



HAL
open science

Sequencing viral siRNAs to identify previously undescribed viruses and viroids in a panel of ornamental plant samples structured as a matrix of pools

Eric Verdin, Catherine Wipf-Scheibel, Patrick Gognalons, François Aller, Mireille Jacquemond, Mark Tepfer

► To cite this version:

Eric Verdin, Catherine Wipf-Scheibel, Patrick Gognalons, François Aller, Mireille Jacquemond, et al.. Sequencing viral siRNAs to identify previously undescribed viruses and viroids in a panel of ornamental plant samples structured as a matrix of pools. *Virus Research*, 2017, 241, pp.19-28. 10.1016/j.virusres.2017.05.019 . hal-01607064

HAL Id: hal-01607064

<https://hal.science/hal-01607064>

Submitted on 26 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

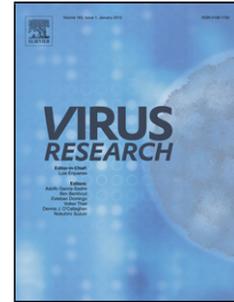


Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Accepted Manuscript

Title: Sequencing Viral siRNAs to Identify Previously Undescribed Viruses and Viroids in a Panel of Ornamental Plant Samples Structured as a Matrix of Pools

Authors: Eric Verdin, Catherine Wipf-Scheibel, Patrick Gognalons, François Aller, Mireille Jacquemond, Mark Tepfer



PII: S0168-1702(16)30788-2
DOI: <http://dx.doi.org/doi:10.1016/j.virusres.2017.05.019>
Reference: VIRUS 97147

To appear in: *Virus Research*

Received date: 30-11-2016
Revised date: 12-5-2017
Accepted date: 24-5-2017

Please cite this article as: Verdin, Eric, Wipf-Scheibel, Catherine, Gognalons, Patrick, Aller, François, Jacquemond, Mireille, Tepfer, Mark, Sequencing Viral siRNAs to Identify Previously Undescribed Viruses and Viroids in a Panel of Ornamental Plant Samples Structured as a Matrix of Pools. *Virus Research* <http://dx.doi.org/10.1016/j.virusres.2017.05.019>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Sequencing Viral siRNAs to Identify Previously Undescribed Viruses and Viroids in a Panel of Ornamental Plant Samples Structured as a Matrix of Pools

Short title: Sequencing viral siRNAs to identify new viruses and viroids in ornamental plants

Eric Verdin ^{a,*}, Catherine Wipf-Scheibel ^a, Patrick Gognalons ^a, François Aller ^b, Mireille Jacquemond ^{a,†}, Mark Tepfer^{a,c,*}

^a Pathologie Végétale, INRA, F-84140 Montfavet, France

^b Fasteris SA, Ch. du Pont-du-Centenaire 109, CH-1228 Plan-les-Ouates, Switzerland

^c Institut Jean-Pierre Bourgin (IJPB), INRA, AgroParisTech, CNRS, Saclay Plant Sciences (SPS), Université Paris-Saclay, F-78026 Versailles, France.

* Corresponding authors

E-mail: Eric.Verdin@inra.fr, Mark.Tepfer@inra.fr

Highlights

- Deep sequencing of siRNAs allowed *de novo* assembly of contigs to identify viruses
- Samples could be pooled before deep sequencing without loss of virus detection
- Detection of 27 previously unknown viruses, including all types of viral genomes
- Ornamental sector could be the source of novel plant diseases in all types of crops

Abstract

Ornamental plants constitute a largely unknown and potentially important source of pathogens affecting not only ornamental plants, but also major crop species. We have carried out studies using high-throughput sequencing of 21-24 nt RNAs from potentially virus-infected ornamental plants, followed by assembly of sequence scaffolds, to identify the virus and viroid genomes present in a panel of 67 plant samples representing 46 species belonging to the main sectors of the ornamental plant industry (cut flowers, pot plants, bulbs). A pilot study demonstrated that samples could be pooled (5 samples per pool), and the overall process simplified without loss of detection of important known pathogens. In a full-scale study, pools of 5 samples were organized in a 5x5 matrix to facilitate attribution of a sequence to a precise sample directly from analysis of the matrix. In the total of 67 samples analyzed in the two studies, partial sequences suggesting the presence of 25 previously unknown viruses and viroids were detected, including all types of virus and viroid genomes, and also showed four cases of known viruses infecting previously undescribed hosts. Furthermore, two types of potential mis-assembly were analyzed, and were shown to not affect the conclusions regarding the presence of the pathogens identified, but show that mis-assembly can affect the results when the objective is determining complete *bona fide* viral genome sequences. These results clearly confirm that ornamental plants constitute a potential source of unknown viruses and viroids that could have a major impact on agriculture, and that sequencing siRNAs of potentially virus- or viroid-infected ornamental plants is an effective means for screening for the presence of potentially important pathogens.

Keywords: Next-generation sequencing, siRNA, virus, viroid, ornamental plant

1. Introduction

Over the past decades, globalization of the ornamental plant industry has brought major changes, including ever increasing numbers of ornamental species, their rapid turnover, and a massive increase

the scope of trade worldwide, which could lead to more severe epidemiological risks due to virus and viroid diseases of ornamentals. The potential seriousness of the impact of ornamental plants harboring viruses and viroids was clearly demonstrated when a serious quarantine pathogen of potato and tomato, Potato spindle tuber viroid (PSTVd), was observed in numerous asymptomatic plants of the widely grown ornamental species *Solanum jasminoides*, first in the Netherlands (Anonymous 2006a, Anonymous 2006b, Anonymous 2006c), then in Italy (Di Serio 2007). Thus, the presence of known and unknown viruses and viroids in ornamental plants has a clear potential for being the unsuspected source of major pathogens of crop species.

Over the past decade, advances in next-generation sequencing have constituted a major breakthrough in plant virology (see Roossinck et al., 2015, and Wu et al., 2015 for review). In particular, all three of the most used strategies (sequencing double-stranded RNA, sequencing the nucleic acids of extracts enriched in virus-like particles, and sequencing virus-derived 21-24 nt siRNAs), have made it possible to detect previously unknown viruses at a remarkable scale, and appropriately these strategies are being considered for inclusion in quarantine measures (Martin et al., 2016). Although each strategy has a clear bias in the types of pathogens detected, siRNA sequencing has the advantage of detecting all types of viral genomes as well as viroids, although double-stranded RNA viruses may be under-represented (Roossinck et al., 2015). These strategies have been widely used to determine the sequence of numerous viruses of crop plants, and have provided a unique view of the diversity of viruses present in natural environments at an ecosystem scale (Roossinck et al., 2015), but to the best of our knowledge they have never been applied to a broad-scope evaluation of the presence of viruses and viroids in ornamental plants.

We have thus carried out a study testing the feasibility of using siRNA sequencing to detect viral and viroid genomes in ornamental plants. We present here a study of 67 samples coming from a total of 46 plant species, most of which expressed possible symptoms of disease. They were from a variety of sources, including producers of ornamental plants and the national wholesale ornamental plant market at Rungis, France. In addition to providing a first, very preliminary global look at the viruses and viroids present in ornamental plants, the experiments were designed to explore various means to simplify protocols and reduce costs, including pooling samples and analyzing pools in a matrix. More than 75% of the samples proved to be infected, and numerous partial sequences of previously unknown virus and viroid genomes were detected. These results suggest that the viruses and viroids of ornamental plants do represent a remarkably unknown territory that would merit further more detailed study.

Future studies based on this strategy could be applied to populations of asymptomatic plants as well as ones displaying symptoms, in order to consider the prevalence of the pathogen studied. This would

would also be important for consideration of implementation of siRNA sequencing as a complement to currently used serological and RT-PCR indexing for the presence of known viruses.

2. Materials and Methods

2.1 Origins of samples and design of matrices

Ornamental plant samples representing the three main sectors in the ornamental plant industry (cut flowers, pot plants and bulbs) were collected during 2012-2014 in France from diverse production facilities, nurseries, experimental trials and the wholesale flower market in Rungis, France. Plants came primarily from France and The Netherlands, but were also imported from Italy, Israel, China, Japan, Sudan and Columbia. Samples were selected based on the presence of potential viral symptoms including mosaic, necrosis, yellowing, discoloration, streak, mottling, and stunting. The 67 samples included 46 plant species, including both monocots and dicots (Suppl Table 1).

2.2 RNA extraction

Total RNAs were isolated using the PureLink® Plant RNA Reagent (Ambion™, Life Technologies) following the specifications of the provider. Each pool of RNA samples was established with equal amounts of total RNAs or small RNAs from 5 different plant species. To facilitate the interpretation of NGS results, the pools were formed by grouping samples from different plant species in order to minimize the risk of pooling samples harboring identical viruses.

2.3 siRNA sequencing

Total RNA samples were sent to Fasteris SA (Plan-les-Ouates, Switzerland) for small RNA library construction and deep sequencing. The library preparation was performed using the Illumina TruSeq® Small RNA Sample Prep Kit. A preliminary polyacrylamide gel size selection of 18-30 nucleotides was performed when sample quality was not sufficient. This additional step can improve the efficiency of library preparation.

An “in-lane” PhiX control spike was included in each lane of the flow-cell. Mapping these spiked control reads against the PhiX reference genome showed that the error rate was 0.21% mismatches. 94.67% of the reads had a quality value (Q30) ≥ 30 , i.e. less than 1 error in 1000 bases. The sequences corresponding to 5' and 3' adaptors were removed according to Fasteris' standard protocols. Except for HBZ-66, most of the inserts found have a length ranging from 18 to 26 bases.

For HBZ-66, 62.58% of inserts have a size ranging from 1 to 17 bases.
Sequence data are available at the NCBI Sequence Read Archive (XXX).

2.4 Bioinformatic analysis

For each sample, the 21-24 nt reads were imported into the CLC Genomics Workbench version 7 (Qiagen, Hilden, Germany), as shown schematically in Suppl Fig. 1. Each 21-24 nt set of reads was assembled *de novo* three times, using the default parameters (Bubble size = 50, Minimum contig length = 100, Mismatch cost = 2, Insertion cost = 3, Deletion cost = 3, Length fraction = 0.5, Similarity fraction = 0.8), except that word length was either 15, 17 or 19. For each sample, the three pools of assembled contigs of minimum length 100 nt were further assembled together *de novo*, but this time with word length 21 to yield two pools, the assembled scaffolds and the remainder contigs that were not incorporated into the scaffolds. BLASTn at NCBI as implemented in the CLC Workbench was used to identify sequences with the highest identity to each scaffold and remainder. The resulting tables were exported, and a case-insensitive filter used to only retain ones in which the description of the closest related sequence (top hit) included "vir", which retained all viral and viroid sequences, while discarding nearly all the others.

