Oligonucleotide microarray with a minimal number of probes for the detection and identification of thirteen genera of plant viruses

Yongjiang Zhang [a,1], Jun Yin [b,1], Guifen Li [a], Mingfu Li [a], Xin Huang [a], Hongjun Chen [a], Wenjun Zhao [a], Shuifang Zhu [a,*]

[a] Institute of Animal and Plant Quarantine, Chinese Academy of Inspection and Quarantine, No. 241 Huixinli, Huixin West Street, Beijing 100029, China
[b] School of Medicine and Medical Science, Conway Institute, University College Dublin, Belfield, Dublin 4, Ireland

ABSTRACT

A major challenge facing agriculture at present is the development of techniques that can screen field samples and other plant materials simultaneously for the presence of many viruses. Microarray techniques show promise in this regard, as their high throughput nature can potentially detect a range of viruses using a single test. In this paper we present an array that can detect a wide spectrum of 169 plant virus species from 13 different genera. The array was constructed using an automated probe design protocol which generated a minimal number of probes to detect viruses at the genus level. The designed arrays showed a high specificity and sensitivity when tested with a set of standard virus samples. Field samples collected from a severe disease outbreak of *Panax notoginseng* farms in Yunnan, China, in 2001 were screened, where a *potyvirus* infection was identified associated with the disease.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Plants can be infected by a wide range of viruses that may cause damage to affected crops and result in significant economic losses to the agricultural industries. Early and accurate pathogen diagnosis of these pathogens is essential for disease control. Many diagnostic techniques have been developed for plant virus detection. These include bioassays, electron microscopy, serological methods such as enzyme-linked immunosorbent assays (ELISA), molecular hybridization and polymerase chain reaction (PCR) (for reviews see James et al., 2006; Mumford et al., 2006).

With the exception of the microarray method, the above mentioned viral detection techniques can only be applied to the detection of specific viruses or viruses that are transmissible to indicator hosts or readily observed with an electron microscope. They lack the capability of discovering new novel viruses that cause emerging viral infections. The risk from emerging viruses has grown significantly due to a number of factors. These may include: (i) new virus species, (ii) established viruses being transferred horizontally to new host species, (iii) viruses invading new territories through international trade. Therefore there is a need for high throughput and broad spectrum strategies for detection, identification and monitoring of plant viruses.

Microarrays show promise in this regard, as their high throughput nature can potentially detect a range of viruses and/or their strains using a single test. Several studies have been published in this area. Boonham et al. (2003) reported identification of *Potato virus Y* (PVY), *Potato virus A* (PVA), *Potato virus X* (PVX) and *Potato virus S* (PVS) in single and multiple infections. Lee et al. (2003) designed a cDNA chip for detection of four cucurbit-infecting tobamoviruses. However, both these studies used cDNA microarrays which are expensive and labor intensive to construct. Bystricka et al. (2005) designed a synthetic oligonucleotide microarray containing thirty-four 40mer probes to detect 6 potato viruses. Pasquini et al. (2008) used several 70mer probes for specific detection and genotyping of strains of *Plum pox virus*. Recently Wei et al. (2009) developed a short oligonucleotide microarray detecting four potyviruses. These plant virus microarrays focused only on detecting viruses from specific plant host(s) or specific virus strains, species or genera (for further review, see Boonham et al., 2007).

For vertebrate virus detection, a few comprehensive microarrays have been developed. These are able to screen a wide range of vertebrate viruses. Wang et al. (2002) reported a microarray (ViroChip), using 70mer probes, capable of detecting simultaneously hundreds of viruses. The successful detection and identification of severe acute respiratory syndrome virus (SARS) using this microarray highlighted the potential of the conserved probe design in detecting emerging viruses (Wang et al., 2003).

* Corresponding author. Tel.: +86 10 64912744; fax: +86 10 85393944.
E-mail address: zhushf@netchina.com.cn (S. Zhu).
[1] These authors contributed equally to the paper.

Recently, a panmicrobial microarray (GreeneChipVr) comprising 29,455 60mer oligonucleotide probes for vertebrate viruses, bacteria, fungi, and parasites has been reported (Palacios et al., 2007). Both of these microarrays used a large number of probes to compensate for the loss of sensitivity and stability from using 60–70mer oligonucleotide probes. This redundancy of probes can significantly increase the cost of the microarray.

A new plant virus microarray was developed for detecting a wide range of plant viruses. The genus-level probe design protocol was revised for the array first described by (Chou et al., 2006). The probe design protocol automatically searched for the conserved sequences of the virus genus to design a minimal number of conserved 70mer probes that accurately detect plant viruses at the genus level. The validation and application of this microarray were confirmed by testing standard known virus samples, as well as field plant samples, where it detected an unknown *potyvirus* infection in the tested diseased plants.

2. Materials and methods

2.1. Viral sequences

Full-length and partial genome sequences of viral genera were obtained from the NCBI Taxonomy Browser (http://www.ncbi.nlm.nih.gov/Taxonomy/) (Table 1). However, due to the difficulty in collecting virus samples, we selected a few plant virus genera as examples of the application of the microarray designed at the genus level. Choices were made based on these standards: (i) the official quarantine viruses in China, (ii) viruses infecting a wide range of plants with high economic value and whose infections cause significant economic losses, (iii) viruses not detected previously by microarray. For example, *Bean pod mottle virus* (BPMV) is a member of the genus *Comovirus* in the family *Comoviridae*. It is one of the most damaging pathogen of *Fabacaeae* plants and responsible for severe loss to soybean, kidney bean and cowpea. BPMV may cause up to 85% loss in mixed infection with *Soybean mosaic virus* (SBMV) (Giesler et al., 2002). BPMV is an important official quarantine virus in China and can be detected by ELISA and RT-PCR (Gu et al., 2002; Joisson et al., 1993), but no microarray-based method has been reported for its detection. Thus *Comovirus* was studied in this investigation.

Following the above mentioned standards, thirteen virus genera were selected which includes 12 genera of positive single-stranded RNA (ssRNA) viruses and 1 genus of negative ssRNA virus (Table 1). Six of the virus genera have multipartite genomes; the other 7,

monopartite genomes. Since the positive strand RNA of ssRNA negative strand viruses exist in the host total RNA, the positive strand sequences of these viruses were used to design probes. The full-length genome sequences were used to design conserved probes for all genera except the genus *Capillovirus*. As the number of full-length genomic sequences for this genus was less than 5 in the database, partial genomic nucleotide sequences were included to determine the conserved probes. A total of 286 viral sequences representing 169 virus species of 13 viral genera were used to design the conserved probes. Additionally 148,537 viral genome sequences and 57,512 plant mRNA sequences were used to identify off-target hybridizations.

