Chapter 29
Study of Plant Exclusive Virus-Derived Small Interfering RNAs

Ajeet Singh and Shailesh Kumar

29.1 Introduction

The emergence of high-throughput, fast, and cost-effective next-generation sequencing (NGS) technology has facilitated the study of small non-coding ribonucleic acids (RNAs) in eukaryotes and their role in RNA silencing mechanisms as a defense response during pathogen infection. Among different causal agents of infection, virus-mediated infections have a tremendous impact on the physiological system, nutritional value, and yield of crop plants (Diener 1963; Bos 1982). Thus, it becomes important to focus on the underlying anti-viral defense mechanisms in plants. Some important natural anti-viral defenses exploit small RNAs in combating the infection in host plants (Hamilton and Baulcombe 1999), preferably termed as virus-induced gene silencing (VIGS). VIGS also assists in the process of chromatin modification, translation process and thus a potent mediator for gene expression regulation bestowing the overall resistance in host plants against the viral defense. This have gained considerable attention in recent years by plant researchers and small interfering RNAs (siRNAs) being an integral component of VIGS have been extensively investigated and studied by plant scientists (Velásquez et al. 2009; Zhu and Guo 2012). Virus-induced infection leads to the production of small non-coding RNA molecules in plants and other diverse eukaryotes as well. This may result in either acquirement of anti-viral immunity or pathogenesis in few cases (Ding and Lu 2011). Major portion of these generates small RNA pool comprises of small interfering RNAs with the length ranging from 21–24 nucleotides (nt); bearing unphosphorylated overhangs of 2 nt at 3′-end. They are considered to be the probable gene expression regulators and component of anti-defence machinery in the host plant (Guo et al. 2016).

In plants, virus-derived siRNAs (vsiRNAs) can be generated from either ribonucleic acid (RNA) or deoxyribonucleic acid (DNA) viruses (Szittya et al.

A. Singh · S. Kumar (✉)
National Institute of Plant Genome Research (NIPGR), New Delhi, India
e-mail: shailesh@nipgr.ac.in

S. Kumar, C. Egbuna (eds.), *Phytochemistry: An in-silico and in-vitro Update*,
https://doi.org/10.1007/978-981-13-6920-9_29

2010). These vsiRNAs can be produced by processing of hairpin-shaped single-stranded RNA (folded structure) or double-stranded RNA genome (for RNA viruses). In the case of DNA viruses, they can be generated from replicative intermediates produced from double-stranded DNA genome in the host cells (Moissiard and Voinnet 2006; Donaire et al. 2009). In plants, the biogenesis of small interfering RNAs takes place with the help of the processing enzyme 'Dicer-like enzymes' (DCLs) (Chapman and Carrington 2007; Chen 2010). Specifically the homologs viz. DCL-2, 3 and 4 are participating directly in the vsiRNAs production while DCL-1 are indirectly involved in the biogenesis of plant vsiRNAs (Zhu and Guo 2012). These DCL homologs contribute in multifarious ways for vsiRNAs production while maintaining mutual balance and coordination with each other. There are two categories of vsiRNAs: primary and secondary vsiRNAs. The primary vsiRNAs are generated by the direct action of DCLs. The association of Argonaute proteins with vsiRNAs leads to the formation of the RNA-induced silencing complex (RISC) (Malpica-Opez et al. 2018). The RISC complemented further with plant RNA-dependent RNA polymerases (RDRs) attacks viral genome. During the initial phase, the viral genome is disintegrated into small fragments of dsRNA by the action of DCLs and further, RDRs convert these primary vsiRNAs into the highly active secondary vsiRNAs during the secondary amplification phase (Vazquez and Hohn 2013).

