Sensitivity of small RNA –based detection of plant viruses

Santala Johanna[1], Jari Valkonen[2]
[1]Finnish Food Safety Authority Evira, [2]University of Helsinki

The aim of this study was to estimate the detection threshold of small RNA-based virus diagnosis in plants, as compared with an established virus species specific polymerase chain reaction (PCR) method. Viruses were propagated in tobacco plants, total RNA was isolated and subjected to reverse-transcription real time PCR (RT-RT-PCR) and small RNA deep-sequencing at different concentrations. The specific small RNA approach used in this study showed similar sensitivity compared to the currently used RT-RT-PCR.

Introduction

Official plant inspections are conducted to prevent spread of plant diseases and pests to areas of plant production. Collected samples are inspected and analyzed in official plant health laboratories. The currently used methods are pest-specific, and analysis of samples is limited to the pests stipulated in legislation. The large number of plant pathogenic pests has made it impossible to test all types of pests possibly present in a sample.

Recently, it has become obvious that plants recognize all types of viruses via an antiviral defense system targeting and processing viral double-stranded RNA to so called small RNAs (sRNA; 21, 22 or 24 nucleotides) to prevent the spread of viruses in the plant. These sRNAs can be extracted from plant tissues and sequenced, and the sequences reassembled into partial or full viral genomes in silico. This method allows detection of all types of plant viruses simultaneously in the plant by comparison with viral sequences available in databases.

Adopting a new method for use in official plant inspection requires validation of specific characteristics in test performance, such as sensitivity, specificity, repeatability and reproducibility. In this study, sensitivity of the sRNA-based virus detection was compered with the currently used reverse transcription real-time polymerase chain reaction (RT-RT-PCR) method.

Materials and methods

Tobacco plants were inoculated with an infectious clone of potato virus Y (PVY), and also potato virus A (PVA) that was engineered to express red fluorescent protein (PVA-RFP), green fluorescent protein (PVA-GFP) or β-glucuronidase (PVA-GUS). Three weeks post inoculation total RNA (including sRNA) was isolated from upper leaves of the plants. The concentrations of PVA and PVY RNA in the tobacco leaves were estimated by quantitative PCR. Ct-values obtained from the tobacco leaves were compared with a standard curve generated with RNA extracted from virus particles of PVA and PVY. For sRNA sequencing, different amounts of viral RNA were added to four RNA pools, each containing 3 µg of total RNA from berry plants in a volume of 15 µl (**Table 1**). Library preparation (TruSeq small RNA kit, Illumina) and sequencing using the 50 base pair single-end mode (HiSeq 2500, Illumina) were conducted by Fasteris SA.

Velvet software (k-mer=15, COV=50) was used to produce contiguous sequences (contigs) from the obtained 21–24 nucleotides (nt) long reads. The contigs corresponding to the reference sequences (RFP, GFP, GUS, PVY) were identified by alignment to NCBI database using BLASTn. Mapping was conducted by aligning all obtained 21-24 reads against the reference sequences using Novoalign software. Sensitivity of the above described sRNA approach was compared with commercially available RT-RT-PCR test (Fera Science, Ltd.) designed to detect PVA and PVY.

Table 1. Composition of the RNA pools (3 µg) subjected to sRNA deep sequencing.

Pool	Amount of viral RNA (ng)			
	PVY	PVA-GFP	PVA-RFP	PVA-GUS
1	1.0	0.00002	0.01	1.0
2	0.01	0.0002	0.2	0.00001
3	0.0001	0.1	0.000001	0.002
4	0.000001	0.000002	0.0001	0.02

This study was supported by Ministry of Agriculture and Forestry, Finland. Published in Frontiers in Microbiology 9:939.

Results

The amounts of RNA of PVY, PVA-GFP, PVA-RFP and PVA-GUS in the leaves of the infected tobacco plants were 15.08 ng, 0.86 ng, 2.20 ng, and 11.12 ng per 1 µg of total RNA, respectively, as determined by RT-qPCR. Subsequently, different amounts of RNA were added to the RNA pools (**Table 1**) and subjected to deep-sequencing. In the Velvet-based de novo analysis, PVY-specific contigs were obtained from the pools amended with total tobacco RNA containing 1 ng or 0,01 ng of PVY RNA, but not from the pools containing 0,0001 ng or 0,000001 ng of PVY RNA (**Table 2**). Contigs specific to the reference sequences of the engineered PVA clones were obtained from the pools containing 0,002 ng or higher amounts of the respective PVA RNA, but not from the pools containing 0,0002 ng or less PVA RNA (**Table 2**). When all the 21 to 24 nt long reads were mapped to the reference sequences with Novoalign, the numbers of alignments diminished with diminishing amounts of viral RNA (**Figure 1**). On the other hand, 0.002 ng or higher amounts of viral RNA resulted in nearly complete coverage (99%) of the respective reference sequence. As little as 0.00002–0.0002 ng of viral RNA was sufficient for covering 25–50% of the reference sequence with virus-derived sRNA reads. The RT-RT-PCR test detected PVA and PVY in the samples containing ≥0.0002 ng of PVA RNA or ≥0.0001 ng of PVY RNA, respectively, in 3 µg of total RNA in a volume of 15 µl.

Table 2. Number of contigs assembled from sRNAs by Velvet (k-mer = 15, COV = 50 nt) and identified as the references sequences by BLASTn in four pools of samples.