2.2. Microarray probe design

The probe design protocol described previously by Chou et al. (2006) has been revised in this investigation. The protocol searched for a minimal number of conserved probes for each virus genus. The conserved regions of the viral genus sequence were identified by pair-wise BLASTN alignments (Altschul et al., 1990) with other viral sequences in the same genus. Seventy nucleotide segments offset by 5 nt were selected from the highest conserved regions. The 70mers were filtered by (i) GC content between 40% and 60%, (ii) <5 continuous mononucleotides (PolyX), and (iii) <8 nt inner hairpin. The 70mers were then aligned to the viral sequences in the virus genus, and the off-target database by pair-wise BLASTN. It has been reported previously that 70mers with ≥85% similarity and ≥15 nt continuous base pairing with the target sequence are sufficient to generate a detectable signal (He et al., 2005). Thus these cutoffs were set as the threshold in the probe-viral sequence pairing. The off-target probe pairing was set as ≤25 nt base pairing with the other viruses or plant mRNA sequences. Probes identified with off-target probe pairing were removed from the dataset.

The major improvement from the protocol of Chou et al. (2006) in this investigation was an iterative seed searching algorithm. First, the top ranked conserved regions were identified and 70mers selected as described above. Then a seed of two non-overlapping 70mer probes with the best coverage of the virus genus was created. One probe was added to the seed each time to maximize the coverage of the genus. The search was finished when all the viral sequences in the genus were covered by the probes, or the size of the set reached 15. The seed searching was repeated three times so that most viruses in the genus were detected by three probes. If less than three probes were selected in the current conserved region, the probe selection was repeated using the lower ranked conserved

Table 1
Plant virus sequences obtained from NCBI Taxonomy Browser and used to design the 70mer oligonucleotide probes for the microarray.

Virus genus	Genome type	No. of genome sequences	No. of nucleotide sequences	No. of species	No. of probes
ssRNA positive strand viruses					
Capillovirus		2	49[b]	4	14
Carlavirus		15	430	15	29
Comovirus		10	69	5	24
Cucumovirus	M[a]	9	879	3	16
Ilarvirus	M	42	477	14	45
Nepovirus	M	20	479	10	28
Pomovirus	M	12	98	4	25
Potexvirus		26	627	26	33
Potyvirus		48	4599	48	45
Sobemovirus		9	245	9	14
Tobamovirus		22	656	22	31
Tobravirus	M	6	91	3	11
ssRNA negative strand viruses					
Tospovirus	M	18	1011	6	30
Total				169	345

[a] M, multipartite genome.
[b] Full-length genomic sequences were used to design probes, except for the genus *Capillovirus*, where partial nucleotide sequences were used.

Table 2
Viruses isolated from plant tissue samples used to test the performance of the microarray.

Pathogen	Genus	Official quarantine virus in China	Main hosts	Microarrays detection reported previously
Arabis mosaic virus (ArMV)	*Nepovirus*	Y[a]	Bean plants, flowering plants, vegetables, fruit trees	N[b]
Bean pod mottle virus (BPMV)	*Comovirus*	Y	Bean plants	N
Prunus necrotic ringspot virus (PNRSV)	*Ilarvirus*	Y	Stone fruit tree (*Prunus, Rosa*)	Y (Lenz et al., 2008)
Plum pox virus (PPV)	*Potyvirus*	Y	Stone fruit tree (*Prunus, Rosa*)	Y (Lenz et al., 2008; Pasquini et al., 2008)
Potato virus X (PVX)	*Potexvirus*	N	*Amaranthaceae, Cruciferae, Solanaceae*	Y (Abdullahi et al., 2005; Boonham et al., 2003; Bystricka et al., 2003, 2005; Lee et al., 2003)
Southern bean mosaic virus (SBMV)	*Sobemovirus*	Y	Bean plants	N
Tobacco mosaic virus (TMV)	*Tobamovirus*	N	*Solanaceae, Cucurbitaceae, Cruciferae, Leguminosae*	Y (Lee et al., 2003)
Tomato ringspot virus (ToRSV)	*Nepovirus*	Y	Horticulture and flowering plants, fruit trees	N
Tobacco ringspot virus (TRSV)	*Nepovirus*	Y	Legume, melon, flowering plants, fruit trees	N
Tomato spotted wilt virus (TSWV)	*Tospovirus*	Y	Ornamentals, vegetables	N

[a] Y = Yes.
[b] N = No.

regions. The whole probe design protocol is implemented in Perl, and can be provided upon request.

After the probes were selected, a pair-wise BLASTN alignment was performed among the probes and the viral sequences in the genus. The probes were assigned to the probe sets according to the viral sequences that they matched.

2.3. Microarray construction

The 70mer oligonucleotide arrays were synthesized on 75 mm × 25 mm glass slides using a SmartArrayer system (CapitalBio, Beijing, China). All probes were spotted three times on the microarray. The construction of the microarray is shown in Supplementary File 1. Reverse complement sequences of the *capillovirus* and *tospovirus* probes were synthesized to analyze the probe strand effect. Three types of control probes were spotted. The 5′ amino-linker oligonucleotide (5′-TTTTTTTTTTTTTTTTCTCATGCCCATGCCGATGC-3′) was used as the positive control. This positive control was a random sequence which did not hybridize with any virus sequences. The positive control targeting sequence was mixed with the samples before hybridization, and used to monitor the hybridization of the microarray. The spotting buffer was used as the negative control. The internal control was 5′ Hex-linker oligonucleotide (5′-GTCACATGCGATGGATCGAGCTCCTTTATCATCGTTCCCACCTTAATGCA-3′). It was spotted on the microarray to monitor the probes linking to the chips. Spotted slides were air dried for 10 min and then fixed by UV exposure at 240 mJ/cm². After being incubated in a 0.5% sodium dodecyl sulfate (SDS) solution at 60 °C for an hour, slides were rinsed in 100 mM Tris (pH 8.0) for 5 min, and kept moist at 4 °C for use.

