Virology 388 (2009) 1–7

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Rapid Communication

Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: A generic method for diagnosis, discovery and sequencing of viruses

Jan F. Kreuze [a,d,*], Ana Perez [c], Milton Untiveros [a], Dora Quispe [a], Segundo Fuentes [c], Ian Barker [c], Reinhard Simon [b]

[a] Applied Biotechnology Laboratory, Germplasm Enhancement and Crop Improved Division, International Potato Center (CIP), Apartado 1558, Lima 12, Peru
[b] Research Informatics Unit, Germplasm Enhancement and Crop Improvement Division, International Potato Center (CIP), Apartado 1558, Lima 12, Peru
[c] Virology laboratory, Integrated Crop Management Division, International Potato Center (CIP), Apartado 1558, Lima 12, Peru
[d] Department of Plant Biology and Forest Genetics, Swedish University of Agricultural Sciences (SLU), P.O. Box 7080, SE-750 07 Uppsala, Sweden

ARTICLE INFO

Article history:
Received 8 January 2009
Returned to author for revision
5 March 2009
Accepted 25 March 2009
Available online 23 April 2009

Keywords:
siRNA
Deep sequencing
Viral diagnostics
Plant viruses
Virus discovery
Sweetpotato
Badnavirus
Mastrevirus
Potyvirus
Crinivirus
Short sequence assembly

ABSTRACT

We report the first identification of novel viruses, and sequence of an entire viral genome, by a single step of high-throughput parallel sequencing of small RNAs from diseased, as well as symptomless plants. Contigs were assembled from sequenced total siRNA from plants using small sequence assembly software and could positively identify RNA, ssDNA and dsDNA reverse transcribing viruses and in one case spanned the entire genome. The results present a novel approach which cannot only identify known viral pathogens, occurring at extremely low titers, but also novel viruses, without the necessity of any prior knowledge.

Introduction

In the field of agriculture, crop losses due to emerging plant diseases, including those of viral origin, are of great current concern particularly in developing countries (Anderson et al., 2004). Strategies to combat plant disease outbreaks often involve early intervention either to stop diseases spreading or to prevent their introduction in the first place. Early detection of the appearance of disease or disease causing agents followed by rapid and accurate identification of these agents is essential if correct control measures are to be deployed. This is particularly true for entirely new diseases where novel control strategies may have to be developed alongside characterization of novel agents. The identification of a virus typically requires the application of a number of methods including physical, biological, serological and molecular methods. Traditional generic methods for identifying and characterizing novel virus diseases include the use of techniques such as electron microscopy or the use of indicator plants as bioassays. These methods are limited in their scope and only permit partial characterization of novel agents. Some nucleic acid based techniques such as dsRNA extraction (Balijja et al., 2008 and references therein) or rolling circle amplification (Haible et al., 2006), although powerful in identifying unknown viral agents, are limited in their scope to RNA or circular DNA viruses respectively. Recently technologies such as diagnostic microarrays and mass spectrometry have been proposed as generic tools for identifying viruses (Mumford et al., 2006; Boonham et al., 2008) although all require some prior knowledge of the agents to be identified. With the advent of next generation high-throughput parallel sequencing platforms, the possibility of random metagenomic sequencing of diseased samples to identify putative pathogens has emerged (Cox-Foster et al., 2007; Quan et al., 2008 and references therein). However elimination of host nucleic acid in these systems is critical to boosting pathogen signals toward the detection threshold, and extremely low titre viruses may be missed with the currently available methods. Enrichment of viral nucleic acids by partial purification of virus-like particles (Melcher et al. 2008; Victoria et al., 2008) partially achieve

* Corresponding author. Germplasm Enhancement and Crop Improvement Division, International Potato Center (CIP), Apartado 1558, Lima 12, Peru. Fax: +51 1 317 5326.
E-mail address: j.kreuze@cgiar.org (J.F. Kreuze).