In most cases, the plant bearing the viral or viroid sequence could be identified by inspection of the matrix of pools. To determine which sample was the source of sequences of interest that appeared only in a horizontal or a vertical pool of a matrix, but not both (orphans), The CLC Genomics Workbench was used to map the reads of the five pools perpendicular to the one bearing the orphan, allowing determination of which perpendicular pool bore siRNAs corresponding to the sequence of interest. The same function was used to determine which sample bore the different potyviral scaffolds and remainder contigs. Inspection of the map read output made it easy to distinguish between two types of cases. Read maps in which the reads mapped to isolated singleton sites on the contig were considered to be spurious, whereas cases in which > 50 % of the contig was covered by blocks of overlapping reads were considered to be positive.

Since these analyses were carried out, pipelines specifically dedicated to assembly of siRNAs to identify viral genes have been described (Barrero et al., 2017; Zheng et al. 2017).

Simplot (Lole et al., 1999) was used to visualize possibly recombinant sequences. The putative recombinant sequence and the two corresponding parental sequences were imported into SimPlot. The following parameters were used: Window: 140 pb, Step: 20 pb, GapStrip: On, Kimura (2-parameter), T/: 2.0.

3. Results

3.1 Pilot study

Ten plants suspected of being virus infected, including nine ornamentals and one crop plant, were selected for a pilot study. The aim was to determine if samples could be pooled without excessive loss of the ability to detect the viruses present, and if the costs could be reduced by eliminating steps in the protocol for siRNA purification and sequencing. These aims were achieved by comparing the results obtained using the standard protocol for siRNA sequencing on the ten individual samples to those obtained with two pools of five samples, each of which was treated according to three simplified protocols designed to reduce costs (Fig 2). Depending on the protocol, simplifications could reduce costs, by 21-36%. The cost calculations taken into account were samples preparation kits, time for preparation and sequencing kit and time for sequencing. Therefore, the reduction of 21% to 36% is made only by reducing the number of library preparation. 2.3 to 4.9 million 21-24 nt reads were obtained for the different samples (Suppl Table 2).

As described in Suppl Fig. 1, the CLC Genomics Workbench was used for *de novo* assembly of contigs using RNAs of 21-24 nt and three different word lengths to generate scaffolds and "remainder" contigs (those that did not enter into scaffold assembly). As expected, compared to the contigs, the resulting scaffolds had higher *N50*, maximum length, and average length values (Suppl Table 2). Among the 10 samples, the maximum scaffold length ranged from 466 to 5692 nt. Assembled sequences were compared to the NCBI non-redundant nucleotide sequence database to identify the scaffolds and remainders with similarity to sequences in databases whose description includes "vir", such that all virus and viroid sequences would be retrieved. Among the viral sequences thus identified, only two did not meet the criterion for significance, $E < 10^{-5}$, and were not considered further.

As shown in Table 3, viral sequences were detected in all ten samples analyzed individually in the pilot study. This included nearly all types of viral genomes, as well as a viroid-like sequence. In some samples, scaffolds longer than 500 nt with >95% identity to a known viral genome clearly indicated the presence of a known viral species (highlighted in red). We also consider the presence of *Dianthus caryophyllus* viroid-like element in carnation HBZ-20 and of Petunia vein clearing virus in petunia HBZ-22 to be certain (highlighted in green), since these pathogens exist as sequences integrated in their host's genome, from which they are transcribed into RNA (Hernandez et al., 1992; Hegedus et al., 2001); thus even if infectious forms of the pathogen were absent, the corresponding siRNAs are expected to be observed. In the cases where scaffolds of >200 nt with less than 90% identity to known viral genomes were observed, as in squash HBZ-24 and wild sweetpea HBZ-17, this provided clear evidence for the presence of an as yet unknown virus. This has been confirmed in the case of the Torradovirus in squash HBZ-24, which has been fully characterized, and proposed to constitute a new

species (Lecoq et al., 2016). Concerning the remaining relatively short viral sequences detected, we propose that they are sufficient to show the presence of a member of the genus in question, but not to determine which species. For instance, in wild sweetpea HBZ-17, several scaffolds and remainders had similar levels of identity to four different luteoviruses, and others to three different ilarviruses, but we consider it likely that only one member of each genus was present. So these situations will be considered to allow identification only to the genus level. In anemone sample HBZ-21, scaffolds and remainders with sequence similarity to five different dsRNA viruses in several genera were observed. Considering that the identities are relatively short (<500 nt), we propose to consider these to show the presence of a dsRNA viral genome.

The analyses of samples HBZ-15 to -24 carried out individually were used as the comparators to determine if we could pool samples in groups of five, and also simplify the siRNA preparation and sequencing strategy. Fig 2 shows the usual and simplified pooling and sequencing strategies. In the reference strategy 1, each sample was treated separately; in strategy 2, total RNAs from 5 samples were pooled before isolation of siRNAs; in strategy 3, siRNAs were pooled after being purified on acrylamide gels; in strategy 4, total RNAs from 5 samples were pooled, siRNAs were not purified before adaptor ligation and cDNA synthesis. The pools were constructed as shown in Table 4; samples HBZ-15 to -19 were pooled, as were HBZ-20 to -24, and each pool analyzed with the three simplified strategies, after which viral sequences were assembled and extracted as previously (Suppl Fig 1).

The third column of Table 4 shows all the viruses that would be predicted to be present in the pools, i.e. all the viruses detected when the samples were analyzed separately (Table 3), and the fourth column of Table 4 presents the expected viruses that were not observed in the pooled analyses. In all pools, whatever sequencing strategy was used, certain viruses were not detected, which can be explained by the greater dilution of the siRNAs representing each virus in the pools relative to the analyses of individual sample. In pools HBZ-11, -13 and -25, the missing viruses were two of four viruses that were detected as short sequences (<500 nt) identified to genus level, and in pools HBZ-12, -14, and -26, the missing viruses were either an integrated virus or a dsRNA viruses. Thus, all the viruses that yielded contigs ≥ 400 nt were easily detected in the pools. Furthermore, since there were no differences between the viruses detected using the three simplified protocols, any of them could be used in the full-scale experiment presented below.

3.2 Full-scale study

The pilot study showed that pooling five samples and using a simplified protocol did not lead to unacceptable loss of ability to detect viral genomes. However, the disadvantage of pooling is that

identification of the pool member positive for a virus would require further analyses, e.g. by virus-specific RT-PCR of each sample constituting the pool. The aim of the full-scale study was to evaluate whether direct identification of the virus-positive pool members could be achieved by organizing pools of five samples in a matrix, as shown in Suppl Fig 3. For instance, if the same virus is detected in horizontal pool B and vertical pool H, it is present in sample 8, at the intersection of the pools. To test this, a total of 75 samples was organized in three 5 x 5 matrices, of which 55 ornamental plant samples, the others being related to other sequencing projects that will not be presented here. Since Matrix 3 was entirely composed of ornamental plants, it will be presented in detail. Each pool of Matrix 3 was composed of five plants of distantly related species, since closely related plants are more likely to be infected by the same virus species, which would make it difficult to determine which sample contributed a particular viral sequence. Sequencing libraries were prepared from the 30 pools corresponding to the three 5x5 matrices according to strategy 4, the simplest one (Fig. 2), but for 3 pools, this yielded no PCR-amplified product, probably due to the presence of inhibitory substances in one or more samples. So the potentially problematic samples were replaced, and strategy 2 was used for all samples, which yielded satisfactory result.

In Matrix 3, the yield of 21-24 nt reads was more variable than in the pilot study, ranging from 7.5 M in HBZ-65 to 1.8 M in HBZ-66 (Suppl Table 5), for which we have no precise explanation.

Consequently the number of viral scaffolds and remainder contigs was also highly variable, but viral sequences were assembled from the reads of all pools (Table 6). Globally, 50% of the scaffolds were viral, and the proportion of scaffolds with low similarity to known sequences ($E > 10^{-5}$) was low, 12.9%. Since the remainder contigs are shorter than the scaffolds, it was expected that the proportion of them with similarity to known sequences below $E > 10^{-5}$ was greater, and fewer of them were identified as viral.

As shown in Table 7, viral sequences were detected in all ten pools of Matrix 3, many of which had >90% identity to known viruses and were considered to be identified to species. In all ten pools there were also sequences of lower sequence identity to known pathogens, which were identified to genus. When the pools were presented in matrix format (Fig. 4), the vast majority of virus/viroid species could be attributed to specific individual samples. For instance, Impatiens necrotic spot virus (INSV) was only detected in pools HBZ-52 and -64, and was thus unambiguously localized to anemone V088. Similarly, two sequences identified to genus could be attributed to a precise sample, a Badnavirus to canna V132 and a Partitivirus to anemone V088.

Further efforts were required to determine from which samples the few remaining virus/viroid sequences originated. For instance, the Alstroemeria mosaic virus (ALMV) and Lily symptomless virus (LSV) sequences in pool HBZ-51 and the Ilarvirus in pool HBZ-66 could be considered orphans, since no scaffolds or contigs of these viruses were assembled from the five pools perpendicular to the

pools in which they were found. However, we were able to unambiguously determine which sample harbored these viruses by mapping the 21-24 nt reads of the five pools perpendicular to the one bearing the orphan sequence. For instance, when the reads of pools HBZ-52 -53, -54, -55, and -66 were mapped individually on the AIMV and LSV scaffolds of pool HBZ-51, reads covering more than 50% of the scaffolds were only observed in pool HBZ-52, thus localizing the two viruses to alstroemeria V098. Similarly, mapping the reads from perpendicular pools showed that the Ilarvirus sequence in pool HBZ-66 was in daisy V139.

The same strategy of read mapping was used to identify which samples bore the sequences identified as belonging to genus Potyvirus in three horizontal and three vertical pools. This was possible because potyviruses are highly diverse, and thus the potyviral sequences observed in the six pools were likely from different members of the genus. For instance, when the reads of pools HBZ-52 -53, -54, -55, and -66 were mapped on the potyviral sequences of pool HBZ-65, reads covering more than 50% of each scaffold were only observed in pools HBZ-52 and -66, thus localizing two potyviruses, one in anemone V089 and a different one in ranunculus V090. In the latter case, the potyvirus sequences could be from the remaining part of the Ranunculus leaf distortion virus (RanLDV) genome, of which there is only a partial sequence in the NCBI database. Similarly, the potyvirus sequences in alstroemeria V098 and ranunculus V095 may correspond to the remaining parts of Alstroemeria mosaic virus (AIMV) and Ranunculus mild mosaic virus (RanMMV), respectively. That the different viruses were correctly attributed to the right sample was confirmed by sequencing RT-PCR products amplified from RNAs of 10 individual samples; in all cases, using primer pairs specific to the scaffolds examined led to the amplification of the corresponding viral sequences, with the expected levels of sequence identity to reference viral genomes in the NCBI nucleotide database (Table 8).

It remained to determine which samples were responsible for the presence of the Tomato spotted wilt virus (TSWV) sequences in two horizontal pools (HBZ-47 and -48) and two vertical pools (HBZ-53 and -55). Since intra-specific sequence diversity among the TSWV sequences was low, it was not possible to distinguish which samples contributed the TSWV sequences by read mapping. TSWV-specific primers were used to evaluate the presence of the virus in the four potentially TSWV-positive samples (canna V132, peony V118, primrose V099, ranunculus V095), of which only the latter could be demonstrated to be TSWV positive. In principle, at least one of the three remaining samples should have been positive, but no TSWV-specific amplification was observed, although the possibility of sample contamination cannot be formally excluded (Kreuze et al., 2009).