2.4. Standard virus samples

Standard virus samples were obtained from the Chinese Academy of Inspection and Quarantine (CAIQ, Beijing, China) and the American Type Culture Collection (ATCC, Manassas, VA, USA) (Table 2). Official quarantine viruses and viruses that were not reported to be detected previously by microarrays were preferentially included in the standard virus study. These viruses were inoculated mechanically into herbaceous host plants. After disease symptoms appeared, infected leaves were collected. The viruses were verified using ELISA and PCR. ELISA was performed by using a commercial kit from ADI, San Antonio, USA, according to the

manufacturer's instructions. After the reactions were completed, OD_{405} of each sample was determined in a micotiter plate reader MULTISKAN ASCENT (THERMO, Vantaa, Finland). ELISA readings were considered positive when the absorbance of the sample wells was at least twofold greater than the mean absorbance of the three healthy controls. For the infected *Panax notoginseng* sample, a universal monoclonal antibody was used in ELISA to detect potyviruses. Reverse Transcription PCR (RT-PCR) was performed by using a commercial kit from TaKaRa, Tokyo, Japan, and performed according to the manufacturer's instructions. BPMV primers were from (Gu et al., 2002). Different PCR primers were designed for the other viruses in the conserved regions of these viruses using Primer 3 (Rozen and Skaletsky, 2000) (Table 3). PCR amplifications were performed for 30 cycles each of 0.5 min at 94 °C, 0.5 min at 55–58 °C according to different primer pairs and 1 min at 72 °C with a final extra 8 min at 72 °C. PCR products were examined by electrophoresis in 1% agarose gels in 0.5× TBE buffer after ethidium bromide staining. The electrophoresis was performed using 100 V for 25 min.

Table 3
PCR primers used to verify the standard virus samples.

Virus	PCR primers
ArMV	ArMVf: 5′-TTGGCAGCGGATTGGGAGTT-3′
	ArMVr: 5′-ATTGGTTCCAGTTGTTAGTGAC-3′
BPMV[a]	BPMVf: 5′-ATAGTTCCATTAGAGGGCGTG-3′
	BPMVr: 5′-AGTGGACCATGCTGTGAGAAAC-3′
PNRSV	PNRSVf: 5′-GGTCCCACTCAGGGCTCAAC-3′
	PNRSVr: 5′-CGCAAAAGTGTCGAAATCTAAATC-3′
PPV	PPVf: 5′-ATCAATGGAATGTGGGTGATG-3′
	PPVr: 5′-GTCTCTTGCACAAGAACTATAACCC-3′
PVX	PVXf: 5′-AGTGGTATGGAACTGGATG-3′
	PVXr: 5′-TTATGGTGGTGGTAGAGTGA-3′
SBMV	SBMVf: 5′-ACAGTCCTGACGCACTCTGAG-3′
	SBMVr: 5′-GTACTCACAGCCGTAGAACTCG-3′
TMV	TMVf: 5′-ATGTCTTACAGTATCACTACTCCATCTCAGTT-3′
	TMVr: 5′-GGTTCGCCTGATTTTCAACTTCTATTAT-3′
ToRSV	ToRSVf: 5′-GACGAAGTTATCAATGGCAGCG-3′
	ToRSVr: 5′-TCCGTCCAATCACGCGAATA-3′
TRSV	TRSVf: 5′-AGTCGGGAAGCTGTATAAACTCA-3′
	TRSVr: 5′-CCCCAAACCAAATATGAACAGT-3′
TSWV	TSWVf: 5′-GCTGGAGCTGAGTATAGCAG-3′
	TSWVr: 5′-AGGATTGGAGCCACTGAC-3′

[a] BPMV primers were from (Gu et al., 2002). The other PCR primers were designed in this investigation, as explained in Section 2.

2.5. Microarray hybridization

Total RNA was extracted from plant samples using TRIzol reagent (Invitrogen, Carlsbad, CA) following the manufacture's protocols. RNA was purified using the NucleotideSpin® RNA clean-up (MN, Duren, Germany). First-strand reverse transcription to cDNA was initiated with a nonamer random primer. The amplified cDNA product was purified with the NucleotideSpin® Extract II (MN, Duren, Germany).

Cy3-dCTP was incorporated into the purified cDNA product with nonamer random primer using the Klenow fragment of DNA polymerase I. Cy3-labeled DNA fragments, pre-heated for 10 min at 80 °C, were added to the microarray for hybridization for 16 h at 42 °C. Slides were washed for 4 min at 42 °C with 2× SSC (1× SSC = 150 mM NaCl, 17 mM sodium citrate, pH 7.0), 0.2% SDS and then washed for 4 min at 42 °C with 0.2× SSC. After being dried by centrifugation, the microarrays were scanned using LuxScan 3.0 (CapitalBio, Beijing, China). Microarray image acquisition and signal analyses were performed using LuxScan 3.0 software.

The sensitivity of microarray detection was evaluated using the total RNA of a *Chenopodium quinoa* leaf infected with PVX. The concentration of RNA was determined by UV spectrophotometer as 20 ng/μl. The RNA was serially diluted 10^0, 10^1 and 10^2, and used in the microarray hybridization and RT-PCR. The hybridization results of the three dilutions were compared to determine the detection sensitivity.

Three sample amplification protocols were evaluated in the study. The first protocol was described above by using a nonamer random primer. The second protocol involved reverse transcribing the RNA with a published primer, PrimerD (5′-GTTTCCCAGTAGGTCTCNNNNNNNNN-3′) (Wang et al., 2002). The resulting product was random-primed with nonamer using Klenow polymerase in the presence of Cy3-dCTP. The third protocol involved the reverse transcription of the RNA with PrimerD and second-strand DNA synthesis was subsequently carried out with Sequenase. The resulting product was random-primed with nonamer by using Klenow polymerase in the presence of Cy3-dCTP.

The microarray data were submitted to GEO (http://www.ncbi.nlm.nih.gov/geo/) with the series accession number of GSE15837.

2.6. Virus identification

The virus identification protocol was adapted from the protocol first described by (Chou et al., 2006). First, the positive probes were identified as those with a feature signal intensity more than three times that of the background intensity, and feature intensity minus background intensity should be more than 800. The genus signal strength is calculated as the sum of the signal intensities of all the positive probes in one given genus divided by the maximum sum of the signal intensities of positive probes of every genus. The species signal strength is calculated as the average signal intensity of the multiple probes for a given virus divided by the maximum average signal intensity of the multiple probes for every virus on the microarray. The genus and species signal strength are both expressed as a percentage. The best prediction is the virus genus or species with the highest percentage, e.g. 100% or 1.

2.7. Detection of and characterization of an unknown viruses

P. notoginseng is a perennial root herb and a rare Chinese medicinal plant with high economic importance. Virus infection has significantly damaged the product quality and quantity of this plant in recent years. There was a severe disease outbreak in *P. notoginseng* farms in Yunnan Province, southwest China in 2001, where the incidence of the disease increased to 60%. Diseased *P. notoginseng* plants showed mosaic and deformation symptoms. These diseased plants were screened using the genus level microarray in order to determine possible infection of these plants with a virus whose genus probes were constructed on the microarray.

