**vd-sRNA profiler: user-focused interface for viroid derived small RNA mapping and profiling**

Charith Raj Adkar-Purushothama1\*, Pavithran Sridharan Iyer2,#, Teruo Sano3, Jean-Pierre Perreault1,\*

1RNA Group/Groupe ARN, Département de Biochimie, Faculté de Médecine des Sciences de la Santé, Pavillon de Recherche Appliquée au Cancer, Université de Sherbrooke, Sherbrooke, QC, Canada

2Département de Physique, Université de Sherbrooke, Sherbrooke, QC, Canada

3Faculty of Agriculture and Life Science, Hirosaki University, Bunkyo-cho 3, Hirosaki 036-8561, Japan

#Current address: Applied Mathematics, Institute for Quantum Computing, University of Waterloo, Waterloo, ON, Canada

\*Corresponding authors:[charith.adkar@USherbrooke.ca](mailto:charith.adkar@USherbrooke.ca) (C.R.A.-P)

Jean-Pierre.Perreault@USherbrooke.ca (J.-P.P.)

Date of Submission: XX, June 2021

**Abstract**

Viroids are circular, highly structured, single-stranded, non-coding, RNA pathogens that are known to infect and cause disease in several plant species. Viroids are known to trigger the host plant's RNA silencing machinery. The detection of viroid-derived small RNAs (vd-sRNA) in viroid infected host plants opened a new avenue of study in viroid-host pathogenicity. Since then, several viroid research groups have studied the vd-sRNA retrieved from different viroid-host combinations. Such studies require the segregation of 21- to -24-nucleotide long small RNAs (sRNA) from a deep-sequencing databank, followed by the separation of the vd-sRNA from any sRNA within this group that showed sequence similarity with either the genomic or the antigenomic strands of the viroid. For visualization, such mapped vd‑sRNAs are then profiled on both the viroid’s genomic and antigenomic strands. Although several commercial interfaces are available, they are all programmed for linear RNA molecules. Hence, viroid researchers must develop a computer program that accommodates the vd-sRNAs that are derived from the circular viroid. This is laborious, and often creates a bottleneck for biologists. In order to overcome this constraint, and to help the viroid research community, a python-based pattern matching interface has been developed so as to be able to profile and map vd-sRNA on a circular genome in this study. Additionally, a feature of matching tolerance has been added so as to be able to map the vd-sRNA derived from the quasi-species. The efficacy of the software was tested using previously reported deep-sequencing data obtained from two independent studies. Clearly this novel software should be a key tool with which to both evaluate the production of vd-sRNA and profile them on their target viroid species.

**Introduction**

Viroids are both the simplest and the smallest known plant pathogens. They consist of a circular, single-stranded, non-coding RNA genome of 246-434 nucleotides (nt) in length [1]. Since viroids do not encode for any known peptides, they rely entirely on their sequence, structure and host factors in order to replicate and induce disease symptoms in their host plants. Upon infection, they induce a wide array of symptoms in their host plants, symptoms such as leaf epinasty, leaf distortion, stunting and flower distortion [2–4]. Due to both their highly base-paired secondary structures and their RNA-RNA mode of replication, viroids are both inducers and targets of RNA silencing [5].

RNA silencing (RNA interference [RNAi]) is a multi-layer defence system which protects plants from invading RNA pathogens such as viruses and viroids [6]. Silencing is triggered by the processing of either a double-stranded or a highly structured RNA by the host’s RNase III-type ribonucleases (i.e., DICER or DICER-LIKE), resulting in the production of small interference RNAs (siRNAs) of 21–24 nt in length [7]. In 2001, two research groups independently reported the presence of viroid-derived small RNAs (vd-sRNA) in plants infected with PSTVd by RNA gel blot assays [8,9], suggesting that viroids are the targets of RNA silencing. The introduction of next-generation sequencing (NGS) technology allowed for large-scale studies on the accumulation of such vd-sRNA in viroid-infected plants [10–13]. The role of vd-sRNA in the RNAi-mediated down-regulation of the host's transcripts and symptom induction has been explored using different host-viroid combinations [14–18]. This has been reviewed elsewhere [19].

Viroid researchers extensively use NGS to study the viroid-host interaction, emphasizing vd-sRNA production [20]. Profiling of vd-sRNAs on both the genomic and antigenomic strands of viroid yields several types of information such as, but not limited to: (i) the proportion of vd-sRNA produced from the genomic and the antigenomic strands of the viroid; (ii) the distribution of vd-sRNA on the genomic and the antigenomic strands; and, (iii) the regions of the viroid that are either susceptible to, or resist, the host’s RNA silencing machinery. Such information could be used to understand host-viroid species relationships [11,13,21], viroid quasi-species [22,23] and to develop viroid resistant plants [24,25]. First and foremost in order to obtain all of this information, is the mapping of deep-sequenced vd-sRNA on the viroid genomic and anti-genomic sequence. The biggest hurdle for such studies is the lack of software foe mapping the sRNA on a circular genome, such as that of viroids. Although several commercial platforms are available, they are all programmed for linear DNA or RNA molecules. Consequently, either viroid researchers develop their own interface, or depend on the work of bioinformaticians in order to progress in their studies.

The mapping of a vd-sRNA to a target viroid’s genomic and/or antigenomic strand sequence is innately tied to the classic example of determining the number of pattern matchings. In computer science, pattern matching is the process of checking a given sequence of characters (in the present scenario, nucleotides) that exists among the provided data. In other words, given two strings and , does string occur as a sub-string of? Knuth, Moris and Pratt first addressed the concept of pattern matching in 1970 [26]. Since then, many proposed solutions to this problem have been proposed. Pattern matching is extensively used in bioinformatics in order to determine the sequence similarity between the subject and query [27].

In the present study, a python language has been used to write the pattern matching program so as to provide a user-friendly and efficient bioinformatics tool to viroid researchers. Python is one of the few? computer languages that is easy to both read and implement. Additionally, it is an open-source language and runs on different platforms such as Mac, Windows and Unix. The development of an interface that was initially tested on an illustrative genomic sequence and sRNAs, and then the developed interface was applied to the previously reported NGS data, is reported.