The presence of orphan scaffolds and remainder contigs (ones appearing in only one pool in the matrix) showed that the depth of sequencing was probably insufficient to detect all viruses present. Since each sample in a matrix was present in one horizontal and one vertical pool, greater sequencing

depth could be achieved by assembling *de novo* the 21-24 nt reads of all two-by-two combinations of horizontal and vertical pools. For example the 21-24 nt reads of pools HBZ-47 and -54 were assembled together to generate deeper sequence results for peony V116. For all 25 populations of two-by-two assembled pools, viral/viroid sequences were identified as previously (Suppl Fig 1), and read mapping was used to determine which sample bore the sequences of any new pathogen not previously observed. In Matrix 3, this showed that sequences related to Mulberry viroid-like RNA were present in Peony V116, and several other similar cases were found in the two other matrices that were partially composed of ornamental samples (not shown).

Viewed globally, considering all 67 plant samples representing 46 species of the pilot and full-scale study, all ten plants of the pilot study were infected, as were 32 of the 57 samples of the full-scale study, and all the major types of viral/viroid pathogens were detected (Table 9 and Suppl Table 10). As could be expected, species of single-stranded plus-sense RNA viruses were most numerous, with more than 30 species detected. Of these, members of the genus Potyvirus were particularly numerous. A single quarantine pathogen was detected, Tobacco ringspot virus (TRSV) present in Cleome HBZ-15 and V129. Considering that ornamental species have been studied much less than crops, it was not surprising to discover numerous previously undescribed virus/viroid sequences (Table 11). And it is of note that we observed four cases of new plant hosts for known viruses (Table 12). Since there are known cases of single- and double-stranded DNA and viroid sequences integrated in host genomes, it should be noted that the results obtained regarding these classes of pathogens do not allow distinguishing between integrated sequences and infectious viruses.

3.3 Potentially recombinant scaffolds

Remarkably, one 444 nt viral scaffold assembled from pool HBZ-25 of the pilot study was clearly chimeric, with different parts having high identity to two completely unrelated viruses. The first 280 nt had 93% identity to RNA 1 of Tobacco ringspot virus (TRSV), and the remaining 166 nt had 100% identity to Tomato spotted wilt virus (TSWV) RNA segment S (Suppl Fig. 5). This was clearly due to an assembly artifact, since these two viruses were present in different samples when analyzed separately (HBZ-15 and HBZ-18, respectively, Table 3). Upon inspection, the two viral genomes share 12 nt of perfect identity (GAAGAAGGGGAA) at the site of junction in the scaffold. Examination of the contigs assembled using different word lengths showed that the misassembly only occurred when a length of 15 nt was used, and not at 17 or 19 nt. For the purpose of this study, detection of viral genomes, the presence of this type of misassembly is not of great concern, but would be if the objective were obtaining complete, authentic sequences of viral isolates.

Another intriguing case was presented by two potentially recombinant scaffolds of Ornithogalum mosaic virus (OrMV) in iris V082. The clearest potential recombinant was scaffold 11 of pool HBZ-46, which was composed of alternating blocks of sequence nearly identical to two divergent isolates of OrMV, JN127345 and FJ573184, that share 78% sequence identity (Suppl Fig. 6). This scaffold could either represent a genuine recombinant isolate of OrMV, or otherwise could be the product of misassembly of siRNAs corresponding to the two parental isolates, present in the same sample. When RT-PCR was carried out on total RNA of iris V082 using primers that spanned two recombination crossovers of scaffold 11, the sequences obtained (positions indicated by V082-24 and -25 in Suppl Fig. 6) were 99% and 98% identical to JN127345 and FJ573184 over their full lengths, respectively, and the crossovers of scaffold 11 were not observed. These results strongly suggest that the potential recombinant scaffold 11 was also produced by artefactual mis-assembly of siRNAs derived from co-infecting isolates of OrMV that are closely related to JN127345 and FJ573184. Examination of the contigs assembled using different word lengths showed that the mis-assembly occurred with all three word lengths (15, 17 or 19 nt), and thus was not due to assembly of contigs to generate the scaffold.

4. Discussion

This feasibility study showed that the strategy of pooling samples and analyzing them in 5x5 matrices can be remarkable effective, since in a total of 67 samples representing 46 plant species, sequences representing 25 potentially new virus and viroid species were detected (Table 11). However, this strategy has limitations; for instance, in Matrix 3, the presence of the same virus (TSWV) in several horizontal and vertical pools made it impossible to directly read from the matrix in which samples TSWV was present. Thus, in future studies using this approach, it will be important to avoid situations in which the same viruses are expected in numerous samples composing a matrix. In contrast, it would be particularly attractive to verify the virus-free status of a collection of samples, since the presumption would be that few individuals would be infected.

We describe two clear cases of *de novo* assembly of artefactual recombinant contigs. Within the goals of the present project, which was primarily to detect viruses, this is not a major problem. However it would be a serious problem if the aim were to determine *bona fide* complete genome sequences. If that is the primary goal, it would be appropriate to confirm sequences by classic sequencing of PCR or RT-PCR products, or otherwise use a strategy that would be based on longer reads, such as those obtained by sequencing double-stranded RNAs the contents of virus-like particles, or ribo-depleted RNAs.

In future studies, it would be highly desirable to achieve greater sequencing depth than those obtained here. This would relieve the problem of orphan scaffolds and remainder contigs, and could also

clarify the status of short contigs with relatively weak sequence similarity to virus or viroid sequences. Since the cost of sequencing can be expected to continue to decrease, this should be a realistic cost-effective goal. Further simplifications that could entail major cost reductions should be possible. Particularly if automation of sample preparation is an obvious goal, eliminating the necessity to purify small RNAs on acrylamide gels would be important, and protocols for doing so have been reported (e.g. Peng et al., 2014). It would also be pertinent to examine whether more sophisticated pooling strategies (e.g. 3-D matrices of pools or the strategy described by Erlich et al., 2009) would reduce the numbers of uncertain attributions of a virus to a precise sample, a problem encountered with TSWV in the full-scale study. In the longer term, a semi-automated system for detecting plant viruses and viroids by siRNA sequencing would be a highly useful tool for following the populations of these pathogens in the ever-changing world of ornamental plants. And furthermore, the strategy for large-scale siRNA sequencing could be of great use for discovering viruses in other types of organisms where siRNA synthesis is an important feature of virus resistance, such as nematodes and insects.

The results presented here justify larger similar studies to evaluate the diversity of viruses and viroids of ornamental plants, particularly ones that are circulating worldwide commercially. However, it is important to consider their implications within a wider framework of epidemiological risk. In this regard, studies on invasive organisms can provide guidance. It has been shown in many studies that the number of successful invasive organisms is related to the volume of commerce, whether invasive plants, insects or plant pathogens are concerned (Paini et al. 2016; Hulme 2009; Westphal et al., 2008). The ongoing increase in commerce in ornamental plants could thus be a reason for real concern in the light of the number of previously undescribed pathogens detected in this study. In terms of the steps in the invasion process (introduction, lag phase, establishment, expansion (Marisco et al., 2010)), our results suggest that the number of virus and viroid species that could be introduced via ornamental plants may be high. But since we selected plants that presented possible symptoms, these results do not represent the prevalence of these pathogens in ornamentals, which would be the equivalent of propagule pressure in terms of invasive species. Furthermore, we know very little about the likelihood of the later steps, such as establishment of viruses once they have been introduced. A critical question may be the degree of access of the main vectors of plant viruses (aphids, whiteflies, thrips, nematodes), to the imported ornamentals. This would be expected to be much lower for imported cut flowers, compared to the longer-lasting pot plants, bulbs and tubers. And these different types of ornamentals also pass through commercial circuits of different lengths: very short for cut flowers, longer for bulbs and tubers, and even longer for plants multiplied locally after import. The length and complexity of these circuits would have an effect on the likelihood of interactions with vectors. The relative rarity of invasive plant viruses being associated with importing ornamental plants may suggest that the actual epidemiological risk would be lower than what appears from the richness of initial introductions we have observed. However, it is important to consider that there is

often a lag phase of varying length between introduction and establishment of invasive organisms, in which case some of the novel pathogens we have observed could become established at some later date. Evaluating the risks associated with viruses harbored by ornamental plants represents an ambitious goal, which would require not only better understanding of their introduction, as could be done by further studies like this one, as well as an estimation of virus/viroid prevalence, but also better understanding of the parameters that would allow evaluation of the potential for establishment and the overall epidemiological risk.

Acknowledgements

We thank Fabien Robert of ASTREDHOR (Association nationale des structures d'expérimentation et de démonstration en horticulture) for allowing us to contact many of the advisors and producers of the ornamental sector who provided samples for this project, and Marie-Laure Winocq, AREXHOR (Association régionale d'expérimentation horticole) who provided particularly numerous ornamental samples. We thank Gregory Girardot, Jonathan Gaudin, and Antoine Lauvernay for assistance with sample preparation. We thank Florian Maumus for excellent suggestion on the manuscript.

The IJPB has benefitted from a French State grant (LabEx Saclay Plant Sciences-SPS, ref. ANR-10-LABX-0040-SPS), managed by the French National Research Agency under an "Investments for the Future" program (ref. ANR-11-IDEX-0003-02). This specific research project was supported by the INRA Plant Health and Environment Department Bioresources Program contract P03006.

References

Anonymous, 2006a. Isolated finding of Potato spindle tuber viroid on *Solanum jasminoides* in the Netherlands. EPPO Reporting Service 07-2006, 2006-142.

Anonymous, 2006b. Isolated finding of Potato spindle tuber viroid on *Solanum jasminoides* in Germany. EPPO Reporting Service 08-2006, 2006-159.

Anonymous, 2006c. Further finding of Potato spindle tuber viroid on ornamental Solanaceae in the Netherlands. EPPO Reporting Service 11-2006, 2006-229.

Barrero, R.A., Napier, K.R., Cunnington, J., Liefing, L. Keenan, S., Frampton, R.A., Szabo, T., Bulman, S., Hunter, A., Ward, L., Whattam, M., Bellgard, M.I. 2017. An internet-based bioinformatics toolkit for plant biosecurity diagnosis and surveillance of viruses and viroids. *BMC Bioinformatics*. 18, doi 10.1186/s12859-016-1428-4.

Di Serio, F., 2007. Identification and characterization of *Potato spindle tuber viroid* infecting *Solanum jasminoides* and *S. rantonnetii* in Italy. *J. Plant Pathol.* 89, 297-300.

Erlich, Y., Chang, K., Gordon, A., Ronen, R., Navon, O., Rooks, M., Hannon, G.J., 2009. DNA Sudoku - harnessing high-throughput sequencing for multiplexed specimen analysis. *Genome Res.* 19, 1243-1253.