vsiRNAs assembles with RISC in a sequence-specific fashion and pair with its homologous complementary viral RNA or DNA genomic transcript strand and thus aid in silencing expression of the viral genome and in this way, they impart anti-viral resistance in the host plant (Szittya et al. 2010; Zhang et al. 2015). Although this pathway of anti-viral defense is vaguely explored, they are supposed to regulate cellular activities epigenetically by mediating DNA methylation in gene promoters (Rodríguez-Negrete et al. 2009). In addition to their crucial role in transcriptional gene silencing (TGS) and post-transcriptional gene silencing (PTGS), artificially synthesized siRNAs can also be very useful for gene knockout and pathways studies associated with gene silencing during varied stress conditions in plants (Guo et al. 2016). The rapid development of Next Generation Sequencing (NGS) technology has been heavily exploited for studying viral genomics, viral ecological studies, virus-host interactions, and evolution of viruses utilizing RNA interference technology with aid of vsiRNAs (Stobbe and Roossinck 2014; Skums et al. 2015). The hairpin construct approach for designing artificial vsiRNAs can be expressed in plant cells and used for targeting against the specific pathogen (Mansoor et al. 2006; Shimizu et al. 2012). By designing multiple hairpin constructs with different viral sources, transgenic plants have been developed which are resistant to a number of viruses (Prins et al. 1995; Bucher et al. 2006). Thus RNA interference (RNAi) technology has proved to be a boon in horticulture and agriculture for developing plants immune to pathogenic viruses (Duan et al. 2012). The biogenesis and mode of action of vsiRNAs are illustrated in Fig. 29.1.

At present, many databases pertaining to siRNAs and virus-derived siRNAs are available. However, these databases are largely focussed on human diseases caused by viruses. For e.g., HIVsirDB (Tyagi et al. 2011), VIRsiRNAdb (Thakur et al. 2012) and siRNAdb (Chalk et al. 2004). Nevertheless, the wide impact of plant vsiRNAs on the physiology of plants cannot be ignored. There is a need of knowledge base dedicated

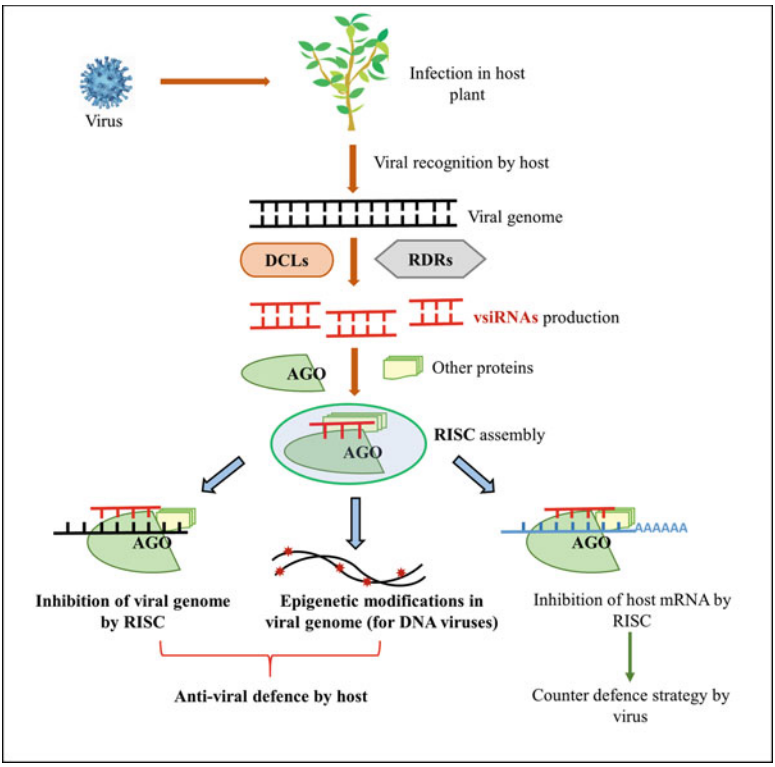

Fig. 29.1 Biogenesis and mode of action of plant vsiRNAs

only for the plant vsiRNAs. This chapter discusses about the database, PVsiRNAdb exclusively for plant vsiRNAs (Gupta et al. 2018). PVsiRNAdb (http://www.nipgr.res.in/PVsiRNAdb) is developed by extensive data mining and harboring information of plant vsiRNAs from literature available till date. The resources available online pertaining to vsiRNAs hold predicted as well as annotated sequences detected in virus-infected plants. This database is developed in such a user-friendly manner for convenience.

29.2 Materials

For this study, data subjected to virus interaction with the plant was collected by data mining of PubMed literature by Gupta et al. and developed a web-based platform named as PVsiRNAdb. It contains information regarding vsiRNA sequences from 20 different viruses infecting 12 different plants which are listed in Table 29.1 with total number of vsiRNAs.