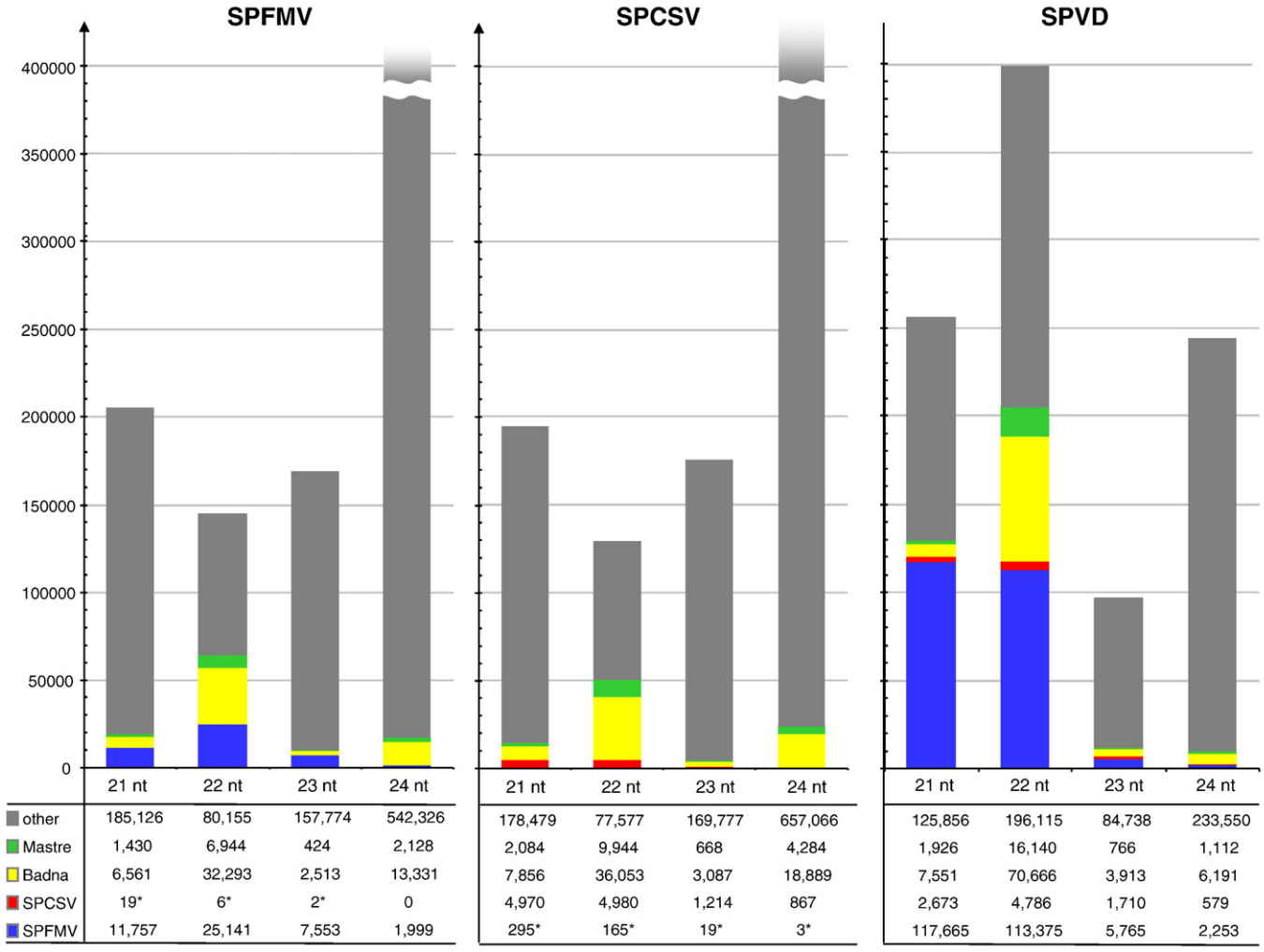
0042-6822/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2009.03.024

this, but are still not entirely generic, since viruses with unstable particles or non-encapsidated agents such as viroids or certain virus strains (e.g. *Tobacco rattle virus* NM strains) cannot be isolated. We hypothesized that it would be possible to identify viruses based on the sequences of viral defense related molecules in plants. RNA silencing (RNAi) is a cytoplasmic cell surveillance system to recognize double stranded RNA and specifically destroy single and double stranded RNA molecules homologous to the inducer, using small interfering RNAs (siRNA) as a guide (Mlotshwa et al., 2008 for a recent review). Viruses are both inducers and targets of RNAi that constitutes a fundamental antiviral defence mechanism in eukaryotic organisms (recently reviewed by Haasnoot et al., 2007a). It is particularly important in plants that use RNAi to recover from virus disease (Covey et al., 1997). We describe the use of high-throughput sequencing of small RNAs (sRNA) from plants to successfully identify the viruses infecting them, including previously unknown viruses, even in extremely low titre symptomless infections.

Results and discussion

In an experimental setup sweetpotato plants were infected with two different RNA viruses individually or in combination. One of the viruses was sweetpotato feathery mottle virus (SPFMV; family *Potyviridae*), an often symptomless potyvirus, which frequently occurs in titres below immunological detection limits when infecting sweetpotato alone. The other virus was sweetpotato chlorotic stunt virus (SPCSV; family *Closteroviridae*), a phloem-limited crinivirus causing mild symptoms, whereas the two viruses infecting together cause the severe synergistic sweetpotato virus disease (SPVD); the single most important disease of sweetpotato worldwide (Karyeija et al., 2000). While the experiment was originally set up to analyze the effects of SPVD on RNA silencing and siRNA accumulation in sweetpotato, we tested if the data obtained could be used also for diagnostic purposes, which is the topic of this manuscript.

