one letter at a time, taking a photo after each letter, and then removing the marker before moving to the next. That is essentially how sequencing-by-synthesis works. Because each cluster emits a measurable signal, the system can accurately determine the DNA sequence for millions of fragments in parallel (Metzker, 2010).

For plant virus detection, paired-end sequencing is often used (Goodwin et al., 2016; & Villamor et al., 2019). A common setup is 2 × 150 bp, meaning the machine reads 150 bases from one end of the fragment and then 150 bases from the opposite end. This doubles the information obtained from each fragment and improves genome assembly, especially for unknown or highly variable viruses (Bankevich et al., 2012). Paired-end reads are particularly useful in de novo assembly because overlapping information from both ends helps reconstruct longer viral contigs.

The NextSeq 500 also performs index sequencing between reads. Index reads are short barcode sequences added during library preparation (Illumina, 2016). These barcodes allow multiple samples to be pooled in one sequencing run a process called multiplexing (Goodwin et al., 2016). After sequencing, the software uses these indices to assign reads back to their original samples, a process called demultiplexing. In plant virus projects, this reduces cost and increases throughput when analyzing multiple cultivars, treatments, or

locations simultaneously (Massart et al., 2017).

During the run, t+he instrument uses Real-Time Analysis (RTA) software for image processing and base calling (Illumina, 2016). Base calling is the computational step where fluorescence signals are translated into nucleotide letters. The software also calculates quality scores (Q-scores), which estimate confidence in each base call (Ewing & Green, 1998; & Goodwin et al., 2016). High Q-scores indicate reliable sequencing data. These metrics are important in plant virology because low-quality reads can interfere with virus assembly and lead to false identifications (Jones et al., 2017).

At the end of sequencing, the output is typically stored in FASTQ files, which contain both the nucleotide sequence and its quality scores (Cock et al., 2010). These files are then used for downstream analysis such as trimming, host-read removal, de novo assembly, and viral database alignment (Grabherr et al., 2011; & Bankevich et al., 2012). In plant virus detection, the FASTQ data form the raw foundation for identifying pathogens. Without this step, there would be no digital evidence of viral presence.

Thus, the sequencing step in Illumina transforms amplified DNA clusters into millions of readable sequence fragments. For plant virus detection, this step provides the depth, accuracy, and scalability needed to detect even low-abundance viral genomes in complex plant samples (Roossinck et al., 2015; & Massart et al., 2017). It is the core engine of NGS-based diagnostics.

Data Analysis

The fourth and final step in the Illumina workflow is data analysis. This is where raw sequencing signals are transformed into biologically meaningful information specifically, the identification of plant viruses present in the sample (Conesa et al., 2016; & Massart et al., 2017). In simple terms, the sequencing machine generates millions of short DNA reads, but these reads are only useful after computational processing. Data analysis converts these fragments into viral genome sequences, taxonomic identifications, and diagnostic conclusions (Knyazev et al., 2021). For plant virus detection, this step is just as important as sequencing itself.

The first stage of analysis begins inside the instrument through Real-Time Analysis (RTA2) software. During sequencing, the Illumina captures images of fluorescent signals emitted from each cluster on the flow cell (Bentley et al., 2008). RTA2 extracts signal intensities, performs base calling (assigning A, T, G, or C), and calculates quality scores (Q-scores) for each base (Ewing & Green, 1998). These results are stored in BCL (base call) files (Illumina, 2016). This on-instrument processing happens automatically while the run is ongoing. Without this step, the fluorescence images would remain uninterpreted raw data.

After sequencing is complete, the BCL files must be converted into FASTQ files, which are the standard format for downstream bioinformatics (Cock et al., 2010). If multiple samples were pooled together using barcodes, the first task is demultiplexing sorting reads into their original samples based on index sequences (Kircher et al., 2012). Tools such as bcl2fastq, BCL Convert, or BaseSpace Sequence Hub perform this conversion. FASTQ files contain both the sequence and the associated quality values for every read. In plant virus studies, these FASTQ files are the starting point for all downstream analysis.