Table 29.1 List of viruses, host plant and total vsiRNAs stored in PVsiRNAdb (Gupta et al. 2018)

Virus	Host plant	Total vsiRNAs
Bamboo mosaic virus	*Dendrocalamus latiflorus* (bamboo)	18
Brassica yellow virus	*Nicotiana benthamiana* (tobacco)	143
Chinese wheat mosaic virus	*Triticum aestivum* (wheat)	19,536
Cotton leaf curl Multan virus	*Gossypium hirsutum* (cotton)	4736
Cucumber green mottle mosaic virus	*Cucumis sativus* (cucumber)	92
Cucumber mosaic virus	*Arabidopsis thaliana* (thale cress)	47
Cymbidium ringspot virus	*Nicotiana benthamiana* (tobacco)	12,305
Grapevine fleck virus and Grapevine rupestris stem pitting associated virus	*Vitis vinifera* (grapevine)	62
Maizechlorotic mottle virus and Sugarcane mosaic virus	*Zea mays* (maize)	260
Pea enation mosaic virus 2	*Nicotiana benthamiana* (tobacco)	137
Potato virus Y	*Solanum tuberosum* (potato)	46,435
Prunus necrotic ring spot virus	*Prunus avium* (cherry)	152
Rice black streaked dwarf virus	*Oryza sativa* (rice)	468
Southern Rice black streaked dwarf virus	*Oryza sativa* (rice)	20,876
Sugarcane mosaic virus	*Zea mays* (maize)	100
Tobacco mosaic virus	*Nicotiana benthamiana* (tobacco)	682
Tobacco rattle virus	*Nicotiana benthamiana* (tobacco)	142
Wheat yellow mosaic virus	*Triticum aestivum* (wheat)	216,018
Zucchini yellow mosaic virus	*Citrullus lanatus* (watermelon)	5

29.3 Methods

29.3.1 Data Collection

An extensive literature search was carried out to excerpt the relevant articles from PubMed (https://www.ncbi.nlm. nih.gov/pubmed). This was carried out by searching queries using a different combination of keywords e.g. viral siRNAs,

plant-virus interaction, siRNA, plant-viral siRNAs etc. Relevant experimental information was extracted by a manual screening of articles. Literature lacking relevant information regarding this study were excluded. Full-text search was done for each of the relevant article having the information of plant-specific vsiRNAs. In addition to this, the relevant information associated with the plant, tissue, PubMed ID (PMID) and PVsiRNA-ID was incorporated along with the collected information of vsiRNAs.

29.3.2 PVsiRNAdb Web Platform

PVsiRNAdb web-interface was built on an Apache Hypertext Transfer Protocol server by using Hypertext Markup Language (HTML), Cascading Style Sheets (CSS), Hypertext Preprocessor (PHP) and JavaScript. MySQL, an object-relational database management system (RDBMS), was used to manage all the data in the backend. It provides commands to retrieve and store the data in the database. All common gateway interface and database interfacing scripts were written in the Hypertext Preprocessor (PHP), and Practical Extraction and Reporting Language (PERL).

29.3.3 Organization of Database

The information in PVsiRNAdb is organized at two levels, primary and secondary (Fig. 29.2). At the primary level queries are searched by specific plant name, virus name, PMID or other options as per the users' requirement.

The information will be displayed according to the number of fields selected by the user. The user can also search multiple queries for virus, plant or PMIDs by performing a batch search. The data at the secondary level can be utilized for the retrieval of further information about primary data. At the secondary level, additional information like experimental details, sequence-related information and details of virus-like name, type of genome and classification can also be fetched for each viral strain. The virus name, genome type as well as classification can also be retrieved for each viral strain. The specific details about any experiment can be inquired by clicking on PMID hyperlink, which will direct the user to the original link of that research article. As structure plays an important role in determining the function of any sequence, the secondary structure of vsiRNA sequences was added to the database using in-house generated PERL scripts for running the Mfold (Zuker 2003) and RNA structure (Reuter and Mathews 2010) packages. Mfold was utilized for calculating minimized energy for the folded structure and structure coordinates were predicted by Draw utility of RNA structure software.