Hegedus, K., Palkovics, L., Toth, E.K., Dallmann, G., Balazs, E., 2001. The DNA form of a retroviroid-like element characterized in cultivated carnation species. *J. Gen. Virol.* 82, 687-691.

Hernandez, C., Daros, J.A., Elena, S.F., Moya, A., Flores, R., 1992. The strands of both polarities of a small circular RNA from carnation self-cleave in vitro through alternative double- and single-hammerhead structures. *Nucleic Acids Res.* 20, 6323-6329.

Hulme, P.E., 2009. Trade, transport and trouble: managing invasive species pathways in an era of globalization. *J. Appl. Ecol.* 46, 10-18.

Kreuze, J.F., Pérez, A., Untiveros, M., Quispe, D., Fuentes, S., Barker, I., Simon, R. 2009. 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. *Virology* 388: 1-7 DOI 10.1016/j.virol.2009.03.024

Lecoq, H., Verdin, E., Tepfer, M., Wipf-Scheibel, C., Millot, P., Dafalla, G., Desbiez, C., 2016. Characterization and occurrence of squash chlorotic leaf spot virus, a tentative new torradovirus infecting cucurbits in Sudan. *Arch. Virol.* 161, 1651-1655.

Li, R., Gao, S., Hernandez, A.G., Wechter, W.P., Fei, Z, Ling, K.S. 2012. Deep Sequencing of Small RNAs in Tomato for Virus and Viroid Identification and Strain Differentiation. *PLoS ONE* 7: e37127.

Lole, K.S., Bollinger, R.C., Paranjape, R.S., Gadkari, D., Kulkarni, S.S., Novak, N.G., Ingersoll, R., Sheppard, H.W., Ray S.C., 1999. Full-Length Human Immunodeficiency Virus Type 1 Genomes from Subtype C-Infected Seroconverters in India, with Evidence of Intersubtype Recombination. *J. Virol.* 73, 152-160.

Marisco, T.D., Burt, J.W., Espeland, E.K., Gilchrist, G.W., Jamieson, M.A., Lindstrom, L., Roderick, G.K., Swope, S., Szucs, M., Tsutsui, N.D., 2010. Underutilized resources for studying the evolution of invasive species during their introduction, establishment and lag phases. *Evol. Appl.* 3, 203-219.

Martin, R. R., Constable, F., Tzanetakis, I.E. Quarantine Regulations and the Impact of Modern Detection Methods. *Ann. rev. phytopathol.* 54, 189-205.

Paini, D.R., Sheppard, A.W., Cook, D.C., De Barro, P.J., Worner, S.P., Thomas, M.B., 2016. Global threat to agriculture from invasive species. *Proc. Natl. Acad. Sci. USA.* 1134, 7575-7579.

Peng, J., Xia, Z., Chen, L., Shi, M., Pu, J., Guo, J., Fan Z., 2014. Rapid and Efficient Isolation of High-Quality Small RNAs from Recalcitrant Plant Species Rich in Polyphenols and Polysaccharides. PLoS ONE e95687.

Roossinck, M. J., Martin, D. P., Roumagnac, P., 2015. Plant virus metagenomics: Advances in virus discovery. *Phytopathol.* 105, 716-727.

Westphal, M.I., Browne, M., MacKinnon, K., Noble, I., 2008. The link between international trade and the global distribution of invasive alien species. *Biol. Invasions* 10, 391-398.

Wu, Q., Ding, S.-W., Zhang, Y., Zhu, S. Identification of viruses and viroids by next-generation sequencing and homology-dependent and homology-independent algorithms. *Ann. rev. phytopathol.* 53, 425-444.

Zheng, Y., Gao, S., Padmanabhan, C., Li, R., Galvez, M., Gutierrez, D., Fuentes, S., Ling, K.-S., Fei, Z. 2017. VirusDetect: An automated pipeline for efficient virus discovery using deep sequencing of small RNAs. *Virology* 500,130-138.

Figure legends

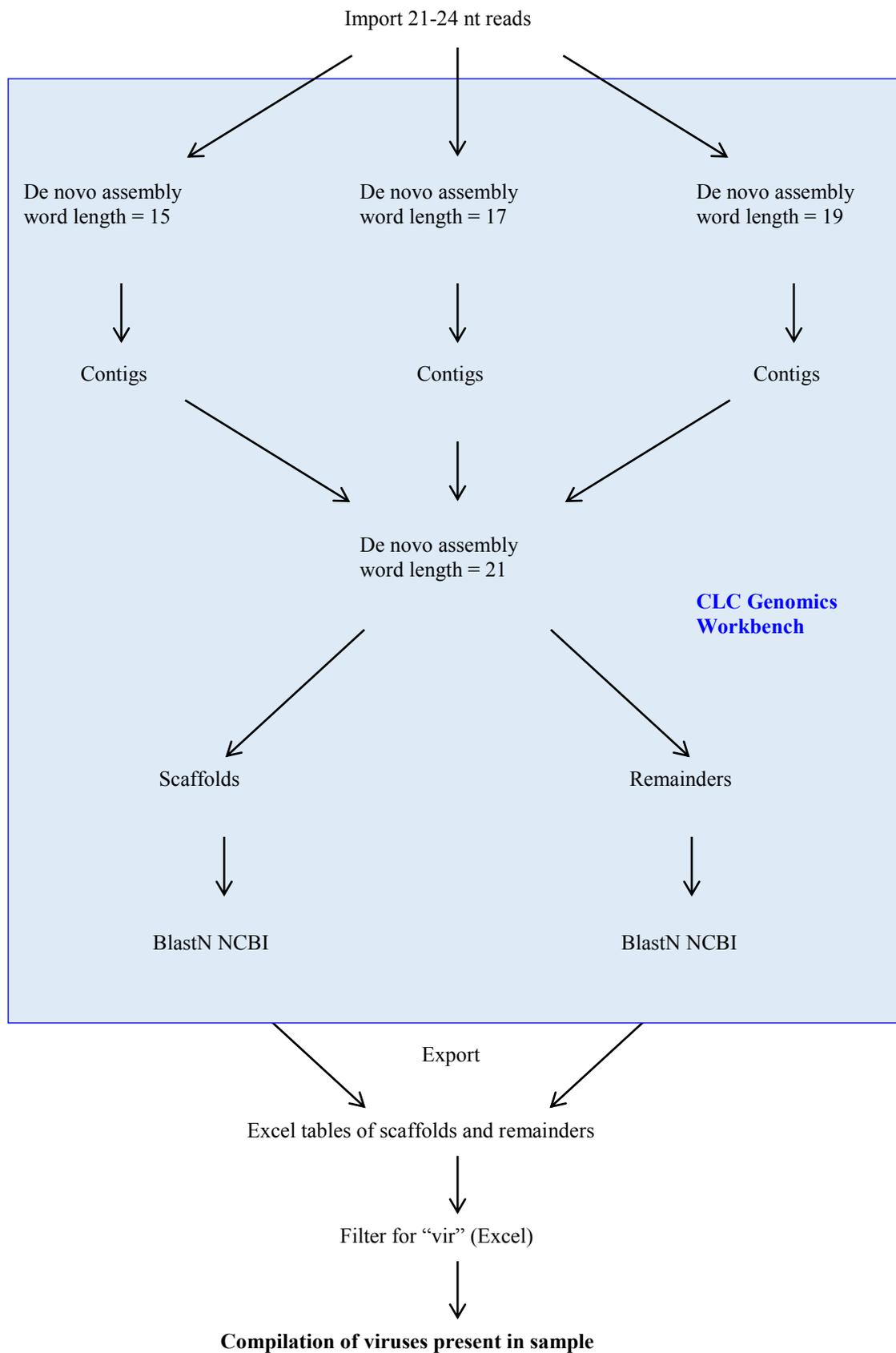


Fig. 1 (suppl) Bioinformatics strategy

Fig. 2. Pooling and sequencing strategies

Usual protocol for sequencing small RNAs:

1. Purification of total RNA samples
2. Isolation of small RNAs on acrylamide gel
3. Ligation of 3' adaptors
4. Ligation of 5' adaptors
5. cDNA synthesis
6. PCR
7. Purification of PCR products on acrylamide gel
8. Sequencing

Strategy	Percent cost reduction
Strategy 1 Usual protocol, each sample treated individually	
Strategy 2 Total RNA samples pooled before isolation of small RNAs (after Step 1)	30%
Strategy 3 Small RNAs pooled after isolation on polyacrylamide gel (after step 2)	21%
Strategy 4 Ligation of 3' adaptors on pooled total RNA samples (after step 1, step 2 omitted)	36%

	F	G	H	I	J
A	1	2	3	4	5
B	6	7	8	9	10
C	11	12	13	14	15
D	16	17	18	19	20
E	21	22	23	24	25

Fig. 3 (suppl) Full-scale study: Example of a matrix of 10 pools (A to J) of 25 samples. If the green virus is present in pools B and H, it is in fact from sample 8. Cost-effectiveness: 10 pools sequenced, rather than 25 samples

		Impatiens necrotic spot virus Ranunculus leaf distortion virus Partitivirus Potyvirus	Petunia vein clearing virus Ranunculus mild mosaic virus Tomato spotted wilt virus Potyvirus	Tobacco rattle virus Tobacco ringspot virus Cycas necrotic stunt virus	Canna yellow streak virus Cycas necrotic stunt virus Tomato spotted wilt virus Bean yellow mosaic virus Potyvirus Badnavirus	Iarvirus
		HBZ-52	HBZ-53	HBZ-54	HBZ-55	HBZ-66
Bean yellow mosaic virus Canna yellow streak virus Ranunculus mild mosaic virus Tobacco rattle virus Tomato spotted wilt virus Potyvirus Badnavirus	HBZ-47	V085 (guzmania)	V095 (ranunculus) Tomato spotted wilt virus Ranunculus mild mosaic virus Potyvirus (may be rest of RMMV)	V116 (peony) Tobacco rattle virus Mulberry viroid-like RNA (V116 scaffold18)	V132 (canna) Tomato spotted wilt virus? Canna yellow streak virus Bean yellow mosaic virus Badnavirus Potyvirus	V139 (daisy) Iarvirus
Tomato spotted wilt virus Cycas necrotic stunt virus	HBZ-48	V086 (dracaena)	V099 (primrose) Tomato spotted wilt virus?	V120 (snapdragon) Cycas necrotic stunt virus	V118 (peony) Tomato spotted wilt virus? Cycas necrotic stunt virus	V141 (holly)
Impatiens necrotic spot Partitivirus	HBZ-64	V088 (anemone) Impatiens necrotic spot virus Partitivirus	V100 (geranium)	V122 (lisianthus)	V133 (chrysanthemum)	V001 (petunia)
Ranunculus leaf distortion virus Petunia vein clearing Potyvirus	HBZ-65	V090 (ranunculus) Ranunculus leaf distortion virus Potyvirus (may be rest of RanLDV)	V102 (petunia) Petunia vein clearing virus	V124 (hibiscus)	V134 (tea)	V089 (anemone) Potyvirus
Alstroemeria mosaic virus Tobacco ringspot virus Tobacco rattle virus Lily symptomless virus Potyvirus	HBZ-51	V098 (alstroemeria) Alstroemeria mosaic virus Lily symptomless virus Potyvirus (may be rest of AIMV)	V106 (dicentra)	V129 (cleome) Tobacco ringspot virus Tobacco rattle virus	V137 (pansy)	V145 (wild tulip)

Fig. 4 Full-scale study: Viruses present in plants in Matrix 3.