We isolated and sequenced sRNAs from infected sweetpotato plants using the Illumina deep sequencing platform. Between 1 and 1.2 million sRNA (1–28 nt) reads were obtained from each of the three different samples and can be downloaded from http://research.cip.cgiar.org/confluence/display/cpx/CIP.sweetpotato.2008.PER.CIPHQ.siRNA-1 . The far majority of sequences were between 21 and 24 nts in size. Sequences smaller or larger than this represented 3% or less of the total. Approximately ~68% of all the reads were unique in the SPCSV and SPFMV infected samples, whereas this was reduced to ~53% in SPVD affected samples, which was also



	SPFMV				SPCSV				SPVD			
	21 nt	22 nt	23 nt	24 nt	21 nt	22 nt	23 nt	24 nt	21 nt	22 nt	23 nt	24 nt
other	185,126	80,155	157,774	542,326	178,479	77,577	169,777	657,066	125,856	196,115	84,738	233,550
Mastre	1,430	6,944	424	2,128	2,084	9,944	668	4,284	1,926	16,140	766	1,112
Badna	6,561	32,293	2,513	13,331	7,856	36,053	3,087	18,889	7,551	70,666	3,913	6,191
SPCSV	19*	6*	2*	0	4,970	4,980	1,214	867	2,673	4,786	1,710	579
SPFMV	11,757	25,141	7,553	1,999	295*	165*	19*	3*	117,665	113,375	5,765	2,253

Fig. 1. Size distribution of virus specific 21–24 nt small RNAs in sweetpotato plants infected by SPFMV, SPCSV and both viruses (SPVD). Total amount of siRNA homologous to different viruses are shown by different colours in the bar graphs according to legend. Scale indicates total amount of sRNAs. Broken and fading bars on 24 nt sequences indicate bar heights go beyond the scale of the figure. The exact amounts of identified sRNAs for each size class and virus are presented in the aligned table below the graphs. Other: sRNAs without similarity to identified viruses; Mastre: approximate Mastrevirus specific sRNAs (obtained by multiplying the number of hits obtained for half [~2.3 kb] the genome sequence by two); Badna: Badnavirus specific sRNAs (corresponding to near complete [~7.3 and ~7.6 kb] genomes of the both badnaviruses). *To compensate for natural sequence variation of the viruses and sequencing errors, up to two nucleotide mismatches were permitted, therefore a few low complexity hits are also identified in samples not infected by the corresponding virus and are considered unspecific.

accompanied by a dramatic relative reduction of the 24 nt RNA in favour of the 21 and 22 nt size classes (Fig. 1). sRNA sequences could be assembled into contigs of up to more than 1000 nts using any of the three tested short sequence assembly programs. The longest contig of 3115 nts corresponding to SPFMV was obtained with VCAKE. VCAKE performed best with fragment size (−k) set to 21 and minimum overlap (−n) set between 11 and 15. SSAKE performed similarly and generally worked well with the parameter −m (minimum overlap) set between 11 and 13, and −t (end trimming) set to 2. Velvet worked best with the hash length set between 13 and 17, whereas the ideal coverage cutoff parameter varied considerably depending on the hash length used. Among the programs tested Velvet was by far the fastest and generally more accurate than SSAKE or VCAKE (since this study, a much faster version, VCAKE 2.0 has been released). Parameters in the latter two programs could however be optimized to generate longer contigs (more than 3 kb for SPFMV) than Velvet could achieve. Searches of nucleotide and protein databases using Blast with the assembled contigs and their corresponding translated peptides successfully identified the expected viruses in each plant, and, surprisingly, also identified several contigs with similarity to badnaviruses (family *Caulimoviridae*; dsDNA reverse transcribing viruses) and mastreviruses (family *Geminiviridae*; ssDNA viruses) (Table 1; see supplemental files online for detailed Blast results of Velvet assemblies). Furthermore, contigs of SPFMV generated in dually infected plants were found to span the entire genome and could be further assembled to generate the complete genomic sequence (Fig. 2A) with an average sequencing depth of 470×. The accuracy of the assembled sequence of SPFMV-Piu was confirmed by PCR and Sanger sequencing and was deposited in GenBank (accession no. FJ155666).

Further investigation of the badna- and mastrevirus specific contigs revealed that they corresponded to at least two distinct badnaviruses and one mastrevirus and covered more than 50% of their respective genomes (Fig. 2B). Phylogenetic analysis using deduced protein sequences of the mastrevirus V1 (movement protein; data not shown) and V2 (coat protein) genes, and the partial badnavirus reverse transcriptase genes, confirmed their relatedness to the genera *Mastrevirus* and *Badnavirus* respectively (Fig. 3). Primers were designed (Fig. 2B and supplemental Table 1), and successfully amplified fragments of the expected sizes filling the gaps between the contigs found by siRNA assembly (data not shown). The obtained PCR fragments were sequenced using the Sanger method and assembled to produce the near complete genomes (partial intergenic region lacking) of both Badnaviruses which were deposited in the GenBank (FJ560943 and FJ560944). The partial Mastrevirus sequences were deposited under accession numbers FJ560945 and FJ560946. The significance of these new, apparently symptomless, viruses remains to be understood.