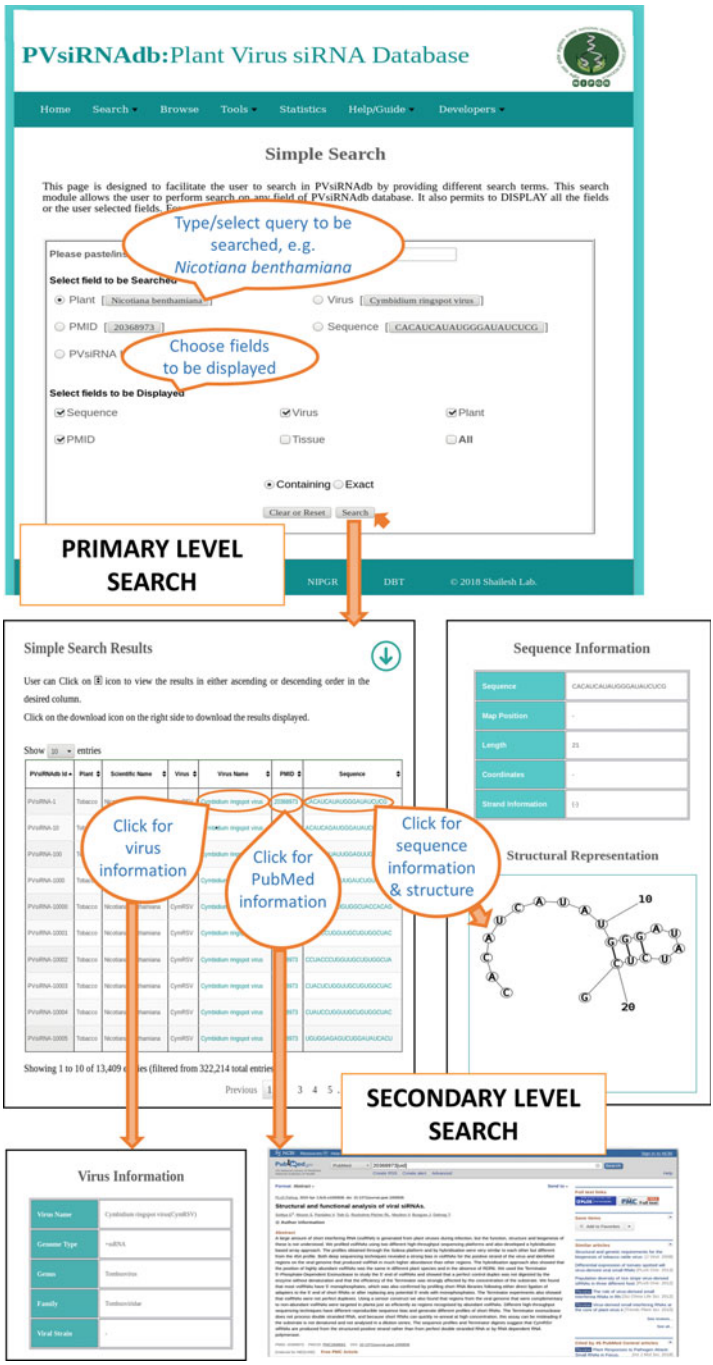

Fig. 29.2 Information at the primary and secondary level of search

Fig. 29.3 Illustration of 'Search' option in PVsiRNAdb (**a**) The representation of 'Simple Search' module. (**b**) The window showing 'Batch Search' module with query example

29.3.4 Features and Tools

In PVsiRNAdb, detailed and comprehensive information is incorporated for each siRNA entry. Apart from the core information including the siRNA sequence, siRNA length, virus name, and plant name, additional information like PMID, plant tissue, mapping coordinates of siRNA to the plant genome and the predicted secondary structures of siRNA may be of high utility to the user. PVsiRNAdb provides two user-friendly options to search for siRNA information i.e., 'Simple Search', and 'Batch Search' (Fig. 29.3a, b).