Microarray hybridization and data analysis protocol were described above. If an unknown virus was identified from a plant sample, it was sequenced and blasted against the NCBI database. Total RNA was extracted from infected plant samples as described by the manual of the TRIzol reagent (Invitrogen, Carlsbad, CA). First-strand cDNA was synthesized using Moloney Murine Leukemia virus reverse transcriptase (Promega, Madison, USA) according to the manufacturer's instructions and with reverse primer as the initial primer. For the unknown virus identified from *P. notoginseng*, 3′-Rapid Amplification of cDNA Ends (3′-RACE) was conducted using the potyvirus degenerate primer (5′-GGXAAYAAYAGYGGXCAZCC-3′, X = A, G, C or T; Y = T or C; Z = A or G). The 3′-RACE protocol was as described (Chen and Adams, 2001). After the first-strand synthesis, the cDNA was amplified by PCR using the virus specific or degenerate primers (Table 3). The PCR products were gel-purified using Gel Extraction Kit (Takara, Tokyo, JAPAN) according to the manufacture's protocol. The purified PCR products were cloned into the pMD18-T vector (Takara, Tokyo, Japan), and then screened with blue-white selection. The clones were confirmed by PCR, and then sent for sequencing to BioAsia Company, Shanghai, China. The sequencing result of the unknown virus identified from *P. notoginseng* was submitted to GenBank with accession number of FJ816101.

3. Results

3.1. Construction of the microarray

The general probe design protocol used was to screen the probes by GC contents, Tm, inner hairpin and off-target hybridization. Subsequently, the probes were ranked according to their hybridization specificity, and also by possible cross hybridization to off-target sequences (Fig. 1). A rigorous seed searching was performed on all the available probes. The selected probe set should utilize a minimal number of probes to attain a maximal coverage of species of the virus genus. This design protocol was implemented in a Perl script which can be automated for all virus genera.

Three hundred and forty five probes were designed for 169 virus species in 13 genera (Table 1). More than 88% of the viral sequences could be identified by more than 3 probes. The top conserved probes were usually selected, which ensured the viruses were covered by a minimal number of probes. This resulted in 2.04 probes per virus for 169 viruses, which is significantly fewer than what was published previously for microarray-based virus detection (Palacios et al., 2007; Wang et al., 2002). The top selected 345 probes were
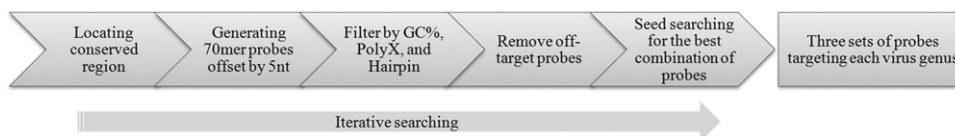

Fig. 1. Probe design workflow of the microarray-based plant virus detection.
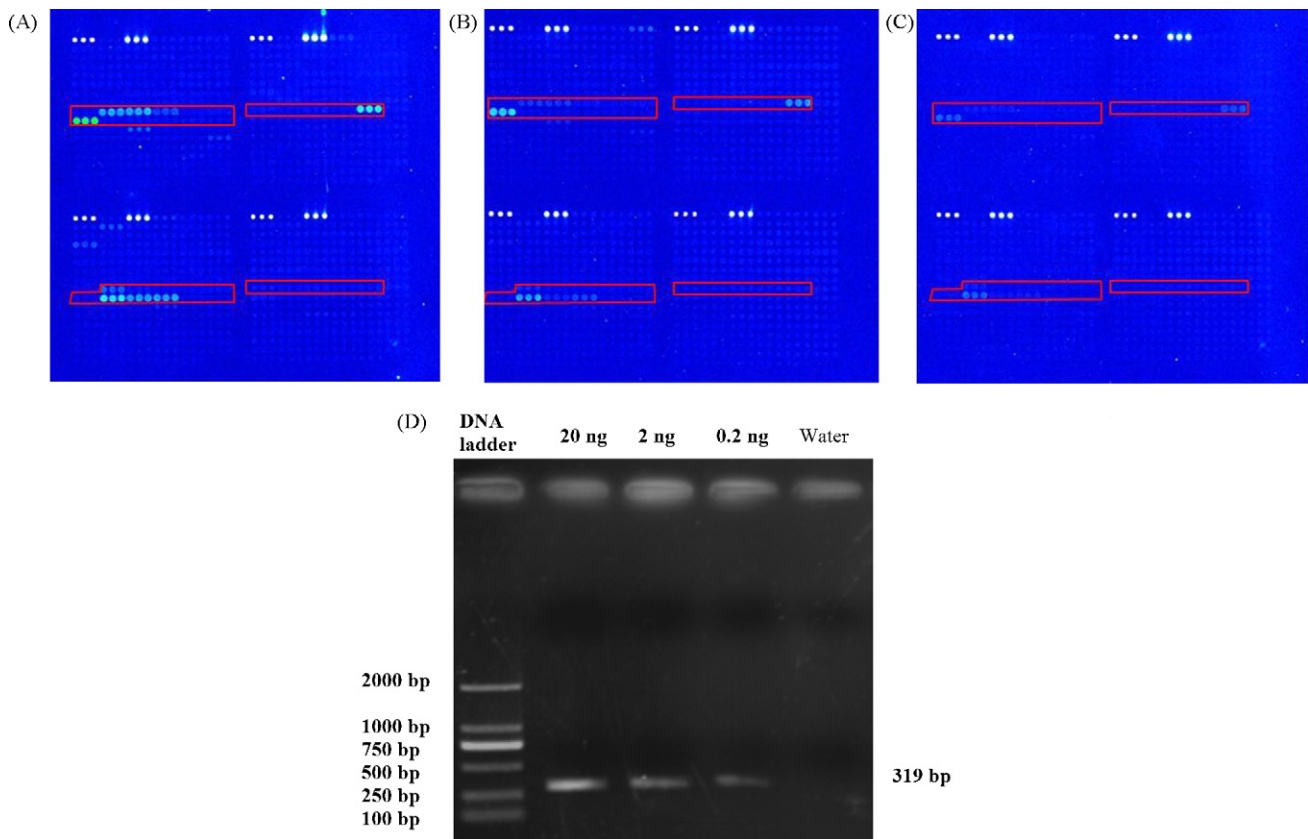
Fig. 2. Sensitivity test of PVX with RNA dilutions of (A) 10^0, (B) 10^1 and (C) 10^2 fold using microarray, and (D) RT-PCR. The array image is a computer generated pseudo-color image. The construction of the microarray is described in Supplementary File 1. The target viruses of each probe are described in the data table of GEO platform GPL8484. Probes targeting *Potexvirus* were highlighted with red rectangles. Probes not included in the red rectangles were probes targeting the other virus genera. Some probes targeting *Potexvirus* were not positive, as can be seen in the lower right rectangle of each image, because they may target the other viruses, other than PVX, in the genus *Potexvirus*.

spotted on the microarray as described in Section 2, and tested for microarray performance.