Pool names are indicated on a mid-grey background. The identity of viral sequences observed in each pool is shown above or to the left of pool names on a light grey background.

In black: viral sequences with >100 nt having >90% identity to a known virus species, localized to a sample by crossing pools.

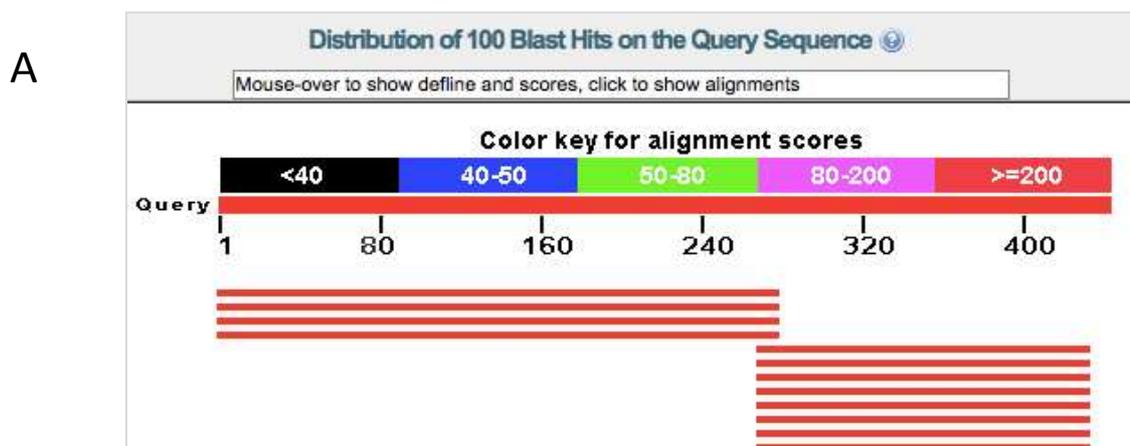
In brown: viral sequences with >100 nt having >90% identity to a known virus species, localized to a sample by mapping reads to scaffolds and remainder contigs.

In red: viral sequences with <90% identity to known virus, identified to genus, localized to a sample by crossing pools.

In blue: viral sequences with <90% identity to known virus, identified to genus, localized to sample by mapping reads to scaffolds and remainder contigs.

In green: viral or viroid sequence observed in *de novo* assembly of sums of reads in pairs of crossed pools, localized to sample by mapping reads.

?: Virus may be present, but not confirmed by RT-PCR.



B

Tobacco ringspot virus partial RdRp gene for polyprotein, genomic RNA
 Sequence ID: [emb|AJ698718.1](https://www.ncbi.nlm.nih.gov/nuclot/emb|AJ698718.1) Length: 761 Number of Matches: 1
 Related Information
 Range 1: 242 to 521 [GenBankGraphics](https://www.ncbi.nlm.nih.gov/nuclot/emb|AJ698718.1) Next Match Previous Match

Alignment statistics for match #1

Score	Expect	Identities	Gaps	Strand
410 bits(454)	9e-111	259/280(93%)	0/280(0%)	Plus/Plus
Query 1	TGGAATTGCCCAAAAATATGCTTTTTAGGAAGAAATTTTGCATTGGGCTTATTCTCT	60		
Sbjct 242	TGGAATTGCCCAAAAACATGCTTTTCAGAAAGAAATTTTGCATTGGGCTTATTCTCTT	301		
Query 61	CAGACATGCGATGGTGTCTGCCCTGCCAGGTAGGTATTGTTGTTTCAGGGCGTGAATGGG	120		
Sbjct 302	CAGACATGCGATGGTGTCTGCCCTGCCAGGTAGGTATTGTTGTTTCAGGGCGTGAATGGG	361		
Query 121	GGCTCCTGATGGACCGTCTTGTCTGCGAAGAAGTTCGGTTGCATATAACTGTGATTATCAA	180		
Sbjct 362	GGCTCCTGATGGATCGTCTTGTCTGCGAAGAAGTTCGGTTGCATATAACTGTGATTATCAA	421		
Query 181	AGTTTGATGGCCTCATGTCTTGTCAGGTGTTGGATGCTATAGGAAAATGGTAAATAAAT	240		
Sbjct 422	AGTTTGATGGTCTCATGTCTTGTCAGGTGTTGGATGCTATAGGAAAATGGTAAATAAAT	481		
Query 241	GCTATTCCAACGCCAACCCTAATTTAAAGAAAGGGGAA	280		
Sbjct 482	GCTATTCCAACGCCAACCCTAATTTAAAGAAAGGGGAA	521		

Tomato spotted wilt virus isolate p105-RB-MaxII segment S, complete sequence
 Sequence ID: [gb|HQ839731.1](https://www.ncbi.nlm.nih.gov/nuclot/gb|HQ839731.1) Length: 2927 Number of Matches: 1
 Related Information
 Range 1: 1985 to 2150 [GenBankGraphics](https://www.ncbi.nlm.nih.gov/nuclot/gb|HQ839731.1) Next Match Previous Match

Alignment statistics for match #1

Score	Expect	Identities	Gaps	Strand
300 bits(332)	1e-77	166/166(100%)	0/166(0%)	Plus/Minus
Query 269	GAAGAAGGGGAAAGAGTATGCTGTCTATACTTAGCTCCAGCAATCCTAATGCTAAAGGAAG	328		
Sbjct 2150	GAAGAAGGGGAAAGAGTATGCTGTCTATACTTAGCTCCAGCAATCCTAATGCTAAAGGAAG	2091		
Query 329	TATTGCTATGGAACATTACAGTGAAACTCTTAACAAGTTCTATGAAATGTTGGGGTTAA	388		
Sbjct 2090	TATTGCTATGGAACATTACAGTGAAACTCTTAACAAGTTCTATGAAATGTTGGGGTTAA	2031		
Query 389	AAAACAGGCAAACTCGCAGAAGTGTCTTAAAAGCAGCTGTAAGTT	434		
Sbjct 2030	AAAACAGGCAAACTCGCAGAAGTGTCTTAAAAGCAGCTGTAAGTT	1985		

Fig. 5 (suppl) Pilot project: HBZ-25 scaffold 61 is a recombinant due to *de novo* mis-assembly, shown by BlastN analysis.

A. HBZ-25 scaffold 61 (444 nt) is a chimera composed of TRSV (first 280 nt) and TSWV (the rest). **B.** When the sequences are aligned, there is an overlap of 12 nt shared identity that exists in the reference sequences (highlighted in yellow).

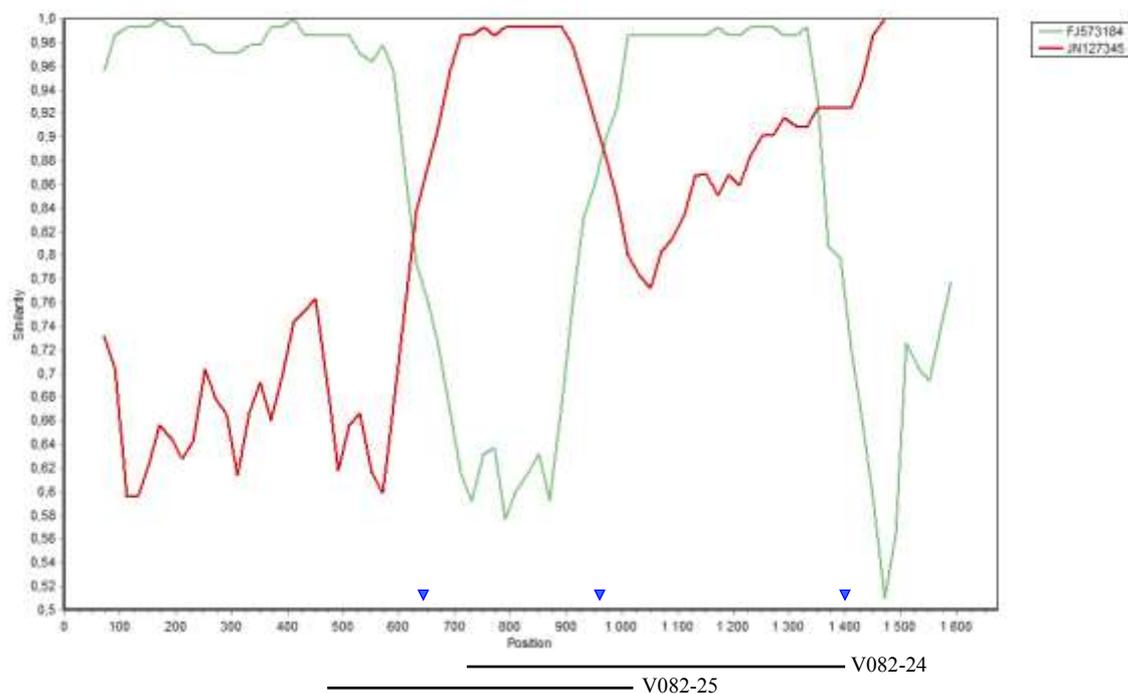


Fig. 6 Full-scale study: SimPlot analysis of HBZ-46 scaffold 11 (nt 469-2135) of Iris V082.

In BlastN analyses, scaffold 11 appeared to be composed of alternating blocks with high similarity to two divergent isolates of OrMV (accessions FJ573184 and JN127345). This was confirmed in the SimPlot (Lole et al., 1999) analysis presented here, in which red and green lines show similarity to FJ573184 and JN127345, respectively. The position of putative crossovers determined from sequence alignments are shown by blue triangles at the bottom of the graph. The position of sequenced RT-PCR amplification products, V082-24 and -25, are shown below the graph.