**Materials and methods**

**Setting the parameters for pattern matching**

The counting version of the standard pattern matching is a problem of enumerating all the occurrences of the string in the string [28,29]. Computational biology provides an archetypal context for counting problems of this sort [27,30,31]. For example, if is considered as being an sRNA and as being a viroid genome, the real-world problem is computing the number of bindings of a vd-sRNA with the viroid genome. When an sRNA nucleotide sequence is represented by a string ,and the viroid genome by a string , computing the number is bindings of the sRNA sequence immediately admits the counting form of the standard pattern matching problem. Furthermore, one can safely assume that and are drawn from the alphabet . Before laying out the problem precisely, in order to solve several instances of the above counting problem, each corresponding to a different viroid RNA nucleotide . Hence, the problem that we would like to solve in this section is stated below.

*Definition 1: Viroid binding problem*

Given the following:

1. Set of N strings where each string has length is sampled from ,
2. Length m string sampled from the alphabet , and
3. Function from strings to in order to identify a matching between two strings. For two equal-length strings and , if and only if and match ~~and otherwise~~.

Compute the size of the sets for each , where:

Equation 1:

The definition of is intentionally unspecified so as to accommodate various scenarios in the present work. That said, is used in four different ways that can be divided into two categories: (i) the matching of in the forward and reverse directions with respect to ; and, (ii) the linear or circular topology of .

(i) the matching of in the forward and reverse directions with respect to :

The first set of variations occurs from choosing an alignment of the viroid genome relative to the sRNA. While the natural alignment, which is called the forward matching, coincides with the definition in “Definition 1”, its counterpart is referred to as the reverse matching set. The forward and reverse cases differ in their rule for deciding when two equal-length strings and, sampled from the same alphabet, are identical. The forward matching setting adopts the natural definition where if and only i , for all , where is the length of . On the contrary, the reverse matching setting places a non-standard requirement: if and only if , for all . Moving forward, only the case of forward matching is discussed as reverse matching adapts the same discussion of forward matching without affecting the conclusion.

Setting the topology is crucial in determining the boundaries for a string that is able to differentiate the mapping of sRNA derived from linear and circular genomes. The case of linear topology (used for sRNA derived from a linear genome), referred to as the linear alignment, refers to an with a finite boundary, that is to say, a length string with following for and no character following . On the contrary, in the case of the circular topology (used for sRNA derived from a circular genome), which is referred to as the circular alignment, is characterized by a gene with periodic boundary conditions, that is, a length string with the characterfollowing the character for . It is straightforward to note that the matching problem on a circular aligned gene trivially reduces to the linear alignment case. This can be realized by appending the gene string with the length substring of starting at position . Consequently, no conceptual distinction will be attributed to this setting, and only the solution for the linear alignment will be analyzed.

**Finding the occurrence of in**

This section presents the solution to forward matching with linear alignment for the pattern matching problem defined in *Definition 1*. First of all, the following useful notations need to be introduced.

1. The *length* of a string is simply the number of characters in . Often is used to denote the size of
2. The *Hamming weight* of a binary sequence is the number of ’s in the sequence. often is used to denote the Hamming weight of

A nsolution to the pattern matching problem entails a systematic search for a match between every sequence in to every substring of . In other words, it makes calls to an elementary routine which compares two strings, and , in order to identify the number of times occurs in . Consequently, the time complexity of the naive solution is . A solution that outperforms the naive one is the goal. This is achieved by mapping the pattern matching problem to a similar size matrix multiplication problem. The latter can be solved efficiently by vectorizing individual operations.

The first ingredient of this mapping is encoding each character in into a unique binary sequence of bits.

Equation 2:

The above encoding naturally implies an encoding for strings composed of the characters from by simply concatenating the binary sequences corresponding to the individual alphabets in the string. Likewise, a nucleotide sequence of length is identified by a binary sequence of bits, and is denoted by .

It is straightforward to notice that distinct characters in Equation 2 correspond to orthogonal binary sequences. Therefore, two nucleotide strings and , each of length , match exactly if and of only if the number of ones in the dot product: is equal to , that is to say . A similar observation can be made in order to account for mismatches wherein the the number of ones in should be at least . To summarize, two length $k$ strings and match approximately with at most mismatches if and only if the Hamming weight of is at most .

The above describes a method with which to check for a match between two strings of equal length, this can be readily extended to strings of unequal lengths such as when and for . To realize this, note that can be found in if and only if there are consecutive characters in that match with . In other words, for some . Furthermore, there are exactly substrings of that can be formed by taking consecutive characters, described by the set . Each of these substrings can be encoded into a binary sequence having bits. The substrings that match with will can be identified by checking if is at least .

Lastly, by combining the tests for matchings between each example of for which and into a single condition, that is to say letting be a matrix whose columns are ,, it turns out that the dot product is a vector of size whose th component satisfies if and only if the substring matches with at all but at most locations. In other words, the number of occurrences of in is determined by the number of entries in whose value is at least . This inference forms the backbone of the solution to the underlying pattern matching problem.

Summarizing the solution to the forward matching problem with linear alignment in definition 1. Recall that the problem here? is defined? by a set of sRNA nucleotides and a gene . The encoding in equation 1 leads to two key observations:

1. is mapped to an binary matrix whose th row denotes the binary encoding of the th nucleotide in the sRNA pool.
2. is mapped to a binary matrix whose th column denotes the binary encoding of the length substring of , starting at position .

These observations imply that the dot product yields an ( binary matrix, where is equal to if and only if the th sRNA nucleotide binds to the gene at position . Finally, taking the sum of all the rows of yields a vector of size where It is easy to realize that is nothing but the number of sRNA nucleotides in which binds to the gene at location . In other words, Equation 1 takes the form . This completely specifies our solution to the pattern matching problem. Casting the pattern matching problem as matrix multiplication is critical [32] to its low runtimes for large number of sRNA nucleotides. Table 1 shows the runtimes for computing the solutions to the pattern matching problem of? definition 1 for a few different sizes of the sRNA pool.