The next step is quality control (QC). Raw reads often contain low-quality bases, sequencing adapters, and technical artifacts (Andrews, 2010). Tools such as FastQC are used to assess read quality, while Trimmomatic or Cutadapt remove adapters and poor-quality regions (Bolger et al., 2014; & Martin, 2011). This cleaning step ensures that only reliable reads are used for assembly and virus detection. In plant virology, poor-quality reads can create false contigs or reduce detection sensitivity, so QC is essential.

Once reads are cleaned, host-derived sequences (plant RNA/DNA) are often removed because they usually represent the majority of reads. This is done by mapping reads against the host reference genome or transcriptome using aligners such as Bowtie 2 (Langmead & Salzberg, 2012). Reads that do

not align to the plant genome are retained as potential viral sequences. This filtering step enriches the dataset for pathogen discovery and reduces computational complexity (Kutnjak et al., 2021). In plant virus detection, host subtraction is critical because viral reads may be only a small fraction of total reads.

The filtered reads are then analyzed using either reference-based mapping or de novo assembly. In reference-based mapping, reads are aligned to known viral genomes in databases such as NCBI GenBank (Sayers et al., 2022). This is useful for identifying known viruses. In de novo assembly, short reads are merged into longer sequences called contigs without using a reference genome (Grabherr et al., 2011). Assemblers such as Trinity, SPAdes, and Velvet are widely used (Zerbino & Birney, 2008; & Bankevich et al., 2012). De novo assembly is especially important for discovering novel or highly divergent plant viruses.

After assembly, contigs are compared against viral sequence databases using similarity search tools such as BLAST (Altschul et al., 1990). This step assigns taxonomic identity to each contig and determines which virus species or strains are present. If a contig shows low similarity to known viruses, it may represent a novel virus candidate (Villamor et al., 2019). Researchers then inspect genome coverage, open reading frames, and conserved domains to confirm classification. This stage is where biological interpretation occurs.

Finally, the detected viruses are often validated using independent laboratory methods such as RT-PCR, qPCR, or Sanger sequencing (Mackay et al., 2002; & Adams et al., 2009). Validation confirms that the computationally detected virus is truly present in the sample and not a sequencing artifact. In scientific studies, this step strengthens confidence in the diagnosis and supports publication-quality evidence. Thus, data analysis does not end with software output it concludes with biological verification.

