Tropical Plant Pathology
https://doi.org/10.1007/s40858-023-00575-8

ORIGINAL ARTICLE

Metagenomic analyses of plant virus sequences in sewage water for plant viruses monitoring

Macária Ferreira Duarte[1] · Ikaro Alves de Andrade[1] · João Marcos Fagundes Silva[1] · Fernando Lucas de Melo[1] · Ana Maria Machado[2] · Alice Kazuko Inoue-Nagata[3] · Tatsuya Nagata[1]

Received: 24 January 2023 / Accepted: 25 March 2023

Abstract

Frequent monitoring of emerging viruses of agricultural crops is one of the most important missions for plant virologists. A fast and precise identification of potential harmful viruses may prevent the occurrence of serious epidemics. Nowadays, high-throughput sequencing (HTS) technologies became an accessible and powerful tool for this purpose. The major discussion of this strategy resides in the process of sample collection, which is usually laborious, costly and nonrepresentative. In this study, we assessed the use of sewage water samples for monitoring the widespread, numerous, and stable plant viruses using HTS analysis and RT-qPCR. Plant viruses belonged to 12 virus families were found, from which *Virgaviridae*, *Solemoviridae*, *Tymoviridae*, *Alphaflexiviridae*, *Betaflexiviridae*, *Closteroviridae* and *Secoviridae* were the most abundant ones with more than 20 species. Additionally, we detected one quarantine virus in Brazil and a new tobamovirus species. To assess the importance of the processed foods as virus release origins to sewage, we selected two viruses, the tobamovirus pepper mild mottle virus (PMMoV) and the carlavirus garlic common latent virus (GarCLV), to detect in processed food materials by RT-qPCR. PMMoV was detected in large amount in pepper-based processed foods and in sewage samples, while GarCLV was less frequent in dried and fresh garlic samples, and in the sewage samples. This suggested a high correlation of virus abundance in sewage and processed food sources. The potential use of sewage for a virus survey is discussed in this study.

Keywords Wastewater · Virome · High-throughput sequencing

Introduction

Pepper mild mottle virus (PMMoV) is a typical tobamovirus, with a single-stranded RNA genome and rod-shaped viral particles. Despite being a plant virus, it is one of the most abundant viruses in human fecal samples (Zhang et al. 2005; Nakamura et al. 2009). PMMoV particles are highly stable, preserving infectivity even after purification from human feces (Zhang et al. 2005). PMMoV has been considered as an useful indicator of human fecal contamination in various settings of urban wastewater worldwide (Rosario et al. 2009; Symonds et al. 2018; Kitajima et al. 2018; Gyawali et al. 2019; Bonanno Ferraro et al. 2021). In contrast, bacterial indicators, such as coliform bacteria, are known as less stable indicators of fecal contamination, due to false positive results (Brownell et al. 2007). In addition to PMMoV, other tobamoviruses and tymoviruses have also been found in human fecal samples in a reasonably high amount, being potential indicators (Zhang et al. 2005; Nakamura et al. 2009). As such, the detection of plant viral genome fragments in sewage water has already been reported by using the high-throughput sequencing (HTS) technology (Cantalupo et al. 2011; Ng et al. 2012; Bačnik et al. 2020). Since 2014, we have been conducting projects to study gastroenteritis viruses in sewage water and have discovered a significant amount of plant virus genomes in our analyses. These findings suggest that sewage virome analysis could be a valuable tool for conducting surveys on plant viruses.

In this study, we analyzed the plant virus genome fragments identified in raw sewage water collected at the Sewage Treatment Plant in Brasilia, Brazil, using the HTS approach.

✉ Tatsuya Nagata
tatsuya@unb.br

[1] Universidade de Brasília, Brasília, DF 70910-900, Brazil

[2] Companhia de Saneamento Ambiental Do DF, ETE-Norte, Brasília, DF 70800-130, Brazil

[3] Embrapa-Hortaliças, Brasília, DF 70351-970, Brazil

Published online: 28 April 2023

To understand the importance of the processed foods as virus release origins to sewage, PMMoV and garlic common latent virus (GarCLV) were chosen for detection by RT-qPCR in the processed and fresh food materials as well as in the sewage samples. The significance of the results of the sewage virome analyses, and the potential use of the metagenomic data for surveys on plant viruses are discussed.