**Profiling of vd-sRNA on both viroid genomic and antigenomic strands**

The key input data for the software are the genome sequence, specified in a text file without line-breaks, and the pool of vd-sRNA, specified in a text file wherein each new line identifies a nucleotide sequence composed of “A”, “T”, “G” and “C” characters. If there are any nucleotide sequences that contain an “N”, they are ignored. In order to accommodate various formats, the program accepts a whole number which specifies the number of lines to skip before reading every nucleotide sequence in the pool file. Additional settings for the matching problem include: (i) tolerance: the maximum number of mismatches allowed in order to accommodate the vd-sRNAs derived from the quasi-species; (ii) topology of the gene, that is to say a Boolean expression that takes the value 0 for linear matching and 1 for circular matching; and, (iii) the number of cores to be used by the program. For example, “gene.txt pool.txt 2 1 1 1” is a complete input specification indicating that the problem of computing bindings in the circular topology condition, while allowing for, at most, one mismatch between the sequence in gene.txt with those in the pool.txt. All such instances of the matching problem can be gathered in a text file and placed in "vbind/data/input". The output format for the solution of the matching problem is a matrix, each of whose rows correspond to a unique length of vd-sRNA sequences in the pool. Each row is of a length equal to the gene sequence, where its i-th column specifies the number of matchings amongst sequences of a given length in the pool with the sequence of the gene starting at position 1.

**Viroid small RNA data**

Two previously published sRNA data sets obtained from tomato plants infected with variants of PSTVd were used in this study. Specifically, the total sRNA sequences obtained from tomato plants infected with PSTVd-I (GenBank Acc. No.: AY937179) or PSTVd-RG1 (GenBank Acc. No.: U23058) having the GEO Acc. No.: GSM1695657 and GSM1717894, respectively [16,33], were used.

**Software accessibility and instructions**

The script is available at <https://github.com/paviudes/vbind>. Example input and running instructions as shown in data/input/example.txt

**Results**

**Overview of sRNA mapping and profiling**

In order to verify the effectiveness of this new tool, a 30‑nt long DNA string (5’- GCT TCA GGG ATC CCC GGG GAA ACC TGG TCG-3’) was used as the “genome,” that is to say that it was the equivalent of the viroid genomic strand (Fig. 1A). From this DNA string, a raondom pool of sRNAs was prepared such that the pool contains at least one non-matching sRNA, one genomic sRNA and one antigenomic sRNAs of both 5- and 6-nts in length. Additionally, a couple of sRNAs were taken between positions 27 and 5 of the DNA string in order to imitate the sRNA derived from the circular genomic strand (Table 1). This pool is equivalent to the NGS data of the sRNAs. Running the patterning matching to find the matching of sRNAs in the pool with genome strand summarizes the results in the terminal as shown in figure 1 (panels A and B). Summary of obtained data includes the name of the gene used, the sRNA pool, the topology, the tolerance level, the cores used for the mapping, a table summarizing the total number of sRNAs of the different lengths, the number of sRNAs matching in the forward direction, the number of sRNAs matching in the reverse direction, the total number of sRNA matching and the percentage of sRNAs mapped for the given sRNA species. The details of each sRNA sequence that occurred in the pool are presented in Fig. S1 as they were generated by the vd-sRNA profiler.

In order to profile the sRNA on both the genomic and antigenomic strands of the viroid genome, the above matched sRNAs were run on the genome strand. As presented in Figure 1C, the X-axis indicates the length of the viroid RNA, while the Y-axis shows the number of matching sRNAs. By testing the above example, it is clear that here developed tool can detect the total sRNA count, the sRNAs matching the genomic and antigenomic strands, thus addressing the fundamental questions of mapping the sRNA from the NGS data on a given genome sequence. Further, profiling of the mapped sRNAs on the illustration DNA strand shows exactly how the vd-sRNA profiler developed in this study could be used to analyse vd-sRNAs obtained from viroid infected plants.

**Mapping of sRNAs obtained from PSTVd-I infected plants**

In order to map and profile the sRNAs on both the genomic and antigenomic viroid RNA strands, an sRNA pool of fragments 15- to 37-nts in length obtained from PSTVd-I infected tomato plants was retrieved from GEO. The data set was processed with the 359-nt long PSTVd-I genome sequence with zero tolerance so as to segregate the genomic and antigenomic matching vd-sRNAs.

As vd-sRNAs of 21- to 24-nt are the those of interest, the matchings obtained for these sizes are presented in Table 2. Out of 4,316,543 sRNA of 21- to 24-nt, a total of 488,176 vd-sRNAs showed 100% sequence similarity with PSTVd‑I sequence, accounting for 11.3% of the total recovered sRNAs. More specifically, 380,731 (8.8%) and 107,445 (2.5%) vd-sRNAs of 21- to 24-nt were derived from the genomic and antigenomic strands of PSTVd-I, respectively. In order to have a better picture of the expression levels of the individual-sized vd-sRNAs, the obtained vd‑sRNAs were normalized per one million reads. Detailed analysis showed that 22-nt long sRNAs had a maximum number of vd-sRNAs (34.3 %), whereas the least expressed vd-sRNA was 24-nt in length (1.2 %). Interestingly, the vd-sRNAs based on individual sizes revealed that the genomic strand produced more 22-nts sRNAs (29.3%), while the antigenomic strand produced more 21-nts sRNAs (7.9%). Overall, more genomic strand-derived sRNAs were covered as compared to the antigenomic strand of PSTVd-I. In other words, (+) vd-sRNA were expressed 3.5 times more than were (-) vd-sRNA. This can be attributed to the higher recovery of the genomic strand of viroid in infected plants as described elsewhere [34]. These data are in agreement with the previous report where the same sRNA pool was used for the analysis [16], confirming the reproducibility of the new software.

**Profiling the mapped sRNAs on PSTVd-I**

In order to examine the regions of PSTVd that produced most vd-sRNA, the above-mapped vd-sRNAs were plotted on both the genomic and the antigenomic strands of PSTVd-I. Hence, each 21- to 24-nt long vd-sRNA, and the cumulative 21- to 24-nt long vd-sRNAs, were profiled on PSTVd-I (Fig. 2). The data presented here clearly identify the regions that produce the most vd-sRNA on both the genomic and antigenomic strands. This result is in agreement with a previous report [16], confirming the reproducibility of this novel profiling tool.