3.2. Evaluation of the microarray performance

The microarray detection sensitivity was evaluated using total RNA of *C. quinoa* leaves infected with PVX, as described in Section 2. The top genus and species predicted for RNA dilutions of 10^0 and 10^1 were *Potexvirus* and PVX, respectively (Fig. 2). For RNA dilutions of 10^2 fold, no probe gave a positive signal. The microarray detection threshold was about 2 ng/μl of total RNA. The RT-PCR amplicons of the three dilutions were all visible on the agarose gel (Fig. 2).

The microarray performance was then tested using three different amplification protocols as described in Section 2. Total RNA extracted from *D. stramonium* infected with TSWV was used for this purpose. The top genus predictions using the three protocols were all *Tospovirus*. However, the first protocol of RNA reverse transcribed using random nonamer primers achieved the largest number of positive probes and the highest signal intensities for TSWV. Thus, the first protocol was used subsequently.

The effect of using reverse complement probes was evaluated by including negative strand probes of *Tospovirus* on the microarray, which are the reverse complement of probes designed in this study. Total RNA from *D. stramonium* infected with TSWV was used in the evaluation. Using positive strand probes and negative strand probes, the top genus and species prediction both were *Tospovirus* and TSWV. However, the number of positive probes and signal intensities for positive strand probes were higher than negative

strand probes. Thus the negative strand probes were not included in the following study.

Several standard virus samples provided by the CAIQ and ATCC were screened using the microarray (Table 2). The virus samples included both ssRNA positive and negative strand viruses with monopartite and multipartite genomes. All virus samples were correctly identified at the genus level, except SBMV (Table 4). SBMV belongs to the genus *Sobemovirus* which was predicted as the second best prediction.

At the species level, SBMV, TMV and PNRSV were not the best predictions for their genera *Sobemovirus*, *Tobamovirus* and *Ilarvirus*, respectively (Table 4). However, *Paprika mild mottle virus* (PaMMV) and *American plum line pattern virus* (APLPV) were the best species prediction for *Tobamovirus* and *Ilarvirus*, respectively. These two predictions were also correct at the genus level. Thus for prediction, nine out of ten cases were correct at the genus level and seven out of ten cases were correct at the species level. We selected two viruses as examples to illustrate the probe design and data analysis in detail.

TRSV was identified accurately at both the genus and species levels from the collected infected leaf samples. The three probes targeting TRSV were designed from the conserved 3′-region of TRSV RNA 1 and RNA 2 (Fig. 3). All three probes were identified as positive in the microarray result (Fig. 4).

The prediction for PNRSV was correct at the genus level as *Ilarvirus*, but PNRSV was the third best at the species level. Two probes were designed for PNRSV, and one of them was positive in the microarray experiment. While 45 probes were designed for *Ilarvirus*, four probes were positive. PNRSV was among the top best species predictions.

Table 4
Genus prediction analyses for the standard virus samples.

Standard virus samples		Genus prediction score		Contradictory genus prediction			Species prediction score		Contradictory species prediction		
Species	Genus	Rank	Relative signal strength	Genus	Rank	Relative signal strength	Rank	Relative signal strength	Species (Genus)	Rank	Relative signal strength
BPMV	*Comovirus*	1	1				1	1			
SBMV	*Sobemovirus*	2	0.483	*Comovirus*	1	1	3	0.673	BPMV (*Comovirus*)	1	1
ArMV	*Nepovirus*	1	1				1	1			
TMV	*Tobamovirus*	1	1				4	0.894	PaMMV[a] (*Tobamovirus*)	1	1
PPV	*Potyvirus*	1	1				1	1			
ToRSV	*Nepovirus*	1	1				1	1			
PNRSV	*Ilarvirus*	1	1				3	0.703	APLPV[b] (*Ilarvirus*)	1	1
PVX	*Potexvirus*	1	1				1	1			
TSWV	*Tospovirus*	1	1				1	1			
TRSV	*Nepovirus*	1	1				1	1			

[a] PaMMV: Paprika mild mottle virus.
[b] APLPV: American plum line pattern virus.

3.3. Analysis of unknown field samples

Diseased plant samples were collected from a severe disease outbreak in *P. notoginseng* farms in Yunnan, China, in 2001, as described in Section 2. Several molecular diagnosis protocols were applied on the collected plant samples, including ELISA and PCR. No virus infection was detected in the samples using these methods. A virus screening was performed on these diseased samples using the microarray platform developed in this study. *Potyvirus* was the best virus genus prediction and PVY was the best species prediction.