Material and methods

Sample preparation and HTS

Raw sewage water was collected at the Sewage Treatment Plant (ETE-Norte) in Asa Norte, Brasilia, Brazil. Samplings were performed four times in different periods: March and May of 2016 (referred as S16-03 and S16-05), and May and August of 2020 (S20-05 and S20-08). The samples (300 mL each) were immediately cooled on ice, transported to the laboratory, and a virus semi-purification protocol was employed for virus enrichment, since virus genomes were expected at low concentration in sewage water. Briefly, sewage water was centrifuged at 5000 × *g* for 20 min. The supernatant was transferred to an ultracentrifuge tube on top of a 20% sucrose cushion, and centrifuged at 140,000 × *g* for 60 min. The pellets were resuspended in ultrapure water and stored at -80 ˚C; or in resuspension buffer of Quick-RNA Fecal/Soil Microbe Microprep Kit (Zymo Research, Irvine, USA) for RNA extraction, according to the manufacturer's instructions. The cDNA libraries (one library for each sample, in a total of four libraries) were constructed after bacterial rRNA removal (Illumina San Diego, USA) at Macrogen Inc. (Seoul, South Korea), and HTS was performed using HiSeq 4000 (samples for 2016) or NovaSeq 6000 (samples for 2020) with 100 bp paired-end in a 5G scale for each library at Macrogen Inc.

Raw reads were initially processed using BBDuk program of BBMap package (Bushnell et al. 2017) to remove low-quality sequences. After trimming, two approaches were used to analyze the plant virus sequences. The first one employed the taxonomic classification program for metagenomics "Kaiju" (Menzel et al. 2016) for the reads classification, performed with default parameters in Linux OS (https://github.com/bioinformatics-centre/kaiju.git) using a Viral Genome RefSeq database, and visualized by Krona viewer (https://github.com/marbl/Krona/wiki/KronaTools) in Linux OS. The second approach consisted in performing de novo assembly of reads using the MEGAHIT assembler (https://academic.oup.com/bioinformatics/article/31/10/1674/177884?login=false) with k-mer of 119. These contigs were exported to Geneious software v.11.0.5 (Dotmatics, Windhill, UK) and the contigs corresponding to virus genomes were identified by BLASTx searches against the RefSeq database of virus genomes in Geneious.

Virus genome sequence assembly and phylogenetic analysis

Contigs identified as plant viruses were extended by the "map to reference" command of the Geneious software using the reads (reads mapping to the contigs) to assemble the near-complete genomes. For phylogenetic analyses, the genomes were aligned with ClustalW in the Geneious software with relevant virus sequences and the phylogenetic trees were constructed using MEGA11 program (Tamura et al. 2021) using the Maximum Likelihood method and General Time Reversible model.

RT-qPCR

To assess the importance of the processed foods as virus release origins to sewage, we selected two viruses, PMMoV (tobamovírus) and GarCLV (carlavirus), to detect in processed food materials by RT-qPCR. The confirmation of the presence of these two viruses was done by RT-qPCR in the original sewage water and in processed and fresh food materials bought in the local market in Brasilia. Potential sources of PMMoV tested were chili pepper sauces, dried pepper flakes, paprika powder, and fresh chili pepper (Malagueta type, *Capsicum frutescens*); and of GarCLV dried garlic flakes and fresh cloves. RNA was extracted from the plant samples using total RNA purification kit (Cellco, São Carlos, Brazil). Total RNA of 500 ng from those materials was subjected to reverse transcription (RT) using SuperScript IV enzyme (Thermo Fisher, Waltham, USA) with a mixture of oligo-dT and random primers. One microliter of 20 µL of RT reaction solution was subjected to quantitative PCR (qPCR) using GoTaq® dye-based qPCR Master Mix (Promega, Madison, USA), with the primers listed in Supplementary Table 1. For absolute quantification of the amplicons, we used the cDNA clone of PMMoV (Junqueira et al. 2014) and the PCR product of GarCLV amplified by primers GarCLV_Control_4341_For and -4809_Rev (Supplementary Table 1) with a series of decimal dilutions ranging from 1 to 0.001 ng per tube as standards. The copy numbers were calculated by the molecular weight of each DNA fragments.

Results

Sewage samples were analyzed by HTS as a surveillance tool for plant viruses. Four samples were collected, in 2016 and 2020. First, the Kaiju analysis showed that in S16-03 sample (March, 2016) a total of 562,993 (out of 55,495,060) reads were classified as virus sequences, being ~ 14% (77,634

reads) identified as plant viruses; for S16-05 sample (May, 2016), 483,461 (out of 52,780,636) reads classified as virus sequences, and ~ 8% (40,775 reads) as plant viruses; for S20-05 (May, 2020), of the 537,608 (out of 57,649,960) virus reads ~ 8% (44,443 reads) classified as plant viruses; and for S20-08 (August, 2020) sample, ~ 9% (70.698) of the reads classified as plant viruses, from a total of 785,413 (out of 73,676,392) virus reads. On average, approximately 10% of the viral reads belonged to plant viruses, suggesting that they were present in a high amount in sewage water.