SPCSV specific siRNA are not readily detectable using radioactive probes in SPCSV infected plants, whereas SPFMV specific siRNAs can be detected in single infection, and are abundant in SPVD affected plants (Kreuze et al., 2008). These findings were confirmed in this experiment by analysis of similarity of the sRNAs to the sequenced viral genomes using the program MAQ (Fig. 1). The number of contigs identified for each virus corresponded well with amount sRNAs found with similarity to them (Fig. 1, Table 1). Results showed that 53–57% of all virus specific sequences corresponded to the 22 nt class of sRNAs, making up 40–50% of the 22 nt sRNAs (Fig. 1). In contrast they represented only about 10% or 35% of 21–24 nt sRNAs in non-SPVD and SPVD affected plants respectively. Because the distribution of sRNAs over the viral genomes was similar among all size classes (data not shown), sequencing only the 22 nt size sRNAs could help to improve the sensitivity of virus detection using this method. We tested this with the current data set by using only the 22 nt RNAs for contig assembly and corresponding Blast. Results indicated that all viruses could readily be detected using only 22 nt sequences (Table 1). Further simulation by assembling random subsets of the 22 nt sequences indicated that as few as 30,000 22 nt sequences were enough to assemble at least one contig recognizable as SPCSV, the virus with the fewest identified siRNAs, in single infection. Much fewer sequences were required for reliable identification of the other viruses (~15,000 for the Mastrevirus, and only a few thousand for SPFMV and the Badnaviruses). In contrast, at least 120,000 21–24 nt sequences were required for similar sensitivity. On the other hand, because of the relative reduction of sRNAs specific to viruses other than SPFMV in SPVD affected plants, about 100,000 22 nt sRNA reads were required to identify SPCSV reliably, and 50,000 for the mastrevirus.

Analysis of polarity and distribution of sRNAs over the viral genomes showed that targeting was not homogenous over the viral genomes, and the positive strand was more profusely targeted on average than the negative strand in all cases (data not shown), which is consistent with previous studies (Donaire et al., 2008 and references therein). This variation affected contig assembly when fewer sequences were analyzed in which case only contigs of highly targeted regions were obtained. The significance of the distribution over the viral genomes, and the dramatic effect seen on the abundance and size of virus specific sRNAs in plants infected by

Table 1
Number of contigs assembled by Velvet using 21–24 nt sRNA, or only 22 nt sRNA sequences, with virus specific hits as identified using Translated Nucleotide Blast (Blastx).

Plant infected with	siRNAs sequenced	Contigs identified	Contigs with Blastx hits 21–24 nt sRNA $k=15$, cov $=30$[a]	Contig sizes 21–24 nt sRNA $k=15$, cov $=30$[a]	Contigs with Blastx hits 21–24 nt sRNA $k=15$, cov $=3$[b]	Contigs with Blastx hits only 22 nt sRNA $k=15$, cov $=3$[b]
SPFMV	1,275,673	Total contigs	239		1633	431
		SPFMV	25	(≤949)	71	78
		SPCSV	0	–	0	0
		Badnavirus	38	(≤256)	62	55
		Mastrevirus	5	(≤210)	6	5
SPCSV	1,271,382	Total contigs	283		1675	285
		SPFMV	0	–	0	0
		SPCSV	10	(≤70)	64	12
		Badnavirus	44	(≤181)	63	44
		Mastrevirus	8	(≤260)	10	6
SPFMV +SPCSV	1,067,577	Total contigs	221		1363	581
		SPFMV	20	(≤1600)	43	51
		SPCSV	2	(≤81)	41	12
		Badnavirus	38	(≤266)	63	51
		Mastrevirus	5	(≤210)	8	5

[a] Parameters optimized for maximum contig sizes.
[b] Parameters optimized for maximum total contig sequence.