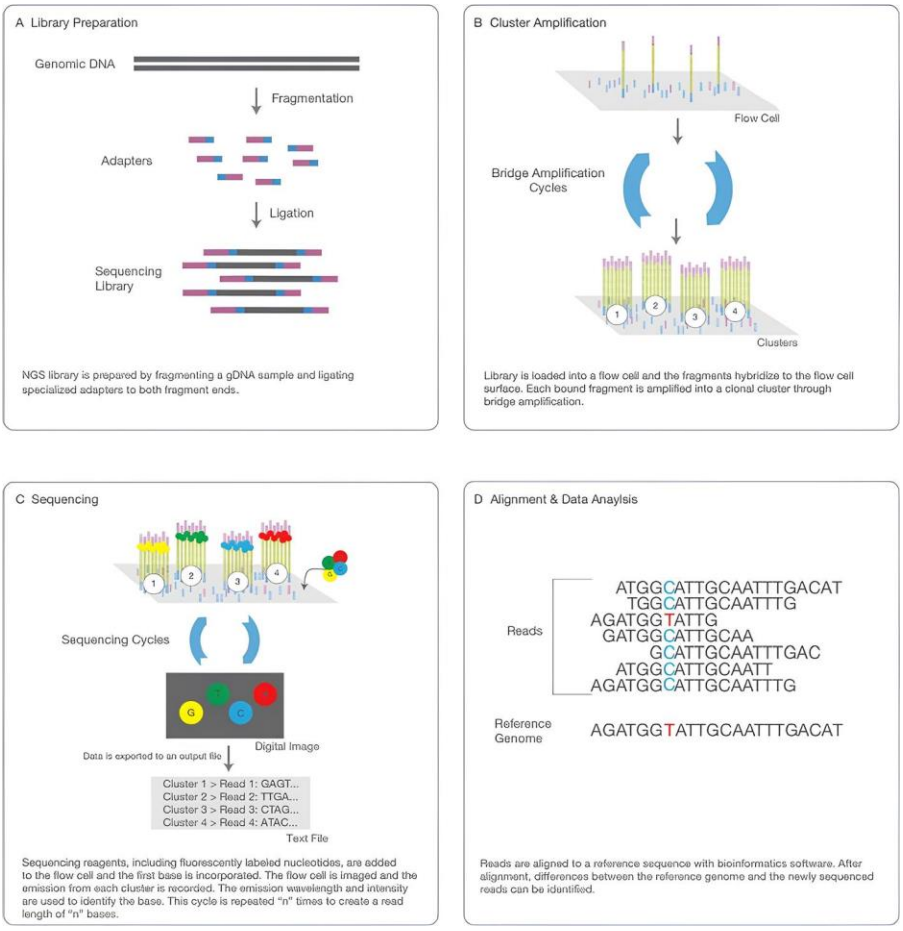

Fig. 2: Illumina Sequencing Workflow

Table2. Illumina platforms and their performance metrics

Illumina Platform	Maximum Read Length (bp)	Sequencing Output	Maximum Number of Reads per Run	Run Time
iSeq 100 System	2 × 150	1.2 Gb	4 million	9–17.5 h
MiniSeq System	2 × 150	7.5 Gb	25 million	4–24 h
MiSeq System	2 × 300	15 Gb	25 million	4–55 h
NextSeq 550 System	2 × 150	120 Gb	400 million	12–30 h
HiSeq 2500 System	2 × 125 / 2 × 250	1000 Gb / 300 Gb	4 billion / 600 million	29 h–6 days / 7–60 h
HiSeq 3000 System	2 × 150	750 Gb	2.5 billion	<1–3.5 days
HiSeq 4000 System	2 × 150	1500 Gb	5 billion	<1–3.5 days
HiSeq X Series	2 × 150	1800 Gb	6 billion	<3 days
NovaSeq 6000 System	2 × 150	6000 Gb	20 billion	16–44 h

Table 3: Applications of Illumina Sequencing in Plant Virus Research

Species	**Strategy**	**Sequencing Platform**	**Virus Studied**	**Application**	**Reference**
Arabidopsis thaliana	Small RNA deep sequencing	Illumina	*Tobacco mosaic virus* (Tobamovirus); *Cauliflower mosaic virus* (Caulimovirus); *Oilseed rape mosaic virus* (Tobamovirus); *Cucumber mosaic virus* (Cucumovirus); *Tobacco rattle virus* (Tobravirus); *Turnip mosaic virus* (Potyvirus)	Virus-host interaction	Qi et al., 2009
Nicotiana benthamiana	Small RNA deep sequencing	Illumina (with 454 FLX)	*Cymbidium ringspot virus* (Tombusvirus)	Virus-host interaction	Donaire et al., 2009
Ipomoea batatas (Sweet potato)	Small RNA deep sequencing	Illumina	*Sweet potato feathery mottle virus* (Potyvirus); *Sweet potato chlorotic stunt virus* (Crinivirus); novel DNA viruses (Badnavirus and Mastrevirus)	Discovery of novel virus; complete genome sequencing	Kreuze et al., 2009
Solanum lycopersicum (Tomato)	Small RNA deep sequencing	Illumina	*Potato spindle tuber viroid* (Pospiviroid); *Pepino mosaic virus* (Potexvirus); novel Potyvirus	Discovery of novel virus; complete genome sequencing	Li et al., 2012
Solanum lycopersicum (Tomato)	Small RNA deep sequencing	Illumina	*Two strains of tomato spotted wilt virus* (Tospovirus); unidentified Tospovirus; squash-infecting DNA virus	Quasi-species analysis	Hagen et al., 2011
Vitis vinifera (Grapevine)	Small RNA deep sequencing	Illumina	*Grapevine rupestris stem pitting-associated virus; Hop stunt viroid; Grapevine yellow speckle viroid 1; Grapevine fleck virus; Grapevine Syrah virus 1*	Discovery of novel virus and virome sequencing	Giampetruzzia et al., 2011
Capsicum annuum (Pepper)	—	Illumina	*Four Potato virus Y variants* (Potyvirus)	Quasi-species analysis	Fabre et al., 2012
Cucumis sativus (Cucumber)	Small RNA deep sequencing	Illumina	*Hop stunt viroid* (Hostuviroid)	Virus-host interaction	Martinez et al., 2010
Oryza japonica (Rice)	Small RNA deep sequencing	Illumina	*Rice stripe virus* (Tenuivirus)	Virus-host interaction	Yan et al., 2010
Gossypium hirsutum (Cotton)	Virus-derived dsRNA deep sequencing	Illumina	*Cotton leafroll dwarf virus* (Polerovirus)	Virus-host interaction	Silva et al., 2011