Table 1 (suppl) Identity and source of plant samples

Sample number	Common name	Latin name	Cultivar	Source	Origin
V001	Petunia	<i>Petunia</i> x 'Surfinia'	Sanguna Rose Vein	Normandy, France	unspecified
V003	Cleome	<i>Cleome spinosa</i>	Senorita Rosalita	Poitou-Charente, France	unspecified
V004	Verbena	<i>Verbena hybrida</i>	Donalena Red Pepper Improved	Normandy, France	Israel
V005	Primrose	<i>Primula acaulis</i>	Starflam Purple	Normandy, France	The Netherlands
V009	Geranium	<i>Pelargonium zonale</i>	Joigny	Ile de France, France	unspecified
V012	Verbena	<i>Verbena hybrida</i>	unspecified	Normandy, France	unspecified
V020	Phlox	<i>Phlox</i> sp.	unspecified	Ile de France, France	unspecified
V021	Cleome	<i>Cleome spinosa</i>	Senorita Rosalita	Aquitaine, France	unspecified
V022	Carnation	<i>Dianthus chinensis</i>	unspecified	Aquitaine, France	France
V026	Geranium	<i>Pelargonium zonale</i>	unspecified	Normandy, France	France
V027	Geranium	<i>Pelargonium zonale</i>	Tango Bravo Light Pink	Normandy, France	unspecified
V029	New Guinea Impatiens	<i>Impatiens hawkeri</i>	unspecified	Normandy, France	France
V031	Verbena	<i>Verbena hybrida</i>	Raspberry	Normandy, France	France
V032	Hydrangea	<i>Hydrangea macrophylla</i>	Teller	Normandy, France	France
V033	Kalanchoe	<i>Kalanchoe</i> sp.	Sephiros	Normandy, France	France
V035	Lilac	<i>Syringa vulgaris</i>	unspecified	Provence, France	France
V036	Wild sweetpea	<i>Lathyrus latifolius</i>	unspecified	Provence, France	France
V037	Periwinkle	<i>Catharanthus roseus</i>	unspecified	Provence, France	France
V039	New Guinea Impatiens	<i>Impatiens hawkeri</i>	unspecified	Aquitaine, France	unspecified
V046	Dahlia	<i>Dahlia pinnata</i>	unspecified	Rhône-Alpes, France	France
V047	Centranthus	<i>Centranthus ruber</i>	unspecified	Languedoc-Roussillon, France	France
V048	Oleander	<i>Nerium oleander</i>	unspecified	Languedoc-Roussillon, France	France
V051	Chrysanthemum	<i>Chrysanthemum hortorum</i>	Lovisette	Normandy, France	France
V054	Cineraria	<i>Cineraria</i> sp.	unspecified	Normandy, France	France
V055	Velvet leaf	<i>Abutilon</i> sp.	Canary Bird	Aquitaine, France	unspecified
V057	Banana	<i>Musa</i> sp.	unspecified	Ile de La Réunion, France	France (La Réunion)
V062	Pomelo	<i>Citrus</i> sp.	unspecified	Ile de La Réunion, France	France (La Réunion)
V063	Cyclamen	<i>Cyclamen</i> sp.	unspecified	Normandy, France	France
V068	Alstroemeria	<i>Alstroemeria</i> sp.	unspecified	Rungis, France ¹	The Netherlands
V069	Hydrangea	<i>Hydrangea</i> sp.	unspecified	Rungis, France	The Netherlands
V072	Lily	<i>Lilium</i> sp.	unspecified	Rungis, France	The Netherlands
V074	Asclepias	<i>Asclepias</i> sp.	unspecified	Rungis, France	Kenya
V077	Carnation	<i>Dianthus caryophyllus</i>	unspecified	Rungis, France	Colombia
V082	Iris	<i>Iris</i> sp.	unspecified	Rungis, France	France
V084	Green bell	<i>Thlaspi</i> sp.	unspecified	Rungis, France	Israel
V085	Guzmania	<i>Guzmania</i> sp.	unspecified	Rungis, France	unspecified
V086	Dracaena	<i>Dracaena</i> sp.	unspecified	Rungis, France	unspecified
V087	Snapdragon	<i>Antirrhinum majus</i>	unspecified	Provence, France	France
V088	Anemone	<i>Anemone</i>	Carmel	Provence, France	Israel

<i>coronaria</i>					
V089	Anemone	<i>Anemone coronaria</i>	Galilee	Provence, France	Israel
V090	Buttercup	<i>Ranunculus asiaticus</i>	Elegance Rose	Provence, France	Italy
V095	Buttercup	<i>Ranunculus asiaticus</i>	Pauline Brique	Provence, France	France
V098	Alstroemeria	<i>Alstroemeria</i> sp.	Alexandra	Pays de la Loire, France	unspecified
V099	Primrose	<i>Primula acaulis</i>	unspecified	Pays de la Loire, France	unspecified
V100	Geranium	<i>Pelargonium zonale</i>	Vulkan 2009	Normandy, France	unspecified
V101	Anemone	<i>Anemone coronaria</i>	Jerusalem Rose Fuschia	Provence, France	Israel
V102	Petunia	<i>Petunia</i> x <i>Surfinia</i>	Purpe Stuti Brillant	Normandy, France	unspecified
V104	Petunia	<i>Petunia hybrida</i>	Fanfare Red	Pays de la Loire, France	The Netherlands
V106	Dicentra	<i>Dicentra spectabilis</i>	unspecified	Pays de la Loire, France	unspecified
V114	Lisianthus	<i>Eustoma grandiflorum</i>	Mariachi	Provence, France	The Netherlands
V116	Peony	<i>Paeonia lactiflora</i>	Red Charm	Provence, France	The Netherlands
V118	Peony	<i>Paeonia lactiflora</i>	Coral Sunset	Provence, France	France
V119	Lisianthus	<i>Eustoma grandiflorum</i>	Bonealis	Provence, France	The Netherlands
V120	Snapdragon	<i>Antirrhinum majus</i>	unspecified	Provence, France	France
V122	Lisianthus	<i>Eustoma grandiflorum</i>	Echo	Provence, France	The Netherlands
V124	Hibiscus	<i>Hibiscus</i> sp.	unspecified	Rhône-Alpes, France	unspecified
V129	Cleome	<i>Cleome spinosa</i>	Senorita rosanita	Poitou-Charente, France	unspecified
V132	Canna	<i>Canna indica</i>	unspecified	Champagne-Ardenne, France	unspecified
V133	Chrysanthemum	<i>Chrysanthemum hortorum</i>	unspecified	Pays de la Loire, France	unspecified
V134	Tea	<i>Camelia sinensis</i>	unspecified	Ile de France, France	unspecified
V137	Pansy	<i>Viola</i> sp.	Pensee Matrix Cassis	Normandy, France	unspecified
V139	Daisy	<i>Bellis perennis</i>	unspecified	Normandy, France	unspecified
V141	Holly	<i>Ilex crenata</i>	Kinme	Rhône-Alpes, France	Japan
V142	Squash	<i>Cucurbita moschata</i>	unspecified	Sudan	Sudan
V145	Wild tulip	<i>Tulipa raddii</i>	unspecified	Rhône-Alpes, France	France
V146	Violet	<i>Viola odorata</i>	unspecified	Provence, France	France

¹ The National Wholesale Market outside of Paris

Table 2 (suppl) Pilot study: quality of *de novo* assembly of contigs and scaffolds

Sample	21-24 nt reads (10 ⁶)	Contigs ¹				Scaffolds ¹			
		<i>N50</i>	Maximum length	Average length	Total length	<i>N50</i>	Maximum length	Average length	Total length
HBZ-11	2.02	178	960	180	33 458	262	1029	245	22 080
HBZ-12	3.88	228	1038	209	32 370	532	1670	351	23 512
HBZ-13	3.48	173	2 025	184	49 361	257	2 241	259	33 687
HBZ-14	3.73	242	1 372	212	31 127	549	1 372	369	23 246
HBZ-15	4.90	231	3 243	213	36 868	512	5 692	371	33 003
HBZ-16	2.65	216	934	207	13 043	336	1 965	321	9 618
HBZ-17	3.16	191	834	186	47 396	309	1 857	270	31 345
HBZ-18	3.37	190	507	187	17 931	225	583	223	14 701
HBZ-19	3.68	144	952	157	51 888	261	1 515	250	28 710
HBZ-20	4.47	148	1 112	160	27 116	245	1 981	251	14 282
HBZ-21	2.34	143	278	145	6 522	201	506	192	6 928
HBZ-22	2.56	149	466	153	19 183	207	466	203	9 115
HBZ-23	3.20	247	1 343	205	23 587	308	2 149	278	18 095
HBZ-24	4.41	815	1 724	244	27 288	3 415	5 690	625	21 245
HBZ-25	3.21	185	1 676	188	61 306	315	2 861	268	42 669
HBZ-26	2.23	235	807	200	20 013	391	1 291	319	20 760

All samples were analyzed with three word lengths: 15, 17, or 19. The values shown in the Contigs columns were those obtained with the word length yielding the highest *N50* value. The scaffolds were obtained by assembling *de novo* the contigs obtained with the three word lengths, using word length 21.

¹ All lengths in nucleotides

Table 3 Pilot study: Analysis of viral scaffolds and remainder contigs of individual plant samples

Sample	Plant species	Closest related virus or viroid sequence (genus or family)	% identity, longest hit length
HBZ-15	Cleome	Tobacco ringspot virus (Nepovirus)	96%, 2524 nt
		Tulip virus X (Potexvirus)	100%, 5679 nt
HBZ-16	Verbena	Angelonia flower break virus (Carmovirus)	98%, 1948 nt
HBZ-17	Wild sweetpea	Pea enation virus (Enamovirus)	93%, 337 nt
		Prune dwarf virus (Iilarvirus)	89%, 411 nt
		Tobacco streak virus (Iilarvirus)	88%, 81 nt
		Viola white distortion-associated virus (Iilarvirus)	96%, 126 nt
		Fragaria chiloensis latent virus (Iilarvirus)	100%, 112 nt
		Bean leafroll virus (Luteovirus)	79%, 163 nt
		Citrus vein enation virus (Luteovirus)	93%, 167 nt
		Rose spring dwarf virus (Luteovirus)	77%, 368 nt
Soybean dwarf virus (Luteovirus)	72%, 437 nt		
HBZ-18	Cineraria	Tomato spotted wilt virus (Tospovirus)	100%, 583 nt
HBZ-19	Asclepias	Citrus endogenous virus (Pararetrovirus)s	96%, 102 nt
HBZ-20	Carnation	Dianthus caryophyllus viroid-like DNA	99%, 131 nt
		Fig cryptic virus virus (Deltapartitivirus)	79%, 176 nt
HBZ-21	Anemone	Blueberry latent virus (Amalgavirus)	93%, 115 nt
		Raphanus sativus cryptic virus 3 (Deltapartitivirus)	93%, 107 nt
		Oryza sativa virus (Endornavirus)	96%, 87 nt
		Yerba mate virus (Endornavirus)	96%, 70 nt
		Vicia faba virus (Partitiviridae)	93%, 146 nt
HBZ-22	Petunia	Petunia vein clearing virus (Badnavirus)	100%, 172 nt
HBZ-23	Lisianthus	Eggplant mottled dwarf virus (Rhabdovirus)	100%, 2149 nt
HBZ-24	Squash	Carrot torrado virus (Torradovirus)	73%, 831 nt

Highlighted viruses considered to be identified with certainty to species level: in red if scaffolds >500 pb and >95% identity, in green if integrated in the host genome.