'Simple Search' allows the user to search the query by providing different search terms like the name of the virus, plant name (scientific or common name), siRNA sequence, PMID, and PVsiRNA-ID. For providing the flexibility in the search module, the 'containing' and 'exact' option has been incorporated. This option also facilitates the user to select the fields to be displayed. A total of five display fields are available for a search term. Three display fields namely the 'Virus name', 'PMID' and 'Sequence' are further linked with their corresponding information. Second option to search in PVsiRNAdb is that of the 'Batch Search' providing the facility to search for multiple queries at a time. The user can extract the information of siRNAs by providing a list of plant names, virus names or PMIDs. In this module, an example list of all the three search terms is provided for the users. The PVsiRNAdb information can be browsed by virus name, plant name or PubMed ID by expanding the respective option in 'Browse' section (Fig. 29.4).

'Tools' section contains three module – 'BLAST', 'SW Align', and 'Mapping' (Fig. 29.5). BLAST module is developed by *blastn* utility of ncbi-blast 2.6.0 (https://blast.ncbi.nlm.nih.gov/Blast.cgi) standalone version. The user provided query sequence(s) can be aligned to blast database i.e., PVsiRNAdb. This module is also provided with the option to select the virus genome and change the E-value for alignment. BLAST result, besides the alignment result, also gives each hit

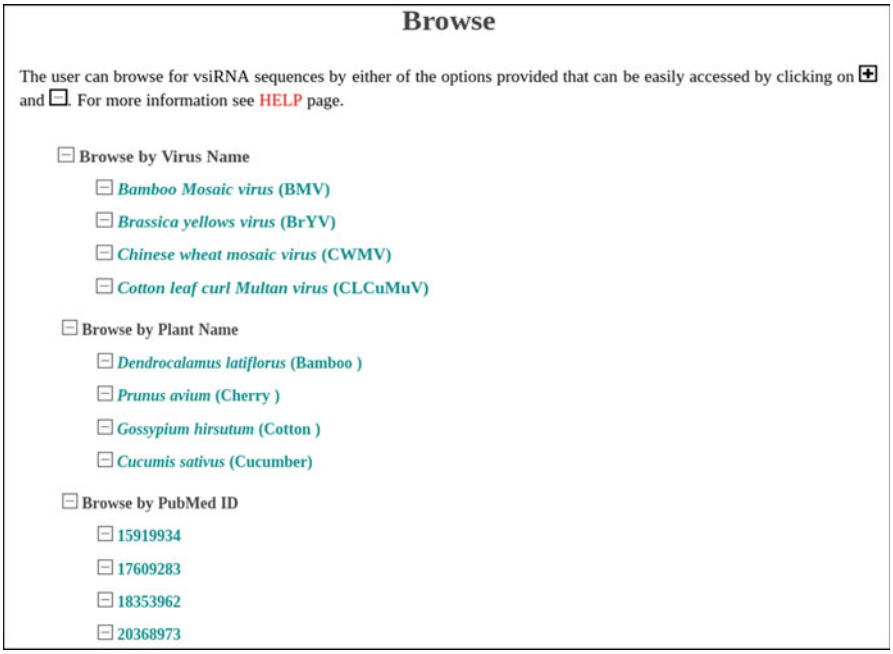

Fig. 29.4 The 'Browse' section of PVsiRNAdb displaying three different options provided

information by a click on it (Fig. 29.5a). 'SW Align' module uses the *water* utility of EMBOSS-6.6.0 (http://emboss.open-bio.org/) to align the query to selected virus siRNA dataset. In-house developed PERL scripts are integrated with the Smith-Waterman algorithm to take the result in the desired pattern (Fig. 29.5b). The 'Mapping' module is designed for the mapping of siRNA sequences, available at PVsiRNAdb to the user-provided sequences e.g. messenger RNA sequences or genomic sequences (Fig. 29.5c). This module uses the *makeblastdb* and *blastn* utility of ncbi-blast 2.6.0 with the PERL script integration. 'Mapping' facility is useful for the designing of specific siRNAs corresponding to the specific viral genome.

Overall statistics of PVsiRNAdb are illustrated in 'Statistics' section in the form of tables and histogram. Any query regarding the use of PVsiRNA web interface is answered by 'Help/Guide' section. This section apart from 'Help' sub-section, also contains two more sub-section i.e., 'Links' and 'References'. From the 'Help' of PVsiRNAdb, the user can understand the working of this database with the help of self-explanatory figures. 'Links' directs the user to important web resources contains information on viral siRNAs. In the 'References' of PVsiRNAdb web interface, all the articles related to vsiRNAs involved plant-virus interaction have been incorporated.