FUTURE PROSPECTUS

- Routine integration into plant diagnostic systems: NGS is expected to become a standard tool in plant disease diagnostics, certification programs, and quarantine services due to its ability to detect multiple pathogens simultaneously with high sensitivity.
- Rapid detection of emerging and novel viruses: Its unbiased sequencing approach will continue to play a crucial role in identifying newly emerging, previously unknown, or highly divergent plant viruses before they cause large-scale outbreaks.
- Advancement of portable and field-based sequencing: The development of compact sequencing platforms will enable on-site virus detection, reducing turnaround time and supporting real-time disease surveillance in agricultural settings.
- Enhanced genome-level characterization: Future NGS applications will improve complete viral genome reconstruction, mutation tracking, and evolutionary analysis, providing deeper insights into virus diversity and epidemiology.

CONCLUSION

Next-Generation Sequencing (NGS) has transformed plant virus detection by providing a rapid, highly sensitive, and unbiased approach for identifying both known and novel viruses in a single assay. It enables the comprehensive analysis of all nucleic acids present in a plant sample, allowing the simultaneous detection of mixed viral infections, viroids, satellite RNAs, and previously uncharacterized viruses without prior knowledge of their genomic sequences. This broad-spectrum capability makes NGS particularly valuable for diagnosing complex disease syndromes in crops that are caused by multiple pathogens. In addition to pathogen detection, NGS supports complete viral genome reconstruction, genetic variation analysis, recombination studies, and phylogenetic investigations, thereby improving our understanding of virus diversity, evolution, and epidemiology. Furthermore, the integration of advanced bioinformatics tools, including *de novo* assembly, BLAST-based sequence comparison, and genome annotation, enhances the accuracy and efficiency of virus identification and characterization, making NGS an indispensable tool in modern plant virology and disease diagnostics.

Acknowledgement:
I would like to sincerely thank my co-authors for their support and kind gesture to complete this manuscript in time.

Funding: NIL.

Conflict of Interest:
There is no such evidence of conflict of interest.

Author Contribution:
All authors have participated in critically revising of the entire manuscript and approval of the final manuscript.

REFERENCES

Adams, I. P., & Fox, A. (2016). Diagnosis of plant viruses using next-generation sequencing and metagenomic analysis. *Plant Pathology, 65*, 1137–1147.

Adams, I. P., Glover, R. H., Monger, W. A., Mumford, R., Jackeviciene, E., Navalinskiene, M., Samuitiene, M., & Boonham, N. (2009). Next-generation sequencing and metagenomic analysis: A universal diagnostic tool in plant virology. *Molecular Plant Pathology, 10*, 537–545.