Sequence analysis of this detected virus was performed. A sequence of 1866 nucleotides excluding the polyA tail at the 3′-terminal region of the genome was determined using 3′-RACE amplification. The sequence included the 3′-terminal part of the NIb, the coat protein, the stop codon and the 3′-untranslated region (UTR) (Fig. 5). Several features in the sequence revealed that the virus is a species of the genus *Potyvirus* as indicated by the cleavage site of the NIb protein and the presence of a potyvirus-like coat protein. The RdRp active site of the NIb protein was SGQPSTVVDNTLMVVIAMTYALLKMGFPAEQHDE-VCQYFVNGDD. The 3′-UTR sequence had high AT content, and included inverted repetitive sequences "GAGGN/N′CCTC". According to a BLAST homology search to the NCBI database, the closest virus to the unknown potyvirus was an unclassified virus identified in *P. notoginseng* in 2006, sharing 98% identity at the protein level and 92% identity at the nucleotide level. The other close virus species were *Angelica virus Y*, *Carrot virus Y*, *Celery mosaic virus* and *Apium virus Y*.
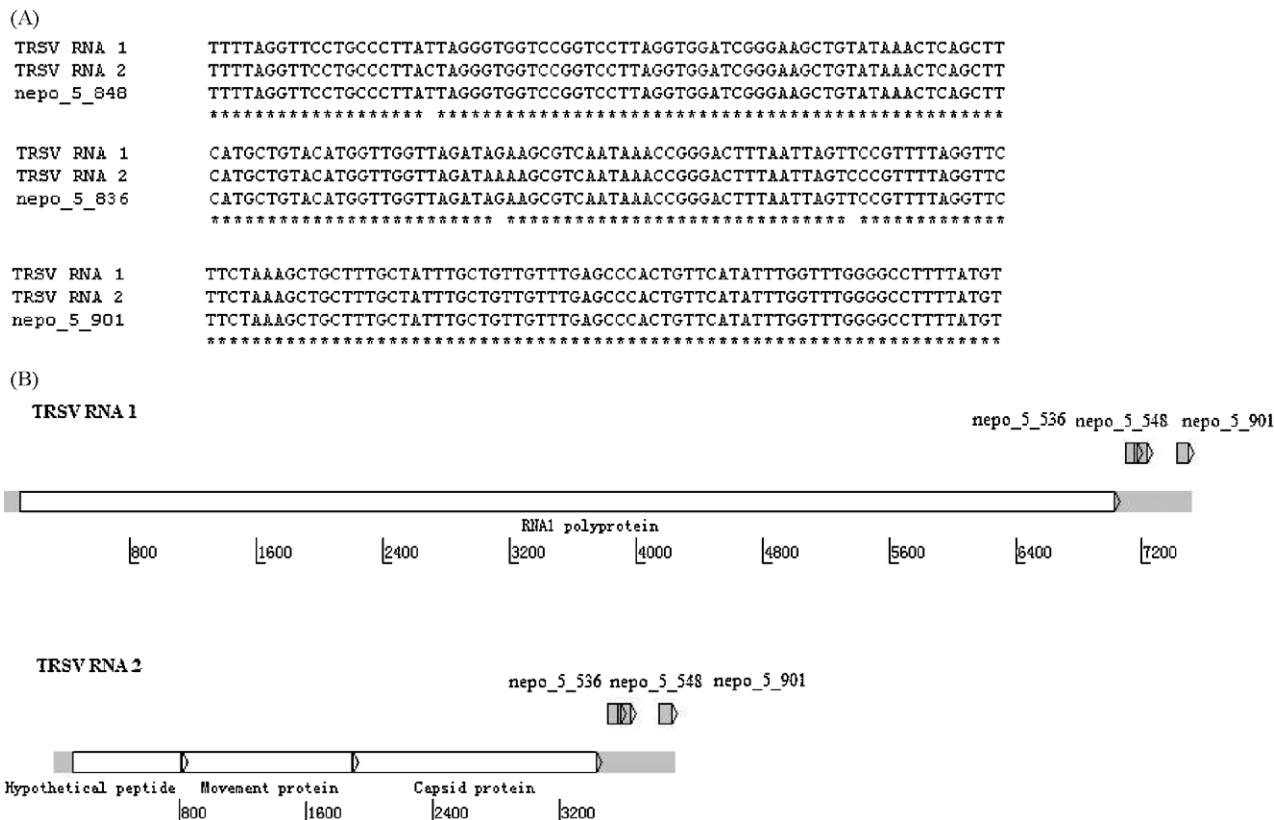
(A)

```
TRSV RNA 1   TTTTAGGTTCCTGCCCTTATTAGGGTGGTCCGGTCCTTAGGTGGATCGGGAAGCTGTATAAACTCAGCTT
TRSV RNA 2   TTTTAGGTTCCTGCCCTTACTAGGGTGGTCCGGTCCTTAGGTGGATCGGGAAGCTGTATAAACTCAGCTT
nepo_5_848   TTTTAGGTTCCTGCCCTTATTAGGGTGGTCCGGTCCTTAGGTGGATCGGGAAGCTGTATAAACTCAGCTT
             *******************  *************************************************

TRSV RNA 1   CATGCTGTACATGGTTGGTTAGATAGAAGCGTCAATAAACCGGGACTTTAATTAGTTCCGTTTTAGGTTC
TRSV RNA 2   CATGCTGTACATGGTTGGTTAGATAAAAGCGTCAATAAACCGGGACTTTAATTAGTCCCGTTTTAGGTTC
nepo_5_836   CATGCTGTACATGGTTGGTTAGATAGAAGCGTCAATAAACCGGGACTTTAATTAGTTCCGTTTTAGGTTC
             *************************  *************************  ********  **************

TRSV RNA 1   TTCTAAAGCTGCTTTGCTATTTGCTGTTGTTTGAGCCCACTGTTCATATTTGGTTTGGGGCCTTTTATGT
TRSV RNA 2   TTCTAAAGCTGCTTTGCTATTTGCTGTTGTTTGAGCCCACTGTTCATATTTGGTTTGGGGCCTTTTATGT
nepo_5_901   TTCTAAAGCTGCTTTGCTATTTGCTGTTGTTTGAGCCCACTGTTCATATTTGGTTTGGGGCCTTTTATGT
             ************ *************************************************************
```

(B)

Fig. 3. Sequence alignment of TRSV and schematic view of genome structure. (A) Probe sequence alignment with TRSV RNA 1 and RNA 2. (B) Genome structure of the TRSV and location of the probes specific to TRSV.
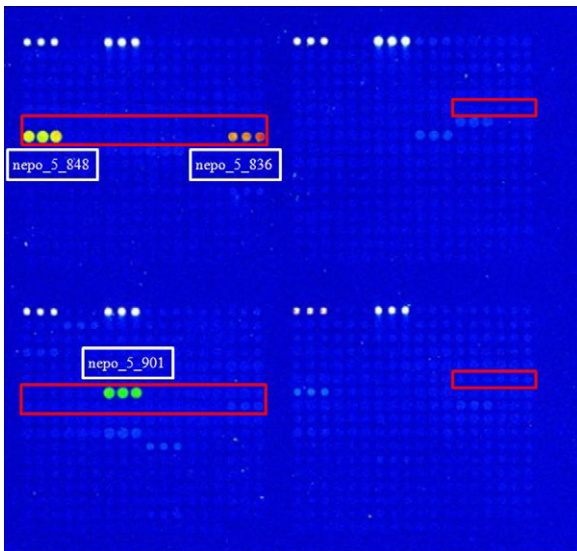
Fig. 4. Detection of TRSV using the microarray. The three probes targeting TRSV were nepo_5_848, nepo_5_836 and nepo_5_901, and all hybridized positively to the virus. The array image is a computer generated pseudo-color image. Probes targeting *Nepovirus* were highlighted with red rectangles. Probes not included in the red rectangles were probes targeting the other virus genera. Some spots targeting the other virus genera were visible on the image, but they were all below the detection threshold, as explained in the Section 2. Some probes targeting *Nepovirus* were not positive, as can be seen in the lower right rectangle of each image, because they may target the other viruses, other than TRSV, in the genus *Nepovirus*.