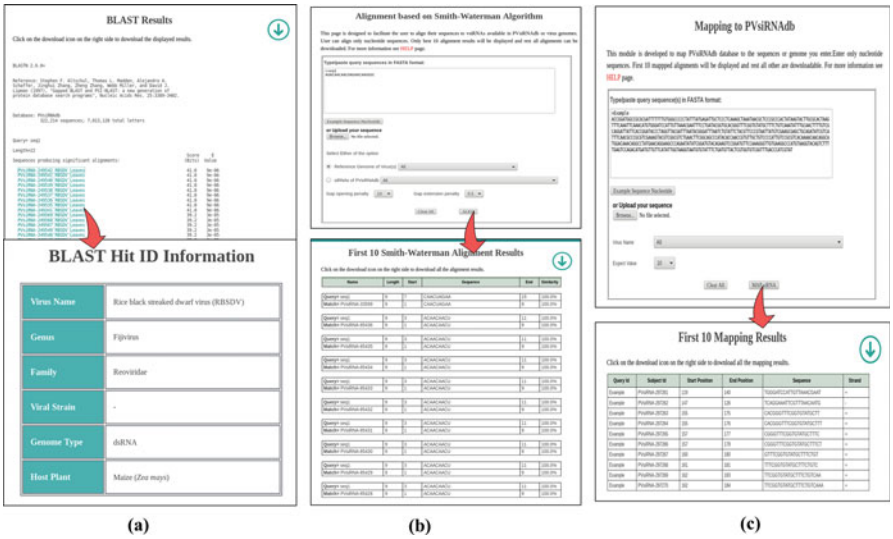

Fig. 29.5 Three modules in 'Tools' section of PVsiRNdb (**a**) 'BLAST' module window showing alignment result and blast hit information. (**b**) The output of 'SW Align' result after a query search. (**c**) A query and their mapping result by 'Mapping' module

Acknowledgment PVsiRNAdb database is developed by Dr. Shailesh Kumar research group (http://www.nipgr.res.in/research/dr_shailesh.php) at National Institute of Plant Genome Research (NIPGR), New Delhi, India.

References

Bos L. Crop losses caused by viruses. Crop Prot. 1982;1:263–82. https://doi.org/10.1016/0261-2194(82)90002-3.

Bucher E, Lohuis D, van Poppel PMJA, et al. Multiple virus resistance at a high frequency using a single transgene construct. J Gen Virol. 2006;87:3697–701. https://doi.org/10.1099/vir.0.82276-0.

Chalk AM, Warfinge RE, Georgii-Hemming P, Sonnhammer ELL. siRNAdb: a database of siRNA sequences. Nucleic Acids Res. 2004;33:D131–4. https://doi.org/10.1093/nar/gki136.

Chapman EJ, Carrington JC. Specialization and evolution of endogenous small RNA pathways. Nat Rev Genet. 2007;8:884–96. https://doi.org/10.1038/nrg2179.

Chen X. Small RNAs – secrets and surprises of the genome. Plant J. 2010;61:941–58. https://doi.org/10.1111/j.1365-313X.2009.04089.x.

Diener TO. Physiology of virus-infected plants. Annu Rev Phytopathol. 1963;1:197–218. https://doi.org/10.1146/annurev.py.01.090163.001213.

Ding S-W, Lu R. Virus-derived siRNAs and piRNAs in immunity and pathogenesis. Curr Opin Virol. 2011;1:533–44. https://doi.org/10.1016/j.coviro.2011.10.028.

Donaire L, Wang Y, Gonzalez-Ibeas D, et al. Deep-sequencing of plant viral small RNAs reveals effective and widespread targeting of viral genomes. Virology. 2009;392:203–14. https://doi.org/10.1016/j.virol.2009.07.005.

Duan C-G, Wang C-H, Guo H-S. Application of RNA silencing to plant disease resistance. Silence. 2012;3:5. https://doi.org/10.1186/1758-907X-3-5.

Guo Q, Liu Q, Smith NA, et al. RNA silencing in plants: mechanisms, technologies and applications in horticultural crops. Curr Genomics. 2016;17:476–89. https://doi.org/10.2174/1389202917666160520103117.