**Accommodating vd-sRNA of the PSTVd-I quasi-species**

Viroids are known to form quasi-species in host plant [22,23,35–37]. That said, a single sequence type could give rise to several hundred or even thousands of sequence variants in host plants. In order to accommodate the vd-sRNA derived from the PSTVd-I sequence variants, in this study the vd-sRNA mapping against both the genomic and antigenomic strands of PSTVd-I was decreased by allowing for at least 1 mismatch. This decrease of the stringency increased the overall matching by 0.4% as compared to that obtained with 0 mismatchs (also called the "0 tolerance"), more precisely from 11.3 % to 11.7 % ( Table 3). In other words, 113,094 and 117,011 vd-sRNAs of 21- to 24-nts long are present per million sRNAs with zero tolerance and 1 mismatch, respectively. Allowing one mismatch increased the percentage of matching from as low as 0.1 % (for 24-nts long vd-sRNA) to much as 1.1 % (for 22-nts long vd-sRNA), as compared to having zero mismatches. However, the overall distribution of (+) vd-sRNA to (-) vd-sRNA remained almost the same as that seen with zero mismatches. Furthermore, as with zero tolerance mapping, 1 mismatch also had the highest number of genomic derived vd-sRNAs for 22-nts long sRNAs, whereas for the antigenomic derived vd-sRNAs the 21-nts long ones did. In order to evaluate the regions of PSTVd that produced the most vd-sRNA, the above mapped vd-sRNAs with 1 mismatch are plotted on both the genomic and the antigenomic strands of PSTVd-I as described above (Fig. 3).

**Evaluating the mapping tool on the vd-sRNAs obtained from PSTVd-RG1 infected plants**

In order to increase the confidence in the programing tool, one more sRNA data set that is publicly available was analyzed. Specifically, the sRNAs sequences obtained from tomato plants infected with the PSTVd-RG1 variant were used for the analysis. Both PSTVd-I and PSTVd-RG1 possess 359-nts long genomes. PSTVd-I and PSTVd-RG1 have 3 mismatches in their genomes and they induce intermediate and lethal disease symptoms in tomato cultivar Rutgers (*Solanum lycopersicum* cv Rutgers) [18]. As described earlier, a total of 730,499 sRNAs of 21- to 24-nts in length were mapped on PSTVd-RG1 in both the forward and the reverse directions in order to segregate the genomic (+) and the antigenomic (-) vd-sRNAs at both 0 and 1 mismatch tolerance. The resulting data was normalized per million reads (Table 4). The results showed that a total of 102,555 vd-sRNAs were recovered per one million reads, and these contained 88,493 (8.8 %) and 14,062 (1.4%) genomic and antigenomic strands derived vd-sRNA, respectively, that showed 100% sequence similarity with PSTVd-RG1. This accounted for 10.3% of all sequenced 21- to 24-nts long sRNAs. Out of all 21- to 24-nts long vd-sRNAs, the 21-nts long vd-sRNAs species are the most highly expressed (25.8%), while the 24-nts long vd-sRNAs are the least expressed (1.2%) for both genomic and anti-genomic strands of PSTVd-RG1.

In order to allow for the vd-sRNA derived from the sequence variants of PSTVd-RG1, one mismatch was allowed in the mapping. Decreasing the stringency increased the number for vd-sRNA by 3.3 % (from 102,555 per million reads with zero mismatches to 135,654 per million reads with one mismatch: Table 5). This was attributed to 2.9% and 0.4% increases in the genomic and antigenomic strands' vd-sRNAs as compared to the zero-tolerance mapping. Analyzing the vd-sRNAs derived from both the genomic and the antigenomic strands revealed that genomic strand-derived sRNAs were expressed at least 6.4 times higher than the antigenomic strand derived sRNAs.

In order to visualize the regions of PSTVd-RG1 that produced vd-sRNAs derived from both the genomic and the antigenomic strands, the mapped vd-sRNAs with both zero and 1 mismatch were plotted on the genomic and antigenomic strands of PSTVd-RG1 (Figure 4). The profiles obtained for 21-, 22-, 23-, 24-nts long vd-sRNA s, and the cumulative 21- to 24-nts long vd-sRNAs, are presented in Figure 4, panels A and B, for zero and 1 mismatch, respectively.

**Discussion**

Although viroids are single-stranded, circular RNA molecules, due to the facts that they: (i) possess sequence complementarity and, consequently, form highly base-paired secondary structures; and, (ii) replicate through either an asymmetric or a symmetric rolling circle mechanism, viroids act as both inducers and targets of the host’s RNA silencing machinery (reviewed in [38]). That said, upon infection, all viroids trigger RNA silencing, and this results in the cleavage of the viroid RNA into sRNAs of 21- to 24-nts in length. The accumulation of these [vd-sRNAs](javascript:;) has been extensively studied in different viroid-host combinations [10,11,13]. Since, viroids are non-coding pathogens, recent works were directed towards understanding the role of such vd-sRNAs in both its pathogenicity and its symptom induction [14–16,21,39–42]. Included in these studies was the prediction of the vd‑sRNA:target mRNA duplex formation and searching for the number of vd-sRNAs in the viroid infected plant that could potentially bind to a given target mRNA. For such studies, it is critical to map the vd-sRNAs obtained from NGS data derived from viroid infected plants on viroid genomic and antigenomic strand. Profiling of the obtained vd-sRNA on both the genomic and antigenomic strands of the viroid helps to identify which regions of viroid produce the most vd-sRNA and which the least. This information indirectly suggests which regions of the viroid are either more susceptible or more resistant to host RNA silencing. Furthermore, this type of information helps in the development of RNAi mediated viroid resistance transgenic plants [43]. In addition, the mapping of vd-sRNA, and their profiling on the viroid’s genomic and antigenomic strand, are very important in understanding the host-viroid interaction.

Viroids, being circular in nature, require a specific computer program in order to be able to map the sRNAs derived from the junction of first and last nucleotides of the viroid. This feature is currently not available in any commercial software. Here, a vd-sRNA profiler has been developed, more precisely a python-based program that is tailormade to address not only this issue, but also to detect both genomic and antigenomic strand derived sRNAs. Besides its capabilities for detecting mappings, the tool is attractive because of its efficiency (Table 6). For instance, the sRNA pool of 5.8 million reads can be solved in 30 minutes using a desktop computer with a single core, whereas using this new tool solving 1 million reads requires only approximately 2 minutes. This gain eliminates the need for the high-end hardware that are often required by commercial programs. Moreover, the fact that vd-sRNA profiler is written in python permits its customization for various studies of both circular and linear RNAs.