Pool[a]	Amount of reference RNA added (ng)[b]	Reference sequence and length (nt)[b]	Contigs specific to reference sequences (BLASTn)		Total no. of sequenced sRNAs (21–24 nt)	Total no. and length of contigs
			Number	Length (nt)		
1	1	PVY (9701)	46	≤926	14,159,712	2013 (≤926)
	1	GUS (1812)	13	≤237		
	0.01	RFP (678)	4	≤110		
	0.00002	GFP (741)	0	–		
2	0.2	RFP (678)	3	≤513	15,856,030	3081 (≤652)
	0.01	PVY (9701)	54	≤220		
	0.0002	GFP (741)	0	–		
	0.00001	GUS (1812)	0	–		
3	0.1	GFP (741)	6	≤235	12,332,818	3255 (≤739)
	0.002	GUS (1812)	5	≤102		
	0.0001	PVY (9701)	0	–		
	0.000001	RFP (678)	0	–		
4	0.02	GUS (1812)	15	≤196	9,856,044	3218 (≤391)
	0.0001	RFP (678)	0	-		
	0.000002	GFP (741)	0	-		
	0.000001	PVY (9701)	0	-		

[a]Pools correspond to those shown in **Table 1**.
[b]Amounts of PVY and PVA RNA added to a mixture of total RNA from leaves of strawberry, raspberry and *Ribes* spp. PVA was engineered to carry a reference gene encoding beta-glucuronidase (GUS) or green (GFP) or red (RFP) fluorescent protein. Lengths of the reference RNAs (in nucleotides) are shown in parentheses.

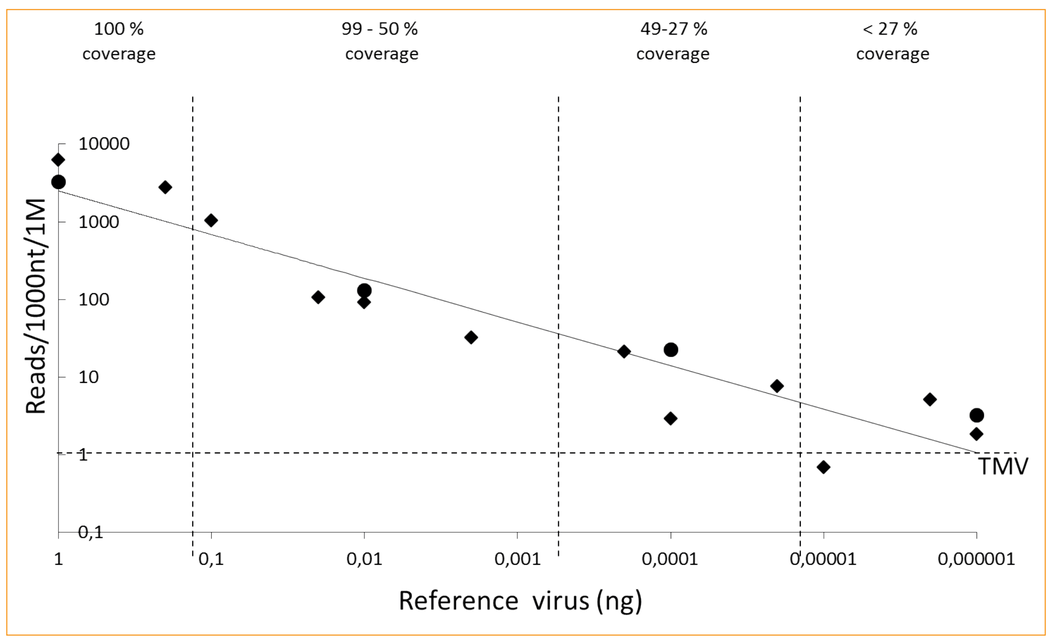

Fig 1. Reference sequence–guided analysis of the obtained 21- to 24-nt reads. The descending trend line (y = 2486,1x$^{0.5616}$; correlation coefficient, 0.8927) illustrates the correlation between the number of aligned reads and the amount of viral RNA. Novoalign software was used to align the sRNA reads to the reference sequences. The y axis shows the number of reads that aligned per 1000 nt of the reference sequence, i.e., genomic sequence of PVY or the marker gene sequences *GFP, GUS* and *RFP,* each carried by a different copy of the PVA genome, per 1 million (1M) total reads. The x axis indicates the amount of PVY genomic RNA or RNA corresponding to *GFP, GUS* or *RFP* engineered into the genome of PVA. The sequenced RNA pool of 15 µl also contained 3 µg of total RNA obtained by mixing RNA from strawberry, raspberry and *Ribes* spp. Circles refer to PVY, whereas quadrilaterals refer to the three reference sequences carried by PVA. Both axes are in a logarithmic scale. The vertical dashed lines were included to visualize how diminishing amounts of viral RNA reduce the coverage of reference sequences by sRNA reads and hence limit the possibility of detecting viruses. The horizontal dashed line indicates the background noise caused by alignment of a few reads to the sequence of *Tobacco mosaic virus* (TMV), which was not included in the samples.

Conclusions

Taken together, the sRNA sequencing approach used in this study and the subsequent analysis of the data by *de novo* assembly of contigs required the equivalent of 30 fg of viral RNA in 50 ng of total RNA for detection of the viruses. However, by mapping all the 21- to 24-nt long reads to existing reference sequences, detection sensitivity reached the level of 0.3 fg of viral RNA in 50 ng of total RNA. The detection limit of PVY and PVA by RT-RT-PCR in this study was 1.5 and 3 fg of viral RNA, respectively, in 50 ng of total RNA per PCR reaction. Hence, prior knowledge about the viruses possibly present in the samples allows sRNA-based diagnosis to reach similarly high sensitivity as RT-RT-PCR in the detection of viruses. Thus, sensitivity of the mapping approach is suitable for use in rutine diagnostics of plant viruses. Additionally, *de novo* analysis of the sequence data can be conducted when no educated presumptions are available about the viruses present in the sample.

Evira

UNIVERSITY OF HELSINKI