Agrios, G. N. (2005). *Plant pathology* (5th ed.). Elsevier Academic Press.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215*, 403–410.

Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data*. Babraham Bioinformatics.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes:

A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology, 19*, 455–477.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., & Bignell, H. R. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature, 456*, 53–59.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible read trimming tool for Illumina NGS data. *Bioinformatics, 30*, 2114–2120.

Cock, P. J. A., Fields, C. J., Goto, N., Heuer, M. L., & Rice, P. M. (2010). The FASTQ file format for sequence data. *Nucleic Acids Research, 38*, 1767–1771.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., & Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology, 17*, 13.

Donaire, L., Wang, Y., González-Ibeas, D., Mayer, K. F., & Aranda, M. A. (2009). Deep sequencing of plant viral small RNAs reveals effective targeting of viral genomes. *Virology, 392*, 203–214.

Ewing, B., & Green, P. (1998). Base-calling of automated sequencer traces using Phred. *Genome Research, 8*, 175–185.

Fabre, F., Montarry, J., Coville, J., Senoussi, R., & Simon, V. (2012). Modelling the evolutionary dynamics of viruses within their hosts: A case study using high-throughput sequencing. *PLoS Pathogens, 8*, e1002654.

García-Arenal, F., Fraile, A., & Malpica, J. M. (2001). Variability and genetic structure of plant virus populations. *Annual Review of Phytopathology, 39*, 157–186.

Giampetruzzi, A., Roumi, V., Roberto, R., Malossini, U., Yoshikawa, N., La Notte, P., Terlizzi, F., Credi, R., & Saldarelli, P. (2011). A new grapevine virus discovered by deep sequencing of virus- and viroid-derived small RNAs in cv. Pinot gris. *Virus Research, 163*, 262–268.

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics, 17*, 333–351.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., & Zeng, Q. (2011). Trinity: Reconstructing transcriptomes from RNA-Seq data without a reference genome. *Nature Biotechnology, 29*, 644–652.

Hagen, C., Frizzi, A., Kao, J., Jia, L., Huang, M., Zhang, Y., & Huang, S. n.d. Using small RNA sequences to diagnose, sequence, and investigate the infectivity characteristics of vegetable-infecting viruses.

Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics, 107*, 1–8.

Hull, R. (2014). *Plant virology* (5th ed.). Academic Press.

Illumina (2016). *NextSeq 500 system guide*. Illumina Inc.

Jo, Y., Choi, H., Kim, S. M., Kim, S. L., Lee, B. C., & Cho, W. K. (2015). Integrated analysis using RNA-Seq data reveals viral genomes in plants. *Plant Pathology Journal, 31*, 135–143.

Jo, Y., Choi, H., Kim, S. M., Kim, S. L., Lee, B. C., & Cho, W. K. (2016). The pepper virome: Natural co-infection of diverse viruses revealed by RNA sequencing. *Scientific Reports, 6*, 29147.

Jo, Y., Lian, S., Cho, J. K., & Cho, W. K. (2017). Viral communities in different

plant species revealed by transcriptome analysis. *Scientific Reports, 7*, 11085.

Jo, Y., Lian, S., Cho, J. K., & Cho, W. K. (2018). Current status of plant virome studies using next-generation sequencing. *Molecules and Cells, 41*, 1–8.

Jones, S., Baizan-Edge, A., MacFarlane, S. A., & Torrance, L. (2017). Viral diagnostics in plants using next-generation sequencing. *Frontiers in Plant Science, 8*, 1770.

Kchouk, M., Gibrat, J. F., & Elloumi, M. (2017). Generations of sequencing technologies from first to next generation. *Biology and Medicine, 9*, 1000395.

Kircher, M., Sawyer, S., & Meyer, M. (2012). Double indexing for multiplex sequencing. *Nucleic Acids Research, 40*, e3.

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

Kukurba, K. R., & Montgomery, S. B. (2015). RNA sequencing and analysis. *Cold Spring Harbor Protocols*, 951–969.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment using Bowtie 2. *Nature Methods, 9*, 357–359.