The solution to the case of a viroid binding problem, as computed by the vd-sRNA profiler, is represented by two matrices with whole number entries: (i) the solution for the forward matchings; and, (ii) the solution for the reverse matchings. In each of the two cases, the matchings are specified by a matrix – for the forward matchings and for the reverse matchings – of size where is the number of distinct nucleotide lengths in the pool and is the length of the gene. Each entry, or , is the number of sRNA sequences in the pool of length indexed by that have a forward or reverse matching, respectively, with a substring of the gene starting at position . In addition to solving the viroid bindings problem, the vd-sRNA profiler is capable of a visual representation of the output by profiling the vd-sRNA on the genomic and antigenomic strands of the viroid (Fig. 2; Fig.3, and Fig. 4). In these figures the X-axis shows the positions on the viroid sequence, the forward matching solutions are represented on the positive Y-axis and the reverse matchings on the negative Y-axis. For each index on the x-axis, the absolute value on the Y-axis for a fixed length is the maximum number of sequences in the pool that match with the substring of the gene starting at position . The area under the curve is filled in for visual appeal purposes. It is important to note that the message conveyed in Figures 2 and 4 is identical to that of previous studies [33,44], even though the latter one adopted a slightly different visual representation for the plot. In these earlier studies [33,44], a visual representation was used wherein the height of the curve at a given position on the X-axis denoted the number of vd-sRNA nucleotides that match with a substring containing the character of the gene at position . This can be derived from the matrices of the vd-sRNA profiler by use of a simple postprocessing routine.

In order to provide a user friendly and the efficient bioinformatics tool to viroid researchers, the vd-sRNA profiler was developed by pattern matching using python language. This software seamlessly allows researchers to map and plot vd-sRNAs on their parent viroid molecule. Choosing the different matching tolerances allows the user to consider and visualize the vd-sRNA derived from viroid quasi-species. This software will help viroid researchers in their studies both evaluating the production of vd-sRNA and its profiling on their target viroid species.

**Supplementary information**

**Fig. S1**. **Data generated from the vd-sRNA profiler showing the details of the sRNA sequences and their occurrence in the pool.** Note that if a single sRNA shows matching in more than one place on the genome sequence, it is then counted as the number of times it showed matching with the genome sequence.

**Author Contributions:**

Conceptualization: C.R.A.-P., P.S.I., T.S. and J.-P.P.; writing—original draft preparation: C.R.A.-P. and PI; writing—review and editing: C.R.A.-P., T.S. and J.-P.P.; supervision: J.-P.P.; funding acquisition: J.-P.P. All authors have read and agreed to the published version of the manuscript.

**Funding**

This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC, Grant Nos. 155219-17) to J-PP. The RNA group was supported by grants from the Université de Sherbrooke. J-PP holds the Research Chair of Université de Sherbrooke in RNA Structure and Genomics, and is a member of the Centre de Recherche du CHUS. The funders had no role in study design, data collection and analysis, decision to publish or in the preparation of the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

**Reference**s

1. Kitabayashi, S.; Tsushima, D.; Adkar-Purushothama, C.R.; Sano, T. Identification and molecular mechanisms of key nucleotides causing attenuation in pathogenicity of Dahlia isolate of potato spindle tuber viroid. *Int. J. Mol. Sci.* **2020**, *21*, doi:10.3390/ijms21197352.

2. Ding, B. The biology of viroid-host interactions. *Annu. Rev. Phytopathol.* **2009**, *47*, 105–131, doi:10.1146/annurev-phyto-080508-081927.

3. Tsagris, E.M.; Tsagris, E.M.; Martínez de Alba, A.E.; Martínez de Alba, A.E.; Gozmanova, M.; Gozmanova, M.; Kalantidis, K.; Kalantidis, K. Viroids. *Cell. Microbiol.* **2008**, *10*, 2168–79, doi:10.1111/j.1462-5822.2008.01231.x.

4. Flores, R.; Hernández, C.; Alba, A.E.M. de; Daròs, J.-A.; Serio, F. Di Viroids and Viroid-Host Interactions. *Annu. Rev. Phytopathol.* **2005**, *43*, 117–139, doi:10.1146/annurev.phyto.43.040204.140243.

5. Pallas, V.; Martinez, G.; Gomez, G. The interaction between plant viroid-induced symptoms and RNA silencing. *Methods Mol. Biol.* **2012**, *894*, 323–343.

6. Ding, B. Viroids: Self-replicating, mobile, and fast-evolving noncoding regulatory RNAs. *Wiley Interdiscip. Rev. RNA* 2010, *1*, 362–375.

7. Dadami, E.; Boutla, A.; Vrettos, N.; Tzortzakaki, S.; Karakasilioti, I.; Kalantidis, K. DICER-LIKE 4 But Not DICER-LIKE 2 may have a positive effect on potato spindle tuber viroid accumulation in nicotiana benthamiana. *Mol. Plant* 2013.

8. Itaya, A.; Folimonov, A.; Matsuda, Y.; Nelson, R.S.; Ding, B. Potato spindle tuber viroid as inducer of RNA silencing in infected tomato. *Mol. Plant. Microbe. Interact.* **2001**, *14*, 1332–1334, doi:10.1094/MPMI.2001.14.11.1332.

9. Papaefthimiou, I.; Hamilton, A.; Denti, M.; Baulcombe, D.; Tsagris, M.; Tabler, M. Replicating potato spindle tuber viroid RNA is accompanied by short RNA fragments that are characteristic of post-transcriptional gene silencing. *Nucleic Acids Res.* **2001**, *29*, 2395–2400, doi:10.1093/nar/29.11.2395.

10. Bolduc, F.; Hoareau, C.; St-Pierre, P.; Perreault, J.-P. In-depth sequencing of the siRNAs associated with peach latent mosaic viroid infection. *BMC Mol. Biol.* **2010**, *11*, 16, doi:10.1186/1471-2199-11-16.

11. Ivanova, D.; Milev, I.; Vachev, T.; Baev, V.; Yahubyan, G.; Minkov, G.; Gozmanova, M. Small RNA analysis of potato spindle tuber viroid infected phelipanche ramosa. *Plant Physiol. Biochem.* **2014**, *74*, 276–282, doi:10.1016/j.plaphy.2013.11.019.

12. Navarro, B.; Pantaleo, V.; Gisel, A.; Moxon, S.; Dalmay, T.; Bisztray, G.; Di Serio, F.; Burgyán, J. Deep sequencing of viroid-derived small RNAs from grapevine provides new insights on the role of RNA silencing in plant-viroid interaction. *PLoS One* **2009**, *4*, e7686, doi:10.1371/journal.pone.0007686.

13. Tsushima, D., Adkar-Purushothama, C.R., Taneda, A., and Sano, T. Changes in relative expression levels of viroid-specific small RNAs and microRNAs in tomato plants infected with severe and mild isolates of Potato spindle tuber viroid. *J. Gen. Plant Pathol* **2015**, *81*, 49–62.