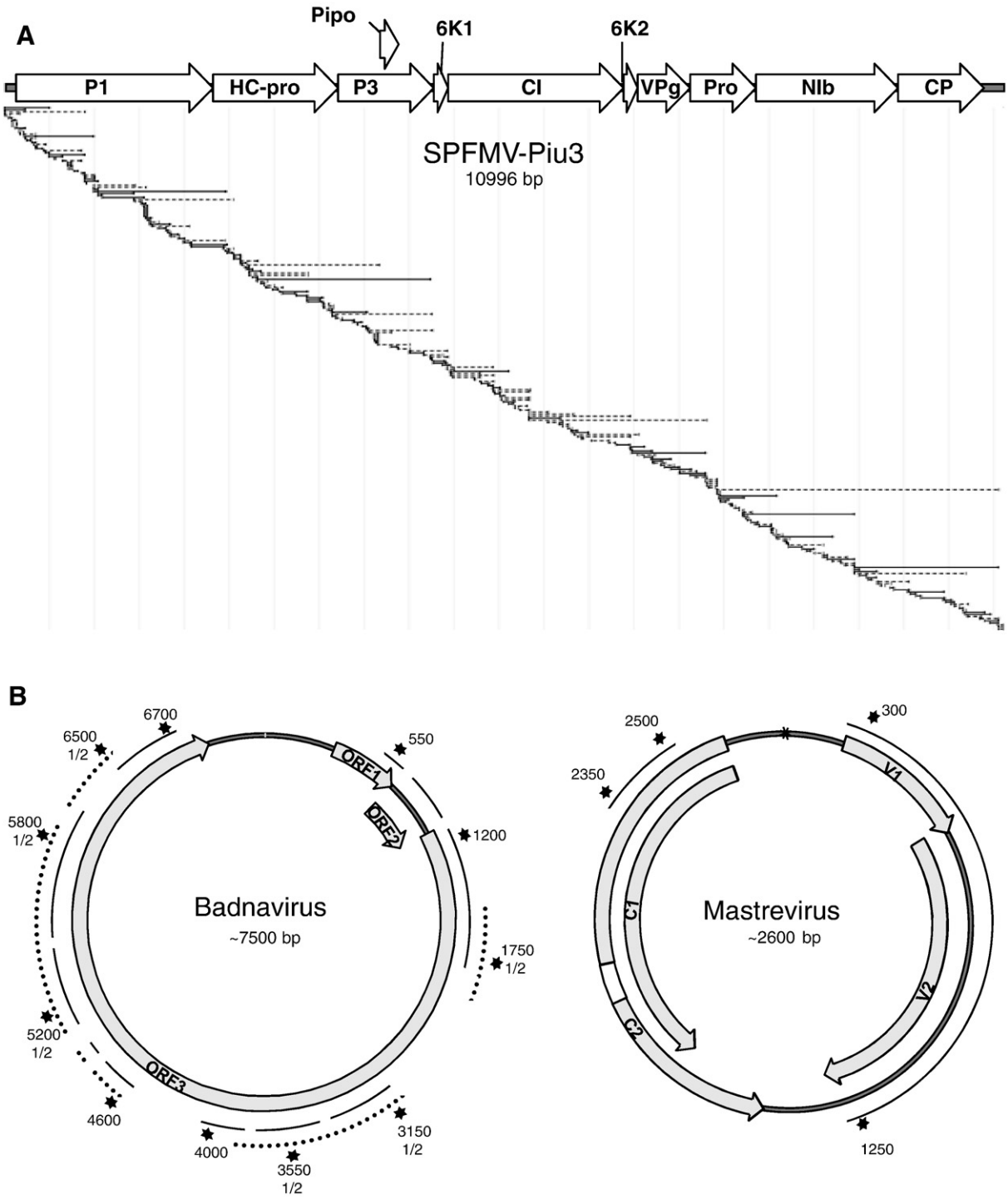


Fig. 2. Coverage of viral genomes by contigs assembled from siRNAs. (A) Schematic representation of the complete genome of SPFMV-Piu, with mature proteins indicated by open arrows and below it the coverage by contigs assembled from siRNAs using various different parameters with the short sequence assembly programs Velvet and VCAKE (see Materials and methods). (B) Schematic representation of generic badnavirus and mastrevirus genomes with open reading frames indicated by open arrows. The regions of their respective genomes covered by contigs assembled from siRNAs are indicated by lines on the outside around the genomes. The overlapping solid and dotted lines around the badnavirus genome indicate different sequences found covering the same genomic regions. Stars indicate the approximate position of primers (forward and reverse) designed based on contig sequences. Numbers next to the stars indicate primer name and 1/2 indicated two different primers were designed based on the different sequences identified.

SPFMV and SPCSV together, may provide important clues towards the mechanism of viral defense against different types of viruses as well as the synergistic disease caused by these two viruses in sweetpotato and will be further addressed in a separate study.

Besides virus specific contigs a number of apparently endogenous contigs were identified, most notably corresponding to ribosomal RNA, but also sequences with similarity to repetitive DNA and retro-elements, or genes closely associated to these elements (data not shown). The majority of contigs however did not produce any significant hits in the databases, most likely reflecting the relatively scarce amount of sweetpotato sequences available to date. The detailed analysis of endogenous small RNAs including miRNA is the topic of a separate study, and will not be further discussed here.