Gupta N, Singh A, Zahra S, Kumar S. PtRFdb: a database for plant transfer RNA-derived fragments. Database. 2018;2018:63.

Hamilton AJ, Baulcombe DC. A species of small antisense RNA in posttranscriptional gene silencing in plants. Science (New York, NY). 1999;286:950–2.

Malpica-Opez N, Rajeswaran R, Beknazariants D, et al. Revisiting the roles of tobamovirus replicase complex proteins in viral replication and silencing suppression. MPMI. 2018;311094:125–44. https://doi.org/10.1094/MPMI-07-17-0164-R.

Mansoor S, Amin I, Hussain M, et al. Engineering novel traits in plants through RNA interference. Trends Plant Sci. 2006;11:559–65. https://doi.org/10.1016/J.TPLANTS.2006.09.010.

Moissiard G, Voinnet O. RNA silencing of host transcripts by cauliflower mosaic virus requires coordinated action of the four Arabidopsis Dicer-like proteins. Proc Natl Acad Sci. 2006;103:19593–8. https://doi.org/10.1073/pnas.0604627103.

Prins M, de Haan P, Luyten R, et al. Broad resistance to tospoviruses in transgenic tobacco plants expressing three tospoviral nucleoprotein gene sequences. Mol Plant-Microbe Interact. 1995;8:85–91. https://doi.org/10.1094/MPMI-8-0085.

Reuter JS, Mathews DH. RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinform. 2010;11:129. https://doi.org/10.1186/1471-2105-11-129.

Rodríguez-Negrete EA, Carrillo-Tripp J, Rivera-Bustamante RF. RNA silencing against geminivirus: complementary action of posttranscriptional gene silencing and transcriptional gene silencing in host recovery. J Virol. 2009;83:1332–40. https://doi.org/10.1128/JVI.01474-08.

Shimizu T, Nakazono-Nagaoka E, Akita F, et al. Hairpin RNA derived from the gene for Pns9, a viroplasm matrix protein of Rice gall dwarf virus, confers strong resistance to virus infection in transgenic rice plants. J Biotechnol. 2012;157:421–7. https://doi.org/10.1016/j.jbiotec.2011.12.015.

Skums P, Artyomenko A, Glebova O, et al. Computational framework for next-generation sequencing of heterogeneous viral populations using combinatorial pooling. Bioinformatics. 2015;31:682–90. https://doi.org/10.1093/bioinformatics/btu726.

Stobbe AH, Roossinck MJ. Plant virus metagenomics: what we know and why we need to know more. Front Plant Sci. 2014;5:150. https://doi.org/10.3389/fpls.2014.00150.

Szittya G, Moxon S, Pantaleo V, et al. Structural and functional analysis of viral siRNAs. PLoS Pathog. 2010;6:e1000838. https://doi.org/10.1371/journal.ppat.1000838.

Thakur N, Qureshi A, Kumar M. VIRsiRNAdb: a curated database of experimentally validated viral siRNA/shRNA. Nucleic Acids Res. 2012;40:D230–6. https://doi.org/10.1093/nar/gkr1147.

Tyagi A, Ahmed F, Thakur N, et al. HIVsirDB: a database of HIV inhibiting siRNAs. PLoS One. 2011;6:e25917. https://doi.org/10.1371/journal.pone.0025917.

Vazquez F, Hohn T. Biogenesis and biological activity of secondary siRNAs in plants. Scientifica. 2013;2013:783253. https://doi.org/10.1155/2013/783253.

Velásquez AC, Chakravarthy S, Martin GB. Virus-induced gene silencing (VIGS) in Nicotiana benthamiana and tomato. JoVE. 2009; https://doi.org/10.3791/1292.

Zhang C, Wu Z, Li Y, Wu J. Biogenesis, function, and applications of virus-derived small RNAs in plants. Front Microbiol. 2015;6:1–12. https://doi.org/10.3389/fmicb.2015.01237.

Zhu H, Guo H. The role of virus-derived small interfering RNAs in RNA silencing in plants. Sci China Life Sci. 2012;55:119–25. https://doi.org/10.1007/s11427-012-4281-3.

Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 2003;31:3406–15.