14. Navarro, B.; Gisel, A.; Rodio, M.E.; Delgado, S.; Flores, R.; Di Serio, F. Small RNAs containing the pathogenic determinant of a chloroplast- replicating viroid guide the degradation of a host mRNA as predicted by RNA silencing. *Plant J.* **2012**, *70*, 991–1003, doi:10.1111/j.1365-313X.2012.04940.x.

15. Eamens, A.L.; Smith, N.A.; Dennis, E.S.; Wassenegger, M.; Wang, M.B. In Nicotiana species, an artificial microRNA corresponding to the virulence modulating region of Potato spindle tuber viroid directs RNA silencing of a soluble inorganic pyrophosphatase gene and the development of abnormal phenotypes. *Virology* **2014**, *450*–*451*, 266–277, doi:10.1016/j.virol.2013.12.019.

16. Adkar-Purushothama, C.R.; Brosseau, C.; Giguère, T.; Sano, T.; Moffett, P.; Perreault, J.-P. Small RNA Derived from the Virulence Modulating Region of the Potato spindle tuber viroid Silences callose synthase Genes of Tomato Plants. *Plant Cell* **2015**, *27*, 2178–94, doi:10.1105/tpc.15.00523.

17. Adkar-Purushothama, C.R.; Sano, T.; Perreault, J.-P. Viroid derived small RNA induces early flowering in tomato plants by RNA silencing. *Mol. Plant Pathol.* **2018**, doi:10.1111/mpp.12721.

18. Adkar-Purushothama, C.R.; Perreault, J.-P. Alterations of the viroid regions that interact with the host defense genes attenuate viroid infection in host plant. *RNA Biol.* **2018**, 1–12, doi:10.1080/15476286.2018.1462653.

19. Adkar-Purushothama, C.R.; Perreault, J. Current overview on viroid-host interactions. *Wiley Interdiscip. Rev. RNA* **2020**, *11*, e1570, doi:10.1002/wrna.1570.

20. Adkar-Purushothama, C.R.; Perreault, J.P. Impact of nucleic acid sequencing on viroid biology. *Int. J. Mol. Sci.* **2020**, *21*, doi:10.3390/ijms21155532.

21. Diermann, N.; Matoušek, J.; Junge, M.; Riesner, D.; Steger, G. Characterization of plant miRNAs and small RNAs derived from potato spindle tuber viroid (PSTVd) in infected tomato. *Biol. Chem.* **2010**, *391*, 1379–1390, doi:10.1515/BC.2010.148.

22. Brass, J.R.J.; Owens, R.A.; Matoušek, J.; Steger, G. Viroid quasispecies revealed by deep sequencing. *RNA Biol.* **2017**, *14*, doi:10.1080/15476286.2016.1272745.

23. Adkar-Purushothama, C.R.; Bolduc, F.; Bru, P.; Perreault, J.-P. Insights Into Potato Spindle Tuber Viroid Quasi-Species From Infection to Disease. *Front. Microbiol.* **2020**, *11*, 1235, doi:10.3389/fmicb.2020.01235.

24. Adkar-Purushothama, C.R.; Kasai, A.; Sugawara, K.; Yamamoto, H.; Yamazaki, Y.; He, Y.-H.; Takada, N.; Goto, H.; Shindo, S.; Harada, T.; et al. RNAi mediated inhibition of viroid infection in transgenic plants expressing viroid-specific small RNAs derived from various functional domains. *Sci. Rep.* **2015**, *5*, 17949, doi:10.1038/srep17949.

25. Kovalskaya, N.; Hammond, R.W. Molecular biology of viroid–host interactions and disease control strategies. *Plant Sci.* **2014**, *228*, 48–60, doi:10.1016/j.plantsci.2014.05.006.

26. Knuth, D.E.; Morris, Jr., J.H.; Pratt, V.R. Fast Pattern Matching in Strings. *SIAM J. Comput.* **1977**, *6*, doi:10.1137/0206024.

27. Gonnet, G.H. Some string matching problems from Bioinformatics which still need better solutions. *J. Discret. Algorithms* **2004**, *2*, 3–15, doi:10.1016/S1570-8667(03)00062-5.

28. Hjalte Wedel Vildhøj Topics in combinatorial pattern matching, Technical University of Denmark, 2015.

29. Chang, W.I.; Lampe, J. Theoretical and empirical comparisons of approximate string matching algorithms. In; 1992; pp. 175–184.

30. Deng, F.; Wang, L.; Liu, X. An efficient algorithm for the blocked pattern matching problem. *Bioinformatics* **2015**, *31*, 532–538, doi:10.1093/bioinformatics/btu678.

31. Bafna, V.; Muthukrishnan, S.; Ravi, R. Computing similarity between RNA strings. In; 1995; pp. 1–16.

32. Le Gall, F. Faster Algorithms for Rectangular Matrix Multiplication. In Proceedings of the 2012 IEEE 53rd Annual Symposium on Foundations of Computer Science; IEEE, 2012; pp. 514–523.

33. Adkar-Purushothama, C.R.; Iyer, P.S.; Perreault, J.-P. Potato spindle tuber viroid infection triggers degradation of chloride channel protein CLC-b-like and Ribosomal protein S3a-like mRNAs in tomato plants. *Sci. Rep.* **2017**, *7*, 8341, doi:10.1038/s41598-017-08823-z.

34. Hutchins, C.J.; Keese, P.; Visvader, J.E.; Rathjen, P.D.; McInnes, J.L.; Symons, R.H. Comparison of multimeric plus and minus forms of viroids and virusoids. *Plant Mol. Biol.* **1985**, *4*, 293–304, doi:10.1007/BF02418248.

35. Gago, S.; Elena, S.F.; Flores, R.; Sanjuán, R. Extremely high mutation rate of a hammerhead viroid. *Science* **2009**, doi:10.1126/science.1169202.

36. López-Carrasco, A.; Ballesteros, C.; Sentandreu, V.; Delgado, S.; Gago-Zachert, S.; Flores, R.; Sanjuán, R. Different rates of spontaneous mutation of chloroplastic and nuclear viroids as determined by high-fidelity ultra-deep sequencing. *PLOS Pathog.* **2017**, *13*, e1006547, doi:10.1371/journal.ppat.1006547.