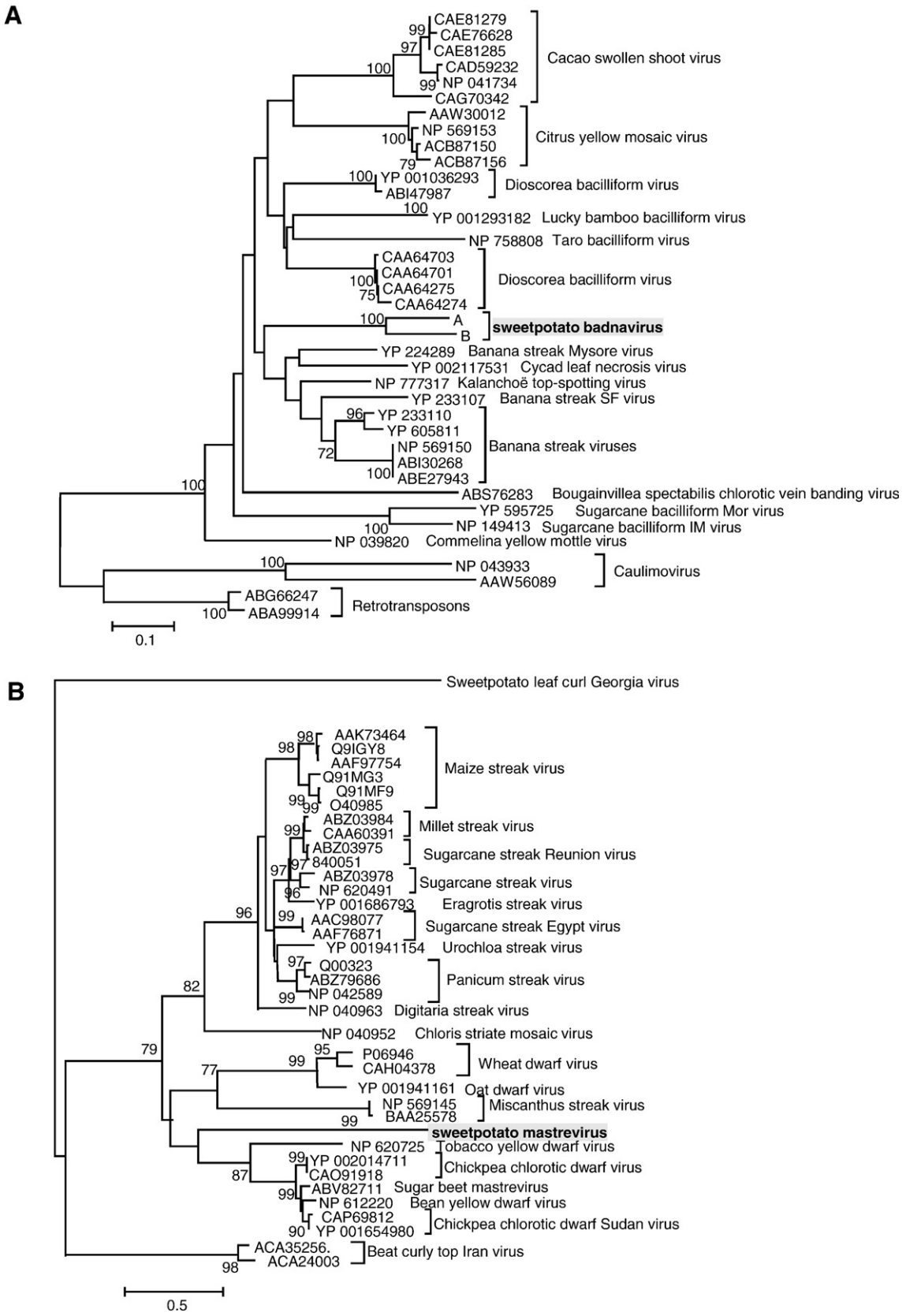


Fig. 3. Phylogenetic tree generated by neighbour joining of (A) partial reverse transcriptase protein alignments of badnaviruses and (B) mastrevirus coat protein (V2) alignments. Values at the nodes indicate percentage of bootstrap support (out of 100 bootstrap replicates) and are indicated if greater that 70. GenBank protein accession numbers are indicated at the end of each branch. In (A) two caulimoviruses and two most closely related retrotransposons are included as outgroups. In (B) the coat protein of the begomovirus Sweetpotato leafcurl Georgia virus was used as an outgroup.

This methodology, as applied in plants, was thus able to detect both RNA and DNA viruses, from widely different families and with different tissue trophisms and intracellular replication sites, even in extremely low titre and seemingly symptomless infections. It offers an entirely generic, specific and apparently sensitive approach to identify plant viruses, as compared to other techniques, which are all in some way limited to a subset of viruses that can be identified or require additional confirmatory steps for virus identification. Although at present the technique is still expensive enough to limit the extent of its application, the obtainable sequence depth from next generation platforms is rapidly increasing, with a corresponding reduction in price per sequenced base. Furthermore, simulations using subsets of data indicated that as few as 30,000 sequences may be required to obtain a reliable diagnosis of the viruses identified in this study. Thus in a single lane of an Illumina genome analyzer, providing 3–4 million reads at the time of our experiment, potentially 100–130 samples could be analyzed simultaneously. As a comparison, current germplasm indexing methods for sweetpotato takes six month to complete, requires ten separate serological tests, one PCR and grafting to indicator hosts and were still unable to identify the new viruses found in this study (the same viruses were confirmed in uninfected control which had previously been indexed). Improving the safety of internationally exchanged germplasm is an obvious application for the use of this type of deep-sequencing approach.