Li, R., Gao, S., Hernandez, A. G., Wechter, W. P., & Fei, Z. (2012). Deep sequencing of small RNAs in tomato for virus and viroid identification and strain differentiation. *PLoS One, 7*, e37127.

Mackay, I. M., Arden, K. E., & Nitsche, A. (2002). Real-time PCR in virology. *Nucleic Acids Research, 30*, 1292–1305.

Maliogka, V. I., Minafra, A., Saldarelli, P., Ruiz-García, A. B., Glasa, M., & Katis, N. I. (2018). Advances in fruit tree virus detection using NGS. *Viruses, 10*, 436.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal, 17*, 10–12.

Martínez, G., Donaire, L., Llave, C., Pallás, V., & Gómez, G. (2010). High-throughput sequencing of hop stunt viroid-derived small RNAs from cucumber leaves and phloem. *Molecular Plant Pathology, 11*, 347–359.

Massart, S., Olmos, A., Jijakli, H., & Candresse, T. (2017). Current impact of high-throughput sequencing in plant virus diagnostics. *Virus Research, 241*, 189–199.

Massart, S., Chiumenti, M., & De Jonghe, K. (2019). Virus detection by high-throughput sequencing of small RNAs. *Phytopathology, 109*, 488–497.

Metzker, M. L. (2010). Sequencing technologies—the next generation. *Nature Reviews Genetics, 11*, 31–46.

Qi, X., Bao, F. S., & Xie, Z. (2009). Small RNA deep sequencing reveals a role for Arabidopsis thaliana RNA-dependent RNA polymerases in viral siRNA biogenesis. *PLoS One, 4*, e4971.

Roossinck, M. J., Martin, D. P., & Roumagnac, P. (2015). Plant virus metagenomics: Advances in virus discovery. *Phytopathology, 105*, 716–725.

Rubio, L., Galipienso, L., & Ferriol, I. 2020. Detection of plant viruses and disease management. *Frontiers in Plant Science, 11*, 1092.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *PNAS, 74*, 5463–5467.

Seguin, J., Otten, P., & Baerlocher, L. (2014). RNA silencing in plant virus infection. *Plant Methods, 10*, 39.

Silva, T. F., Romanel, E. A. C., Andrade, R. R. S., & Farinelli, L. (2011). Profile of small interfering RNAs from cotton plants infected with cotton leafroll

dwarf virus. *BMC Molecular Biology*, *12*, 40.

Vasanthan, E., Kundargi, C. Y., Gaddennavar, K. D., Bachaspati, S., Huded, S., Yadav, K., Marappan, K., & Pradhan, R. (2025). "Crosstalk in Plant Defense: Bridging Abiotic and Biotic Stresses". *plant cell biotechnology and molecular biology 26*(11-12):101–109.

Villamor, D. E. V., Ho, T., Al Rwahnih, M., Martin, R. R., & Tzanetakis, I. E. (2019). High throughput sequencing for plant virus detection. *Phytopathology*, *109*, 716–725.

Wang, Y. M. (2022). Plant viral disease detection: From molecular diagnosis to advanced sensing technologies. *Remote Sensing*, *14*, 1542.

Xiao, H., Shrestha, D., & Tzanetakis, I. E. (2019). Detection of plant viruses using next-generation sequencing. *Viruses, 11*, 183.

Yan, F., Zhang, H., Adams, M. J., Yang, J., Peng, J., Antoniw, J. F., Zhou, Y., & Chen, J. (2010). Characterization of siRNAs derived from rice stripe virus in infected rice plants by deep sequencing. *Archives of Virology, 155*, 935–940.

Zerbino, D. R., & Birney, E. (2008). Velvet: De novo assembly using short reads. *Genome Research, 18*, 821–829.

Zhao, G. (2015). RNA-seq-based transcriptome profiling. *Genomics, Proteomics* and *Bioinformatics, 13*, 293–301.