37. Tessitori, M.; Rizza, S.; Reina, A.; Causarano, G.; Di Serio, F. The genetic diversity of Citrus dwarfing viroid populations is mainly dependent on the infected host species. *J. Gen. Virol.* **2013**, *94*, 687–693, doi:10.1099/vir.0.048025-0.

38. Pallás, V., Martínez, G., and Gómez, G. The interaction between plant viroid induced symptoms and RNA silencing. *Methods Mol. Biol.* **2012**, *894*, 323–343.

39. Bao, S.; Owens, R.A.; Sun, Q.; Song, H.; Liu, Y.; Eamens, A.L.; Feng, H.; Tian, H.; Wang, M.-B.; Zhang, R. Silencing of transcription factor encoding gene StTCP23 by small RNAs derived from the virulence modulating region of potato spindle tuber viroid is associated with symptom development in potato. *PLoS Pathog.* **2019**, *15*, e1008110, doi:10.1371/journal.ppat.1008110.

40. Adkar‐Purushothama, C.R.; Perreault, J. Current overview on viroid–host interactions. *WIREs RNA* **2020**, *11*, doi:10.1002/wrna.1570.

41. Thibaut, O.; Claude, B. Innate immunity activation and RNAi interplay in citrus exocortis viroid-tomato pathosystem. *Viruses* **2018**, *10*, 587, doi:10.3390/v10110587.

42. Aviña-Padilla, K.; Rivera-Bustamante, R.; Kovalskaya, N.Y.; Hammond, R.W. Pospiviroid infection of tomato regulates the expression of genes involved in flower and fruit development. *Viruses* **2018**, *10*, 516, doi:10.3390/v10100516.

43. Adkar-Purushothama, C.R.; Kasai, A.; Sugawara, K.; Yamamoto, H.; Yamazaki, Y.; He, Y.-H.; Takada, N.; Goto, H.; Shindo, S.; Harada, T.; et al. RNAi mediated inhibition of viroid infection in transgenic plants expressing viroid-specific small RNAs derived from various functional domains. *Sci. Rep.* **2015**, *5*, 17949, doi:10.1038/srep17949.

44. Adkar-Purushothama, C.R.; Brosseau, C.; Giguère, T.; Sano, T.; Moffett, P.; Perreault, J.-P. Small RNA Derived from the Virulence Modulating Region of the Potato spindle tuber viroid Silences callose synthase Genes of Tomato Plants. *Plant Cell* **2015**, *27*, 2178–2194, doi:10.1105/tpc.15.00523.

**Tables**

**Table 1. Summary of sRNA sequences used in the test run of the vd-sRNA profiler.**

| sRNA length | Tolerance: 0 | | Tolerance: 1 | | Non-matching sRNA |
| --- | --- | --- | --- | --- | --- |
| Forward matching | Reverse matching | Forward matching | Reverse matching |
| 5-nt | GGGAT | ACCAC | ACCCG | AGATT | AGGGA |
|  | ATCCC | CCACG | GGATC | CACGA | CAAGT |
|  | AAACC | CCGTG\* | TTGAC\* | CCCTG\* | GGGGC |
| 6-nt | TTCAGG | AGTTCC | ACAAAA | CAGAGG | ACGCAG |
|  | CGGGGA | GTGAAC\* | GGCTCA | GTGGAC\* | GTAGAT |
|  | CGGCTT\* | CCGTGA\* | CACCGA\* | CCGCGA\* | CGGAAA |

\*sRNAs derived between nucleotide positions 28 and 5, representing the sRNA of a circular genome.

**Table 2. Summary of the sRNAs identified by NGS from PSTVd-I inoculated tomato plants cv. Rutgers at 0 mismatch with the PSTVd-I genome.**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **sRNA length** | **Total sRNA in the pool** | **Tolerance: 0** | | | **Normalized reads (per million reads)** | | | | | | | **(+):(-) vd-sRNA** |
| **(+) vd-sRNA** | **(-) vd-sRNA** | **Total vd-sRNA** | **total sRNA** | **(+) vd-sRNA** | | **(-) vd-sRNA** | | **total vd-sRNA** | |
| **No. of reads** | **%** | **No. of reads** | **%** | **No. of reads** | **%** |
| 21-nt | 700,131 | 105,294 | 55,304 | 160,598 | 162,197 | 24,393 | 15.0 | 12,812 | 7.9 | 37,205 | 22.9 | 1.9 |
| 22-nt | 839,033 | 245,515 | 42,665 | 288,180 | 194,376 | 56,878 | 29.3 | 9,884 | 5.1 | 66,762 | 34.3 | 5.8 |
| 23-nt | 622,373 | 10,064 | 3,154 | 13,218 | 144,183 | 2,331 | 1.6 | 731 | 0.5 | 3,062 | 2.1 | 3.2 |
| 24-nt | 2,155,006 | 19,858 | 6,322 | 26,180 | 499,243 | 4,600 | 0.9 | 1,465 | 0.3 | 6,065 | 1.2 | 3.1 |
| 21-24-nt | 4,316,543 | 380,731 | 107,445 | 488,176 | 1,000,000 | 88,203 | 8.8 | 24,891 | 2.5 | 113,094 | 11.3 | 3.5 |

**Table 3. Summary of the sRNAs identified by NGS from PSTVd-I inoculated tomato plants cv. Rutgers at 1 mismatch with the PSTVd-I genome.**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **sRNA length** | **Total sRNA in the pool** | **Tolerance: 1** | | | **Normalized reads (per million reads)** | | | | | | | **(+):(-)**  **vd-sRNA** |
| **(+) vd-sRNA** | **(-) vd-sRNA** | **Total vd-sRNA** | **total sRNA** | **(+) vd-sRNA** | | **(-) vd-sRNA** | | **total vd-sRNA** | |
| **No. of reads** | **%** | **No. of reads** | **%** | **No. of reads** | **%** |
| 21-nt | 700,131 | 109,114 | 57,707 | 166,821 | 162,197 | 25,278 | 15.6 | 13,369 | 8.2 | 38,647 | 23.8 | 1.9 |
| 22-nt | 839,033 | 251,128 | 45,522 | 296,650 | 194,376 | 58,178 | 29.9 | 10,546 | 5.4 | 68,724 | 35.4 | 5.5 |
| 23-nt | 622,373 | 10,984 | 3,511 | 14,495 | 144,183 | 2,545 | 1.8 | 813 | 0.6 | 3,358 | 2.3 | 3.1 |
| 24-nt | 2,155,006 | 20,556 | 6,561 | 27,117 | 499,243 | 4,762 | 1.0 | 1,520 | 0.3 | 6,282 | 1.3 | 3.1 |
| 21-24-nt | 4,316,543 | 391,782 | 113,301 | 505,083 | 1,000,000 | 90,763 | 9.1 | 26,248 | 2.6 | 117,011 | 11.7 | 3.5 |