Table 4 Pilot study: Analysis of viral scaffolds and remainder contigs of pooled samples

	Pool composition	RNA extraction and sequencing strategy	Viruses that should be present (sum of viruses in individual samples)	Viruses not detected
HBZ-11	HBZ-15 HBZ-16 HBZ-17 HBZ-18 HBZ-19	2	Angelonia flower break virus (Carmovirus) Tobacco ringspot virus (Nepovirus) Tomato spotted wilt virus (Tospovirus) Tulip virus X (Potexvirus) Enamovirus Luteovirus Ilarvirus Pararetrovirus	Enamovirus Pararetrovirus
HBZ-13	HBZ-15 HBZ-16 HBZ-17 HBZ-18 HBZ-19	3	Angelonia flower break virus (Carmovirus) Tobacco ringspot virus (Nepovirus) Tomato spotted wilt virus (Tospovirus) Tulip virus X (Potexvirus) Enamovirus Luteovirus Ilarvirus Pararetrovirus	Enamovirus Pararetrovirus
HBZ-25	HBZ-15 HBZ-16 HBZ-17 HBZ-18 HBZ-19	4	Angelonia flower break virus (Carmovirus) Tobacco ringspot virus (Nepovirus) Tomato spotted wilt virus (Tospovirus) Tulip virus X (Potexvirus) Enamovirus Luteovirus Ilarvirus Pararetrovirus	Enamovirus Pararetrovirus
HBZ-12	HBZ-20 HBZ-21 HBZ-22 HBZ-23 HBZ-24	2	Eggplant mottled dwarf virus (Rhabdovirus) <i>Dianthus caryophyllus</i> viroid-like DNA Petunia vein clearing virus (Badnavirus) Torradovirus dsRNA virus	Petunia vein clearing virus (Badnavirus) dsRNA virus
HBZ-14	HBZ-20 HBZ-21 HBZ-22 HBZ-23 HBZ-24	3	Eggplant mottled dwarf virus (Rhabdovirus) <i>Dianthus caryophyllus</i> viroid-like DNA Petunia vein clearing virus (Badnavirus) Torradovirus dsRNA virus	Petunia vein clearing virus (Badnavirus) dsRNA virus
HBZ-26	HBZ-20 HBZ-21 HBZ-22 HBZ-23 HBZ-24	4	Eggplant mottled dwarf virus (Rhabdovirus) <i>Dianthus caryophyllus</i> viroid-like DNA Petunia vein clearing virus (Badnavirus) Torradovirus dsRNA virus	Petunia vein clearing virus (Badnavirus) dsRNA virus

Highlighted viruses considered to be identified with certainty to species level: in red if scaffolds >500 pb and >95% identity, in green if integrated in the host genome.

Table 5 (suppl) Full-scale study: Quality of *de novo* assembly of contigs and scaffolds in Matrix 3

Sample	21-24 nt reads (10^6)	Contigs ¹				Scaffolds ¹			
		N50	Maximum length	Average length	Total length	N50	Maximum length	Average length	Total length
HBZ-47	4.9	416	3085	281	76 064	733	5036	462	72 119
HBZ-48	3.8	504	2863	300	36 220	1124	2863	482	31 309
HBZ-51	2.5	183	913	190	31 680	340	1152	277	22 425
HBZ-52	3.9	525	2047	322	38 981	900	4118	531	34 540
HBZ-53	4.0	481	3474	294	51 271	1277	3887	535	39 605
HBZ-54	3.3	231	1100	214	31 289	363	1100	305	25 310
HBZ-55	5.5	470	3944	302	58 948	879	4465	506	52 573
HBZ-64	6.2	792	4581	320	41 426	1439	6063	530	30 227
HBZ-65	7.5	296	3426	265	31 929	690	4259	471	24 498
HBZ-66	1.8	193	1714	194	13 585	304	2087	292	10 800

¹ All lengths in nucleotides

Table 6 Full-scale study: Matrix 3, overview of yield

Pool	Number scaffolds	$E > 10^{-5}$	NSS	Number viral scaffolds	Number remainders	$E > 10^{-5}$	NSS	Number viral remainders
HBZ-47	156	20	2	103	114	38	1	27
HBZ-48	65	7	1	31	74	24	0	29
HBZ-51	81	4	0	41	86	15	0	18
HBZ-52	65	8	0	35	52	19	0	6
HBZ-53	74	18	0	23	117	45	1	5
HBZ-54	83	2	0	49	65	9	0	23
HBZ-55	104	10	0	79	97	52	1	31
HBZ-64	57	15	1	12	103	45	2	1
HBZ-65	52	7	0	18	82	37	0	7
HBZ-66	37	4	1	0	33	17	0	1
Totals	774	95	5	391	823	301	5	148

Sequences are considered viral if $E < 10^{-5}$ to a viral sequence identified by CLC BlastN at NCBI.

NSS: no significant similarity ("not available" according to CLC Workbench)

12.3% scaffolds and 36.6% remainders had $E > 10^{-5}$

Table 7 Full-scale study: Viral sequences identified in Matrix 3 pools

Pool	Known viral species present¹	Genera to which new viral species belong²	Viral presence shown from mapping reads on scaffolds and contigs
HBZ-47	<ul style="list-style-type: none"> • Bean yellow mosaic virus • Canna yellow streak virus • Ranunculus mild mosaic virus • Tobacco rattle virus • Tomato spotted wilt virus 	<ul style="list-style-type: none"> • Potyvirus • Badnavirus 	<ul style="list-style-type: none"> • HBZ-47 reads mapped on potyvirus of HBZ-53 • HBZ-47 reads mapped on potyvirus of HBZ-55
HBZ-48	<ul style="list-style-type: none"> • Cycas necrotic stunt virus • Tomato spotted wilt virus 	<ul style="list-style-type: none"> • Nepovirus 	<ul style="list-style-type: none"> •
HBZ-51	<ul style="list-style-type: none"> • Alstroemeria mosaic virus • Lily symptomless virus • Tobacco rattle virus • Tobacco ringspot virus 	<ul style="list-style-type: none"> • Potyvirus 	<ul style="list-style-type: none"> • HBZ-51 reads mapped on Alstroemeria mosaic of HBZ-52 • HBZ-51 reads mapped on potyvirus of HBZ-52
HBZ-52	<ul style="list-style-type: none"> • Impatiens necrotic spot virus • Ranunculus leaf distortion virus 	<ul style="list-style-type: none"> • Partitivirus • Potyvirus 	<ul style="list-style-type: none"> • HBZ-52 reads mapped on Alstroemeria mosaic of HBZ-51 • HBZ-52 reads mapped on Lily symptomless of HBZ-51 • HBZ-52 reads mapped on potyvirus of HBZ-65
HBZ-53	<ul style="list-style-type: none"> • Petunia vein clearing virus • Ranunculus mild mosaic virus • Tomato spotted wilt virus 	<ul style="list-style-type: none"> • Potyvirus 	<ul style="list-style-type: none"> • HBZ-53 reads mapped on potyvirus of HBZ-47
HBZ-54	<ul style="list-style-type: none"> • Cycas necrotic stunt virus • Tobacco rattle virus • Tobacco ringspot virus 	<ul style="list-style-type: none"> • Nepovirus 	<ul style="list-style-type: none"> • HBZ-54 reads mapped on potyvirus of HBZ-47
HBZ-55	<ul style="list-style-type: none"> • Bean yellow mosaic virus • Canna yellow streak virus • Cycas necrotic stunt virus • Tomato spotted wilt virus 	<ul style="list-style-type: none"> • Badnavirus • Nepovirus • Potyvirus 	<ul style="list-style-type: none"> • HBZ-55 reads mapped on potyvirus of HBZ-47
HBZ-64	<ul style="list-style-type: none"> • Impatiens necrotic spot virus 	<ul style="list-style-type: none"> • Partitivirus 	
HBZ-65	<ul style="list-style-type: none"> • Petunia vein clearing virus • Ranunculus leaf distortion virus 	<ul style="list-style-type: none"> • Potyvirus 	<ul style="list-style-type: none"> • HBZ-65 reads mapped on potyvirus of HBZ-52
HBZ-66		<ul style="list-style-type: none"> • Ilarvirus 	<ul style="list-style-type: none"> • HBZ-66 reads mapped on potyvirus of HBZ-65

¹ Identified to species if there are scaffolds or contigs with >100 nt having >90% identity to the sequence of a known virus. ² Identified to genus if there are scaffolds or contigs with >100 nt having <90% identity to a viral sequence, and $E < 10^{-5}$.

Table 8 Confirmation of the presence of specific viral sequences by RT-PCR

Sample	Plant species	Study	Virus species or genus	% identity to reference	PCR product length (nt)
V033	Kalanchoe	Matrix 2	Kalanchoe mosaic virus	97	824
V057	Banana	Matrix 2	Banana mild mosaic virus	100	74
V082	Iris	Matrix 2	Iris mild mosaic virus	97	819
V082	Iris	Matrix 2	Ornithogalum mosaic virus	98	455
V090	Buttercup	Matrix 3	Ranunculus leaf distortion virus	99	814
V090	Buttercup	Matrix 3	Potyvirus	72	482
V090	Buttercup	Matrix 3	Potyvirus	79	479
V095	Buttercup	Matrix 3	Ranunculus mild mosaic virus	95	58
V095	Buttercup	Matrix 3	Potyvirus	70	470
V095	Buttercup	Matrix 3	Tomato spotted wilt virus	99	811
V098	Alstroemeria	Matrix 3	Alstroemeria mosaic virus	93	60
V098	Alstroemeria	Matrix 3	Potyvirus	72	233
V102	Petunia	Matrix 3	Potyvirus	68	640
V114	Lisianthus	Pilot	Eggplant mottled dwarf virus	97	435
V116	Peony	Matrix 3	Tobacco rattle virus	99	347
V132	Canna	Matrix 3	Bean yellow mosaic virus	86	363

RT-PCR was carried out with primers specific to the scaffold or remainder contig to be confirmed.