It remains to be verified if similar results can be obtained with viruses from other families and in other plants, but reports of virus specific siRNA from several plant species infected by many different viruses (Donaire et al., 2008; Mlotshwa et al., 2008 and references therein) suggest it may be universally applicable. Because the production of significant amounts of siRNA from viral genomes in response to infection has been reported also from invertebrate animals (Chotkowski et al., 2008) and fungi (Zhang et al., 2008) this methodology may be applicable to these organisms as well. On the other hand substantial populations of virus specific siRNAs have not yet been reported for mammalian viruses (Haasnoot et al., 2007b and references therein). The apparent sensitivity combined with increased throughput obtained by massive parallel sequencers may eventually lead to the technique becoming widely applicable.

Materials and methods

Plant material and virus strains

Pathogen-tested *in vitro* plants of the Peruvian sweetpotato landrace 'Huachano' (accession no. CIP420065) were obtained from the germplasm collection of the International Potato Center (CIP, Lima, Peru) and transferred to an insect-proof greenhouse at CIP. The isolate of SPFMV (SPFMV-Piu) and SPCSV (SPCSV m2–47) used for plant inoculation in this study were originally obtained from sweetpotato plants infected in the field in Peru (Cuellar et al., 2008). They were maintained in *I. setosa* plants in an insect-proof greenhouse at CIP, Lima, Peru. The viruses were inoculated to sweetpotato plants via grafting. Infection was confirmed by index grafting onto the susceptible indicator plants of *I. setosa* and double or triple antibody sandwich enzyme linked immunosorbent assay for SPFMV and SPCSV respectively as described in Kreuze et al. (2008).

Nucleic acid extraction and sequencing

Total RNA was isolated from 3 g of fresh leaf material using Trizol (Invitrogen, CA, USA) following the manufacturer's instructions. The amount and quality of the RNA were checked using a spectrophotometer (Nanodrop, Thermo Fisher Scientific, Waltham, MA, USA) and agarose gel electrophoresis. Lyophilized RNA was sent to Fasteris Life Sciences SA (Plan-les-Ouates, Switzerland) for processing and sequencing on the Illumina Genome Analyzer. Briefly, the processing by Illumina consisted of the following successive steps: acrylamide gel purification of RNA bands corresponding to the size range 20–30 nts, ligation of 3′ and 5′ adapters to the RNA in two separate subsequent steps each followed by acrylamide gel purification, c-DNA synthesis followed by acrylamide gel purification, and a final step of PCR amplification to generate DNA colonies template library for Illumina sequencing. DNA was extracted using the CTAB method. PCR amplification of virus specific fragments was performed using Taq DNA polymerase (Promega) according to the manufacturer's recommendations together with virus specific primers (supplemental Table 1). Sequencing of PCR amplified fragments using the Sanger method was performed by Macrogen (Seoul, Korea).

Sequence analysis

For siRNA sequence assembly three different short read assemblers were tested: SSAKE v3.2 (Warren et al., 2007), VCAKE v1.0 (Jeck et al., 2007) and Velvet v0.6.04 (Zerbino and Birney, 2008). Different, overlapping, contigs were produced depending on the program used and the parameters set, and they could be further assembled into greater contigs using the program ContigExpress included in the Vector NTI package (Invitrogen, Carlsbad, CA). Assembled contigs were used to search the GenBank/EMBL/DDBJ database using BLASTn (nucleotide blast) or BLASTx (translated nucleotide blast). Primers (supplemental Table 1) were designed for amplification and Sanger sequencing based on the identified viral contigs using Vector NTI (Invitrogen). Alignment and phylogenetic analysis were performed using the MEGA 4 package (http://www.megasoftware.net/). Coverage and distribution of virus specific contigs by siRNAs was determined using the program MAQ (http://maq.sourceforge.net) under default parameters, and results were exported to Microsoft Excel for further analysis. The total amount of identified mastrevirus specific siRNAs was multiplied by two as an approximation of the total mastrevirus siRNAs, because only half the mastrevirus sequence was available.

Acknowledgments

This research was supported by the Howard Buffet Foundation. We are grateful to Marc Ghislain and Roger Beachy for supporting this study.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.virol.2009.03.024.

References

Anderson, P.K., Cunningham, A.A., Patel, N.G., Morales, F.J., Epstein, P.R., Daszak, P., 2004. Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. Trends Ecol. Evol. 10, 535–544.

Balijja, A., Kvarnheden, A., Turchetti, T., 2008. A non-phenol–chloroform extraction of double-stranded RNA from plant and fungal tissues. J. Virol. Methods 152, 32–37.