**Table 4. Summary of the sRNAs identified by NGS from PSTVd-RG1 inoculated tomato plants cv. Rutgers at 0 mismatch with the PSTVd-RG1 genome.**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **sRNA length** | **Total sRNA in the pool** | **Tolerance: 0** | | | **Normalized reads (per million reads)** | | | | | | | **(+):(-)**  **vd-sRNA** |
| **(+) vd-sRNA** | **(-) vd-sRNA** | **Total vd-sRNA** | **total sRNA** | **(+) vd-sRNA** | | **(-) vd-sRNA** | | **total vd-sRNA** | |
| **No. of reads** | **%** | **No. of reads** | **%** | **No. of reads** | **%** |
| 21-nt | 160,262 | 35,533 | 5,778 | 41,311 | 219,387 | 48,642 | 22.2 | 7,910 | 3.6 | 56,552 | 25.8 | 6.1 |
| 22-nt | 145,487 | 24,528 | 3,913 | 28,441 | 199,161 | 33,577 | 16.9 | 5,357 | 2.7 | 38,934 | 19.5 | 6.3 |
| 23-nt | 124,625 | 1,448 | 197 | 1,645 | 170,603 | 1,982 | 1.2 | 270 | 0.2 | 2,252 | 1.3 | 7.4 |
| 24-nt | 300,125 | 3,135 | 374 | 3,509 | 410,849 | 4,292 | 1.0 | 512 | 0.1 | 4,804 | 1.2 | 8.4 |
| 21-24-nt | 730,499 | 64,644 | 10,272 | 74,906 | 1,000,000 | 88,493 | 8.8 | 14,062 | 1.4 | 102,555 | 10.3 | 6.3 |

**Table 5. Summary of the sRNAs identified by NGS from PSTVd-RG1 inoculated tomato plants cv. Rutgers at 1 mismatch with the PSTVd-RG1 genome.**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **sRNA length** | **Total sRNA in the pool** | **Tolerance: 1** | | | **Normalized reads (per million reads)** | | | | | | | **(+):(-)**  **vd-sRNA** |
| **(+) vd-sRNA** | **(-) vd-sRNA** | **Total vd-sRNA** | **total sRNA** | **(+) vd-sRNA** | | **(-) vd-sRNA** | | **total vd-sRNA** | |
| **No. of reads** | **%** | **No. of reads** | **%** | **No. of reads** | **%** |
| 21-nt | 160,262 | 46,752 | 7,461 | 54,213 | 219,387 | 64,000 | 29.2 | 10,214 | 4.7 | 74,214 | 33.8 | 6.3 |
| 22-nt | 145,487 | 32,779 | 5,121 | 37,900 | 199,161 | 44,872 | 22.5 | 7,010 | 3.5 | 51,882 | 26.1 | 6.4 |
| 23-nt | 124,625 | 2,049 | 280 | 2,329 | 170,603 | 2,805 | 1.6 | 383 | 0.2 | 3,188 | 1.9 | 7.3 |
| 24-nt | 300,125 | 4,162 | 491 | 4,653 | 410,849 | 5,697 | 1.4 | 672 | 0.2 | 6,370 | 1.6 | 8.5 |
| 21-24-nt | 730,499 | 85,742 | 13,353 | 99,095 | 1,000,000 | 117,375 | 11.7 | 18,279 | 1.8 | 135,654 | 13.6 | 6.4 |

**Table 6. The runtime of the vd-RNA profiler on a single core of 3.2 GHz**

|  |  |  |  |
| --- | --- | --- | --- |
| **Size of the sRNA pool** | **Size of the viroid genome** | **Tolerance** | **Runtime in seconds** |
| 5875050 | 359 | 0 | 1450 |
| 5875050 | 359 | 1 | 1800 |
| 730499 | 359 | 0 | 128 |
| 730499 | 359 | 1 | 131 |

**Figure legends**

**Figure 1. vd-sRNA profiler data output for the test sample.** (**A**) Summary of the mapping data obtained from “0” tolerance and (**B**) “1” tolerance. (C) Profiling of the mapped sRNAs on the 30-nt long circular test genome at both the 0” and the “1” tolerance for 5- and 6-nts long sRNAs.

**Figure 2. Profiling of the vd-sRNAs recovered from PSTVd-I infected tomato plants at zero mismatches.** Sequence profiles of the (+) and the (−) PSTVd-sRNA populations recovered from the leaf tissues of infected tomato plants. Panels (A), (B), (C), (D) and (E) represent the profiles of the (+) and the (−) PSTVd-I derived sRNAs of sizes 21-, 22-, 23-, 24 and the cumulative of those of 21- to 24-nts, respectively. The data were normalized to reads per million.

**Figure 3**. **Profiling of the vd-sRNAs recovered from PSTVd-I infected tomato plants at 1 mismatch.** Sequence profiles of the (+) and the (−) PSTVd-sRNA populations recovered from the leaf tissues of infected tomato plants. Panels (A), (B), (C), (D) and (E) represent the profiles of the (+) and the (−) PSTVd-I derived sRNAs of sizes 21-, 22-, 23-, 24 and the cumulative of those of 21- to 24-nts, respectively. The data were normalized to reads per million.

**Figure 4**. **Profiling of the vd-sRNAs recovered from PSTVd-I infected tomato plants at zero and 1 mismatches.** Sequence profiles of the (+) and the (−) PSTVd-sRNA populations recovered from the leaf tissues of infected tomato plants. Panels (A), (B), (C), (D) and (E) represent the profiles of the (+) and the (−) PSTVd-I derived sRNAs of 0 mismatches of sizes 21-, 22-, 23-, 24- and the cumulative of those of 21- to 24-nts, while, Panels (F), (G), (H), (I) and (J) represent the profiles of the (+) and the (−) PSTVd-I derived sRNAs of 1 mismatch of sizes 21-, 22-, 23-, 24- and the cumulative of those of 21- to 24-nts, respectively. The data were normalized to reads per million.