Table 9 Pilot and full-scale study: Types of viruses and viroids detected

Type of pathogen	Number of genera	Genera (number of species) observed
(+) ssRNA virus	15	Carlavirus (1), Carmovirus (1), Closterovirus (1), Cucumovirus (1), Enamovirus (1), Ilarvirus (2), Luteovirus (1), Nepovirus (2), Polerovirus (2), Potexvirus (1), Potyvirus (13), Tobravirus (1), Tombusvirus (2), Torradovirus (1), Unassigned Betaflexiviridae (1)
(-) ssRNA virus	2	Rhabdovirus (1), Tospovirus (2)
dsRNA virus	3	Amalgavirus (1), Endornavirus (2), Partitivirus (3)
dsDNA virus (including EPRVs)	3	Badnavirus (4), Caulimovirus (2), endogenous pararetrovirus (1)
ssDNA virus	1	Begomovirus (1)
Viroid	1	Hostuviroid (1), Viroid-like DNA (1)
Total	25	50

Table 10 (suppl) Pilot and Full-scale study: Complete compilation of viral/viroid sequences observed

Sample	Species	Study	Virus/viroid sequence identified (acronym)	Genus
V003	Cleome	Pilot	Tobacco ringspot virus (TRSV)	Nepovirus
V003	Cleome	Pilot	Tulip virus X (TVX)	Potexvirus
V012	Verbena	Pilot	Angelonia flower break virus (AnFBV)	Carmovirus
V036	Wild sweetpea	Pilot	Iarvirus	Iarvirus
V036	Wild sweetpea	Pilot	Enamovirus	Enamovirus
V036	Wild sweetpea	Pilot	Allexivirus	Allexivirus
V036	Wild sweetpea	Pilot	Luteovirus	Luteovirus
V054	Cineraria	Pilot	Tomato spotted wilt virus (TSWV)	Tospovirus
V054	Cineraria	Pilot	Soymovirus	Soymovirus
V074	Asclepias	Pilot	Pararetrovirus	Pararetrovirus
V077	Carnation	Pilot	Dianthus caryophyllus viroid-like DNA (CarSAVd)	Viroid-like DNA
V077	Carnation	Pilot	Deltapartitivirus	Deltapartitivirus
V101	Anemone	Pilot	Deltapartitivirus	Deltapartitivirus
V101	Anemone	Pilot	Amalgavirus	Amalgavirus
V101	Anemone	Pilot	Endornavirus	Endornavirus
V104	Petunia	Pilot	Petunia vein clearing virus (PVCV)	Badnavirus
V114	Lisianthus	Pilot	Eggplant mottled dwarf virus (EMDV)	Rhabdovirus
V142	Squash	Pilot	Squash chlorotic leaf spot virus (SCLSV)	Torradovirus
V004	Verbena	Matrix 1	Angelonia flower break virus (AnFBV)	Carmovirus
V005	Primrose	Matrix 1	none	
V009	Geranium	Matrix 1	Pelargonium flower break virus (PFBV)	Carmovirus
V009	Geranium	Matrix 1	Pelargonium line pattern virus (PLPV)	Tombusvirus
V009	Geranium	Matrix 1	Pelargonium vein banding virus (PVBV)	Badnavirus
V020	phlox	Matrix 1	none	
V021	Cleome	Matrix 1	Cucumber mosaic virus + satRNA (CMV)	Cucumovirus
V022	Carnation	Matrix 1	Carnation etched ring virus (CERV)	Caulimovirus
V022	Carnation	Matrix 1	Tomato spotted wilt virus (TSWV)?	Tospovirus
V026	Geranium	Matrix 1	Pelargonium flower break virus (PFBV)	Carmovirus
V026	Geranium	Matrix 1	Pelargonium vein banding virus (PVBV)	Badnavirus
V026	Geranium	Matrix 1	Tomato spotted wilt virus (TSWV)?	Tospovirus
V029	NG Impatiens	Matrix 1	Impatiens necrotic spot virus (INSV)	Tospovirus
V029	NG Impatiens	Matrix 1	Tomato spotted wilt virus (TSWV)?	Tospovirus
V031	Verbena	Matrix 1	Tomato spotted wilt virus (TSWV)?	Tospovirus
V119	Lisianthus	Matrix 1	none	
V146	Violet	Matrix 1	Viola white distortion associated virus (VWDaV)	Iarvirus
V027	Geranium	Matrix 2	Pelargonium vein banding virus (PVBV)	Badnavirus
V032	Hydrangea	Matrix 2	Potyvirus?	Potyvirus

V033	Kalanchoe	Matrix 2	Impatiens necrotic spot virus (INSV)	Tospovirus
V033	Kalanchoe	Matrix 2	Kalanchoe mosaic virus (KMV)	Potyvirus
V033	Kalanchoe	Matrix 2	Potyvirus	Potyvirus
V033	Kalanchoe	Matrix 2	Rhabdovirus	Rhabdovirus
V035	Lilac	Matrix 2	Caulimovirus	Caulimovirus
V035	Lilac	Matrix 2	Begomovirus	Begomovirus
V037	Periwinkle	Matrix 2	Endornavirus	Endornavirus
V037	Periwinkle	Matrix 2	Polerovirus	Polerovirus
V039	NG impatiens	Matrix 2	Tomato spotted wilt virus (TSWV)?	Tospovirus
V046	Dahlia	Matrix 2	Dahlia latent viroid (DLVd)	Viroid
V046	Dahlia	Matrix 2	Dahlia mosaic virus (DMV)	Caulimovirus
V046	Dahlia	Matrix 2	Tomato spotted wilt virus (TSWV)?	Tospovirus
V047	Centranthus	Matrix 2	Caulimovirus	Caulimovirus
V048	Oleander	Matrix 2	none	
V051	Chrysanthemum	Matrix 2	Potyvirus?	Potyvirus
V051	Chrysanthemum	Matrix 2	Tomato spotted wilt virus (TSWV)?	Tospovirus
V055	Velvet leaf	Matrix 2	Tomato spotted wilt virus (TSWV)	Tospovirus
V055	Velvet leaf	Matrix 2	viroid	Viroid
V057	Banana	Matrix 2	Banana mild mosaic virus (BanMMV)	Betaflexiviridae
V057	Banana	Matrix 2	Banana streak virus (BSV)	Badnavirus
V062	Pomelo	Matrix 2	Citrus tristeza virus (CTV)	Closterovirus
V062	Pomelo	Matrix 2	Pararetrovirus	Pararetrovirus
V063	Cyclamen	Matrix 2	Tomato spotted wilt virus (TSWV)?	Tospovirus
V063	Cyclamen	Matrix 2	Potyvirus	Potyvirus
V068	Alstroemeria	Matrix 2	Tomato spotted wilt virus (TSWV)?	Tospovirus
V069	Hydrangea	Matrix 2	none	
V072	Lily	Matrix 2	none	
V082	Iris	Matrix 2	Iris mild mosaic virus (IMMV)	Potyvirus
V082	Iris	Matrix 2	Ornithogalum mosaic virus (OrMV)	Potyvirus
V082	Iris	Matrix 2	Polerovirus	Polerovirus
V082	Iris	Matrix 2	Potyvirus	Potyvirus
V082	Iris	Matrix 2	Tomato spotted wilt virus (TSWV)?	Tospovirus
V084	Green bell	Matrix 2	Potyvirus?	Potyvirus
V084	Green bell	Matrix 2	Tomato spotted wilt virus (TSWV)	Tospovirus
V087	Snapdragon	Matrix 2	Tomato spotted wilt virus (TSWV)?	Tospovirus
V001	Petunia	Matrix 3	none	
V085	Guzmania	Matrix 3	Potyvirus?	Potyvirus
V086	Draceana	Matrix 3	none	
V088	Anemone	Matrix 3	Impatiens necrotic spot virus (INSV)	Tospovirus
V088	Anemone	Matrix 3	Partitivirus	Partitivirus
V089	Anemone	Matrix 3	Potyvirus	Potyvirus
V090	Buttercup	Matrix 3	Potyvirus (may be rest of RanLDV)	Potyvirus
V090	Buttercup	Matrix 3	Ranunculus leaf distortion virus (RanLDV)	Potyvirus
V095	Buttercup	Matrix 3	Potyvirus (may be rest of RanMMV)	Potyvirus

V095	Buttercup	Matrix 3	Ranunculus mild mosaic virus (RanMMV)	Potyvirus
V095	Buttercup	Matrix 3	Tomato spotted wilt virus (TSWV)	Tospovirus
V098	Alstroemeria	Matrix 3	Alstroemeria mosaic virus (AlMV)	Potyvirus
V098	Alstroemeria	Matrix 3	Lily symptomless virus (LSV)	Carlavirus
V098	Alstroemeria	Matrix 3	Potyvirus (may be rest of AlMV)	Potyvirus
V099	Primrose	Matrix 3	Tomato spotted wilt virus (TSWV)	Tospovirus
V100	Geranium	Matrix 3	none	
V102	Petunia	Matrix 3	Petunia vein clearing virus (PVCV)	Badnavirus
V102	Petunia	Matrix 3	Potyvirus	Potyvirus
V106	Dicentra	Matrix 3	Potyvirus?	Potyvirus
V116	Peony	Matrix 3	Potyvirus?	Potyvirus?
V116	Peony	Matrix 3	Tobacco rattle virus (TRV)	Tobravirus
V116	Peony	Matrix 3	Viroid-like DNA	Viroid-like DNA
V118	Peony	Matrix 3	Cycas necrotic stunt virus	Nepovirus
V118	Peony	Matrix 3	Tomato spotted wilt virus (TSWV)?	Tospovirus
V120	Snapdragon	Matrix 3	Cycas necrotic stunt virus (CNSV)	Nepovirus
V122	Lisianthus	Matrix 3	none	
V124	Hibiscus	Matrix 3	none	
V129	Cleome	Matrix 3	Tobacco rattle virus (TRV)	Tobravirus
V129	Cleome	Matrix 3	Tobacco ringspot virus (TRSV)	Nepovirus
V132	Canna	Matrix 3	Badnavirus	Badnavirus
V132	Canna	Matrix 3	Bean yellow mosaic virus (BYMV)	Potyvirus
V132	Canna	Matrix 3	Canna yellow streak virus (CaYSV)	Potyvirus
V132	Canna	Matrix 3	Potyvirus	Potyvirus?
V132	Canna	Matrix 3	Tomato spotted wilt virus (TSWV)?	Tospovirus
V133	Chrysanthemum	Matrix 3	none	
V134	Tea	Matrix 3	Potyvirus	Potyvirus
V137	Pansy	Matrix 3	Potyvirus?	Potyvirus
V139	Daisy	Matrix 3	Illarvirus	Illarvirus
V141	Holly	Matrix 3	none	
V145	Wild tulip	Matrix 3	none	

Table 11 Pilot and full-scale study: Previously undescribed pathogen sequences detected

Type of pathogen	Host plant	Viral genus
(+) ssRNA virus	Hydrangea	Potyvirus
	Periwinkle	Polerovirus
	Iris	Polerovirus
	Ranunculus	Potyvirus
	Canna	Potyvirus
	Daisy	Ilarvirus
	Anemone	Potyvirus
	Wild sweetpea	Luteovirus
	Wild sweetpea	Ilarvirus
	Squash	Torradovirus
(-) ssRNA virus	Kalanchoe	Rhabdovirus
dsRNA virus	Periwinkle	Endornavirus
	Anemone	Partitivirus
	Carnation	Partitivirus
	Anemone	Partitivirus
	Anemone	Endornavirus
	Anemone	Amalgavirus
dsDNA virus (including EPRVs)	Centranthus	Caulimovirus
	Pomelo	Pararetrovirus (endogenous?)
	Lilac	Caulimovirus
	Canna	Badnavirus
	Asclepias	Pararetrovirus (endogenous?)
ssDNA virus	Lilac	Begomovirus
Viroid	Abutilon	Hostuviroid
	Peony	Viroid-like RNA

Table 12 Pilot and full-scale study: New hosts of previously described viruses

Sample	Host plant	Virus identified	Genus
V003	Cleome	Tobacco ringspot virus (TRSV)	Nepovirus
V003	Cleome	Tulip virus X (TVX)	Potexvirus
V114	Lisianthus	Eggplant mottled dwarf virus (EMDV)	Rhabdovirus
V055	Velvet leaf	Tomato spotted wilt virus (TSWV)	Tospovirus