4. Discussion

Detection of viral infections is a major challenge in disease identification and control in plants. The existing methods, such as ELISA, molecular hybridization and PCR, are labor intensive and less efficient when screening for a large range of known viruses (James et al., 2006; Mumford et al., 2006). To overcome the shortage in these methods, we have designed a newly developed plant virus detection microarray platform at the virus genus level. The probe design strategy involved selecting conserved 70mer probes, using the minimal number of probe combinations.

ViroChip and GreeneChipVr are two well known virus detection microarray platforms, able to detect 140 and 1710 vertebrate viruses respectively (Palacios et al., 2007; Wang et al., 2002). Yet, only a few microarrays have been reported for detecting plant virus infections at the genus level. The largest number of plant genera detected using microarray, as reviewed (Boonham et al., 2007), was six potato viruses from four genera (Bystricka et al., 2005). Another microarray was constructed to detect seven fruit tree viruses from five genera (Lenz et al., 2008). Our microar-

ray platform was designed to detect 169 plant virus species in 13 genera, greatly expanding the detection of plant viruses by microarrays.

The selection of probes is a major concern in the microarray design. A large number of probes may provide good statistical power, but there will be a certain amount of redundancy of the probes and the cost of the microarray will be increased. Chou et al. (2004) showed that a probe set of three 70mer probes will generate consistent signals and reliable results. Therefore, by generating a minimal number of three probes targeting each virus genus, we aimed to achieve a consistent detection result while reducing the redundancy of the probes. For screening unknown viruses, it might be useful to use 4 or 5 probes per virus, to minimize the risk of not detecting a novel virus in the genus.

About 1600 probes were used for 140 virus species in ViroChip (Wang et al., 2002). The average number of probes per virus was 11.43. About 9500 probes were used for 1710 virus species in the GreeneChipVr (Palacios et al., 2007), which had an average of 5.54 probes per virus. Our microarray achieved a ratio of 2.04, which is considerably less than the above two virus detection microarrays. This ratio is smaller than three, since the top conserved probes were always selected in serial searching.

Currently, about 80 plant virus genera are reported in ICTVdb (http://www.ncbi.nlm.nih.gov/ICTVdb/) containing approximately 1000 virus species with published genome sequences. Using our probe design protocol, only 2400 probes are needed to detect all the plant virus genera (data not shown). Thus it should be feasible to detect all the plant virus genera with a single microarray using approximately 2400 probes.

In the sensitivity test of the microarray, an RNA dilution of 1:10 was detected by the microarray. While the microarray showed a negative result for the dilution of 1:100, RT-PCR detected the 1:100 dilution. This sensitivity of this microarray is similar to previous microarray sensitivities (Boonham et al., 2003, 2007). The level of microarray sensitivity was roughly equivalent to TAS-ELISA, 10-fold higher than conventional ELISA, although still much lower than PCR. Agindotan and Perry (2007) had PLRV being detected in RNA dilution of 1:100 using their nylon membrane based macroarray. As shown in their report, this sensitivity is comparable to DAS-ELISA. Thus refinements of the microarray target amplification procedure are needed in future to increase the sensitivity of microarray detection, and to prevent possible false negative detection.

In the microarray performance test using standard viruses, only BPMV was not correctly identified at the genus level. The best genus prediction for BPMV was *Comovirus*, with the top two species predictions as SBMV RNA 1 and RNA2 (*Comovirus*). BPMV was the second best at the genus and the third best species prediction. This could potentially be explained by multiple viral infections in the plant, or due to a mix of the samples by chance during the microarray experiment. In future development of the virus
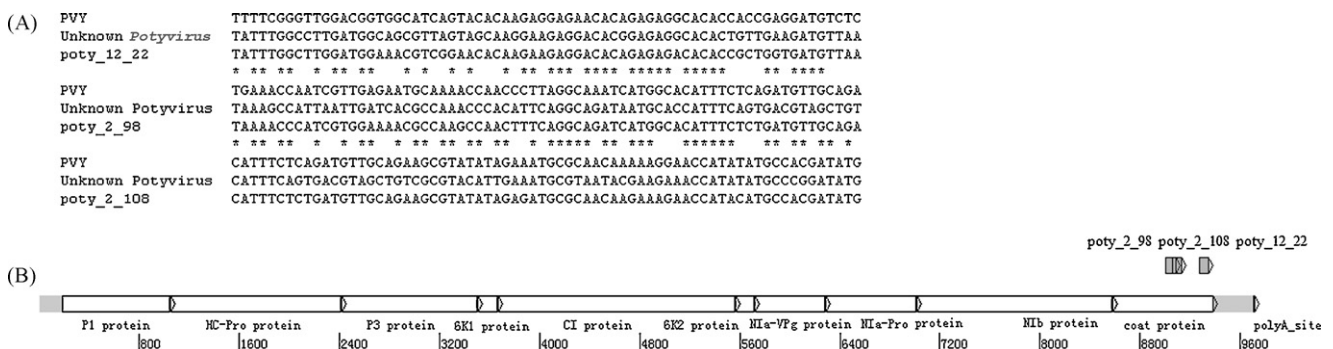

Fig. 5. Sequence alignment of the unknown *Potyvirus* and schematic view of genome structure. (A) Probe sequence alignment with PVY and the unknown *Potyvirus* detected from *P. notoginseng* samples. (B) Genome structure of the unknown *Potyvirus* and location of the probes. The genome structure and coordinates were simulated from PVY.

detection microarray, multiple virus infections will be taken into consideration.

Virus screening of field samples using the genus level microarray was particularly rewarding. For the *P. notoginseng* samples showing disease symptoms, while the traditional serotype based method, ELISA, failed to detect any virus, the microarray identified the unknown virus as a species of the genus *Potyvirus*. None of the closest viruses, e.g. the unclassified viruses identified in *P. notoginseng*, *Angelica virus Y*, and *Carrot virus Y*, had published genomic sequences. Thus no sequence from these viruses was used to design the conserved probes for the genus *Potyvirus* in the microarray. It suggests that the detection microarray system does not need prior knowledge of newly emerged viruses, thus it will not be limited by selection of specific primer sequences (i.e. PCR) or antibodies (i.e. ELISA). This test highlights the potential use of the microarray in detecting newly emerging plant viruses. This is similar to the detection of SARS as a novel coronavirus using the ViroChip prior to the sequencing of the SARS genome sequence (Wang et al., 2003).