Boonham, N., Glover, R., Tomlinson, J., Mumford, R., 2008. Exploiting generic platform technologies for the detection and identification of plant pathogens. Eur. J. Plant Pathol. 121, 355–363.

Chotkowski, H.L., Ciota, A.T., Jia, Y., Puig-Basagoiti, F., Kramer, L.D., Shi, P.Y., Glaser, R.L., 2008. West Nile virus infection of Drosophila melanogaster induces a protective RNAi response. Virology 377 , 197–206.

Covey, S.N., Al-Kaff, N.S., Langara, A., Turner, D.S., 1997. Plants combat infection by gene silencing. Nature 285, 781–782.

Cox-Foster, D.L., Conlan, S., Holmes, E.C., Palacios, G., Evans, J.D., Moran, N.A., Quan, P.-L., Briese, T., Hornig, M., Geiser, D.M., Martinson, V., vanEngelsdorp, D., Kalkstein, A.L., Drysdale, A., Hui, J., Zhai, J., Cui, L., Hutchison, S.K., Simons, J.F., Egholm, M., Pettis, J.S., Lipkin, W.I., 2007. A metagenomic survey of microbes in honey bee colony collapse disorder. Science 318 , 283–286.

Cuellar, W.J., Tairo, F., Kreuze, J.F., Valkonen, J.P.T., 2008. Analysis of gene content in sweet potato chlorotic stunt virus RNA1 reveals the presence of the p22 RNA silencing suppressor in only a few isolates: implications for viral evolution and synergism. J. Gen. Virol. 89, 573–582.

Donaire, L., Barajas, D., Martinez-Garcıa, B., Martinez-Priego, L., Pagan, I., Llave, C., 2008. Structural and genetic requirements for the biogenesis of tobacco rattle virus-derived small interfering RNAs. J. Virol. 5167–5177.

Haasnoot, J., Westerhout, E.M., Berkhout, B., 2007a. RNA interference against viruses: strike and counterstrike. Nature Biotech. 12, 1435–1443.

Haasnoot, J., de Vries, W., Geutjes, E-J., Prins, M., de Haan, P., Berkhout, B., 2007b. The Ebola virus VP35 protein is a suppressor of RNA silencing. PLoS Path. 3, 794–803.

Haible, D., Kober, S., Jeske, H., 2006. Rolling circle amplification revolutionizes diagnosis and genomics of geminiviruses. J. Virol Meth. 135, 9–16.

Jeck, W.R., Reinhardt, J.A., Baltrus, D.A., Hickenbotham, M.T., Magrini, V., Mardis, E.R., Dangl, J.L., Jones, C.D., 2007. Extending assembly of short DNA sequences to handle error. Bioinformatics 23, 2942–2944.

Karyeija, R.F., Kreuze, J.F., Gibson, R.W., Valkonen, J.P.T., 2000. Synergistic interactions of a potyvirus and a phloem-limited crinivirus in sweet potato plants. Virology 269, 26–36.

Kreuze, J.F., Samolski Klein, I., Untiveros Lazaro, M., Cuellar Chuquiyuri, W.J., Lajo Morgan, G., Cipriani Mejía, P.G., Ghislain, M., Valkonen, J.P.T., 2008. RNA silencing-mediated resistance to a crinivirus (*Closteroviridae*) in cultivated sweetpotato (*Ipomoea batatas* L.) and development of sweetpotato virus disease following co-infection with a potyvirus. Mol. Plant. Pathol. 9, 589–598.

Melcher, U., Muthukumar, V., Wiley, G.B., Min, B.E., Palmer, M.W., Verchot-Lubicz, J., Ali, A., Nelson, R.S., Roe, B.A., Thapa, V., Pierce, M.L., 2008. Evidence for novel viruses by analysis of nucleic acids in virus-like particle fractions from Ambrosia psilostachya. J. Virol. Methods 152, 49–55.

Mlotshwa, S., Pruss, G.J., Vance, V., 2008. Small RNAs in viral infection and host defense. Trends Plant Sci. 13, 375–382.

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

Quan, P., Briese, T., Palacios, G., Lipkin, W.I., 2008. Rapid sequence-based diagnosis of viral infection. Antiviral Res. 79, 1–5.

Victoria, J.G., Kapoor, A., Dupuis, K., Schnurr, D.P., Delwart, E.L., 2008. Rapid identification of known and new RNA viruses from animal tissues. PLOS Pathol. 4, e1000163.

Warren, R.L., Sutton, G.G., Jones, S.J.M., Holt, R.A., 2007. Assembling millions of short DNA sequences using SSAKE. Bioinformatics 4, 500–501.

Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18, 821–829.

Zhang, X.M., Segers, G.C., Sun, Q.H., Deng, F.Y., Nuss, D.L., 2008. Characterization of hypovirus-derived small RNAs generated in the chestnut blight fungus by an inducible DCL-2-dependent pathway. J. Virol. 82, 2613–2619.