In conclusion, we have developed a 70mer oligonucleotide microarray for detection and identification of 13 genera of plant viruses. This microarray detected the largest number of plant virus genera reported so far. We also aimed to use a minimal number of probes, thus minimizing the cost and probe redundancy of the microarray. The microarray platform developed in this study illustrated the feasibility of detecting a wide range of plant viruses at the genus level, and can be used to detect unknown emerging viruses. This technology can be applied to other plant and vertebrate virus genera. This diagnostic tool will certainly have an important use for discovery and identification of plant viruses.

Acknowledgments

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jviromet.2010.03.010.

References

Abdullahi, I., Koerbler, M., Stachewicz, H., Winter, S., 2005. The 18S rDNA sequence of *Synchytrium endobioticum* and its utility in microarrays for the simultaneous detection of fungal and viral pathogens of potato. Appl. Microbiol. Biotechnol. 68, 368–375.

Agindotan, B., Perry, K.L., 2007. Macroarray detection of plant RNA viruses using randomly primed and amplified complementary cDNAs from infected plants. Phytopathology 97, 119–127.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Boonham, N., Tomlinson, J., Mumford, R., 2007. Microarrays for rapid identification of plant viruses. Annu. Rev. Phytopathol. 45, 307–328.

Boonham, N., Walsh, K., Smith, P., Madagan, K., Graham, I., Barker, I., 2003. Detection of potato viruses using microarray technology: towards a generic method for plant viral disease diagnosis. J. Virol. Methods 108, 181–187.

Bystricka, D., Lenz, O., Mraz, I., Dedic, P., Sip, M., 2003. DNA microarray: parallel detection of potato viruses. Acta Virol. 47, 41–44.

Bystricka, D., Lenz, O., Mraz, I., Piherova, L., Kmoch, S., Sip, M., 2005. Oligonucleotide-based microarray: a new improvement in microarray detection of plant viruses at the genus level. J. Virol. Methods 128, 176–182.

Chen, J., Adams, M.J., 2001. A universal PCR primer to detect members of the Potyviridae and its use to examine the taxonomic status of several members of the family. Arch. Virol. 146, 757–766.

Chou, C.C., Chen, C.H., Lee, T.T., Peck, K., 2004. Optimization of probe length and the number of probes per gene for optimal microarray analysis of gene expression. Nucleic Acids Res. 32, e99.

Chou, C.C., Lee, T.T., Chen, C.H., Hsiao, H.Y., Lin, Y.L., Ho, M.S., Yang, P.C., Peck, K., 2006. Design of microarray probes for virus identification and detection of emerging viruses at the genus level. BMC Bioinformatics 7, 232.

Giesler, L.J., Ghabrial, S.A., Hunt, T.E., Hill, J.H., 2002. Bean pod mottle virus: a threat to U.S. soybean production. Plant Disease, 86.

Gu, H., Clark, A.J., de Sa, P.B., Pfeiffer, T.W., Tolin, S., Ghabrial, S.A., 2002. Diversity among isolates of bean pod mottle virus. Phytopathology 92, 446–452.

He, Z., Wu, L., Li, X., Fields, M.W., Zhou, J., 2005. Empirical establishment of oligonucleotide probe design criteria. Appl. Environ. Microbiol. 71, 3753–3760.

James, D., Varga, A., Pallas, V., Candresse, T., 2006. Strategies for simultaneous detection of multiple plant viruses. Can. J. Plant Pathol. 28, 16–29.

Joisson, C., Kuster, F., Plaue, S., Van Regenmortel, M.H., 1993. Antigenic analysis of bean pod mottle virus using linear and cyclized synthetic peptides. Arch. Virol. 128, 299–317.

Lee, G.P., Min, B.E., Kim, C.S., Choi, S.H., Harn, C.H., Kim, S.U., Ryu, K.H., 2003. Plant virus cDNA chip hybridization for detection and differentiation of four cucurbit-infecting Tobamoviruses. J. Virol. Methods 110, 19–24.

Lenz, O., Petrzik, K., Spak, J., 2008. Investigating the sensitivity of a fluorescence-based microarray for the detection of fruit-tree viruses. J. Virol. Methods 148, 96–105.

Mumford, R., Boonham, N., Tomlinson, J., Barker, I., 2006. Advances in molecular phytodiagnostics—new solutions for old problems. Eur. J. Plant Pathol. 116, 1–19.

Palacios, G., Quan, P.L., Jabado, O.J., Conlan, S., Hirschberg, D.L., Liu, Y., Zhai, J., Renwick, N., Hui, J., Hegyi, H., Grolla, A., Strong, J.E., Towner, J.S., Geisbert, T.W., Jahrling, P.B., Buchen-Osmond, C., Ellerbrok, H., Sanchez-Seco, M.P., Lussier, Y., Formenty, P., Nichol, M.S., Feldmann, H., Briese, T., Lipkin, W.I., 2007. Panmicrobial oligonucleotide array for diagnosis of infectious diseases. Emerg. Infect. Dis. 13, 73–81.

Pasquini, G., Barba, M., Hadidi, A., Faggioli, F., Negri, R., Sobol, I., Tiberini, A., Caglayan, K., Mazyad, H., Anfoka, G., Ghanim, M., Zeidan, M., Czosnek, H., 2008. Oligonucleotide microarray-based detection and genotyping of Plum pox virus. J. Virol. Methods 147, 118–126.

Rozen, S., Skaletsky, H., 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol. Biol. 132, 365–386.

Wang, D., Coscoy, L., Zylberberg, M., Avila, P.C., Boushey, H.A., Ganem, D., DeRisi, J.L., 2002. Microarray-based detection and genotyping of viral pathogens. Proc. Natl. Acad. Sci. U.S.A. 99, 15687–15692.

Wang, D., Urisman, A., Liu, Y.T., Springer, M., Ksiazek, T.G., Erdman, D.D., Mardis, E.R., Hickenbotham, M., Magrini, V., Eldred, J., Latreille, J.P., Wilson, R.K., Ganem, D., DeRisi, J.L., 2003. Viral discovery and sequence recovery using DNA microarrays. PLoS Biol. 1, E2.

Wei, T., Pearson, M.N., Blohm, D., Nolte, M., Armstrong, K., 2009. Development of a short oligonucleotide microarray for the detection and identification of multiple potyviruses. J. Virol. Methods 162, 109–118.