Further, analyses performed by the Kaiju software showed that in S16-03 sample 46% (35,938 reads) of the plant virus reads corresponded to viruses in the family *Virgaviridae*, 30% (23,379) to *Solemoviridae* and 8% (5,965) to *Tymoviridae* (Fig. 1). In the sample S16-05, the analysis showed that 43% (17,755) of the plant virus reads were classified as members of *Solemoviridae*, 35% (14,272) of *Virgaviridae,* and 10% (3,887) of *Betaflexiviridae* (Fig. 1). For the analysis performed with sample of S20-05, 23% (10,209) were classified as members of *Alphaflexiviridae*, 22% (9,679) as *Tymoviridae,* and 18% (7,796) as *Virgaviridae* (Fig. 1). In S20-08 sample, the reads of 36% (25,425) were of *Virgaviridae*, 22% (15,178) of *Alphaflexiviridae,* and 10% (6,164) of *Tombusviridae* (Fig. 1). In general, members of genus *Tobamovirus* were the most abundant, followed by those of *Sobemovirus*, *Potexvirus*, *Carlavirus*, *Tymovirus* and *Closterovirus*.

BLASTx searches of the contigs assembled by MEGAHIT showed that in samples collected in 2016 (Fig. 2a), viruses of nine plant virus families were identified, particularly those of *Betaflexiviridae* (25%), *Alphaflexiviridae* (17%), *Closteroviridae* (16%), and *Virgaviridae* (14%). In 2020, also nine plant families were identified (Fig. 2b): *Alphaflexiviridae* (26%), *Betaflexiviridae* (15%), *Virgaviridae* (15%), and *Secoviridae* (12%) (Fig. 2b).

For the species classification analysis, the threshold limit of 90% amino acid identity to the reference sequence was considered for the sake of simplicity. The identities were confirmed manually and individually, analyzing BLASTx alignments to exclude errors in the identification. Short contig alignment with the size less than 100 aa were not considered to eliminate possible false positive. Some contigs showed lower query coverages with translated query sequences by BLASTx due to the alignment to the read-through domains of the reference sequence (e.g., melon necrotic spot virus, MNSV, in Table 1); in other cases, the contig region was adjacent to two ORFs, resulting in split of coverage for two ORFs (resulting in lowering coverage). In all cases, the BLASTn analysis confirmed > 90% coverage with high identities. A total of 25 plant virus species were identified in the samples collected in 2016 (Table 1). The same viruses were present in 2020, except for two potexviruses, Cymbidium mosaic virus and Schlumbergera virus X (Table 1). It is important to note that the number of contig hits with known viruses was not correlated with the abundance of the virus genomic fragments in the sewage samples. For example, in the sample S20-05 (May, 2020), GarCLV had 18 contig hits, but only 914 reads (mean coverage of 10.7 to the whole genome) were mapped to the genome. On

Family	Genus	S16-03	S16-05	S20-05	S20-08
Alphaflexiviridae	*Allexivirus*	211	167	1,557	1,522
	Potexvirus	3,174	271	8,652	13,656
Betaflexiviridae	*Foveavirus*	23	29	15	34
	Carlavirus	3,222	3,858	2,526	5,130
Bromoviridae	*Ilarvirus*	5	1	15	5
	Cucumovirus	27	29	133	214
Closteroviridae	*Ampelovirus*	18	24	122	89
	Crinivirus	32	21	9	16
	Closterovirus	1,167	1,217	2,070	6,544
Geminiviridae	*Begomovirus*	2	3	10	41
Potyviridae	*Potyvirus*	376	317	385	3,055
Secoviridae	*Nepovirus*	12	4	13	22
	Comovirus	304	172	933	922
Solemoviridae	*Polerovirus*	67	47	62	125
	Sobemovirus	23,312	17,708	877	1,008
Tombusviridae	*Macanavirus*	4	4	28	67
	Luteovirus	5	4	0	3
	Alphacarmovirus	8	11	5	29
	Umbravirus	9	6	18	37
	Tombusvirus	17	4	1	42
	Aureusvirus	44	16	38	287
	Pelarspovirus	63	26	79	151
	Betacarmovirus	94	38	123	215
	Machlomovirus	184	60	327	593
	Gammacarmovirus	272	394	2,038	1,745
	Panicovirus	598	208	2,383	2,995
Tospoviridae	*Orthotospovirus*	12	8	1	49
Tymoviridae	*Maculavirus*	35	26	392	389
	Tymovirus	5,930	722	9,287	1,257
Virgaviridae	*Tobravirus*	7	6	7	28
	Tobamovirus	35,931	14,266	7,789	25,397

Number of reads
0 >1,000

Fig. 1 A heatmap table showing the taxonomic classification of plant virus reads identified by the Kaiju software in raw sewage samples of March and May 2016, and May and August 2020. Bluish color background indicates higher numbers of reads in the category

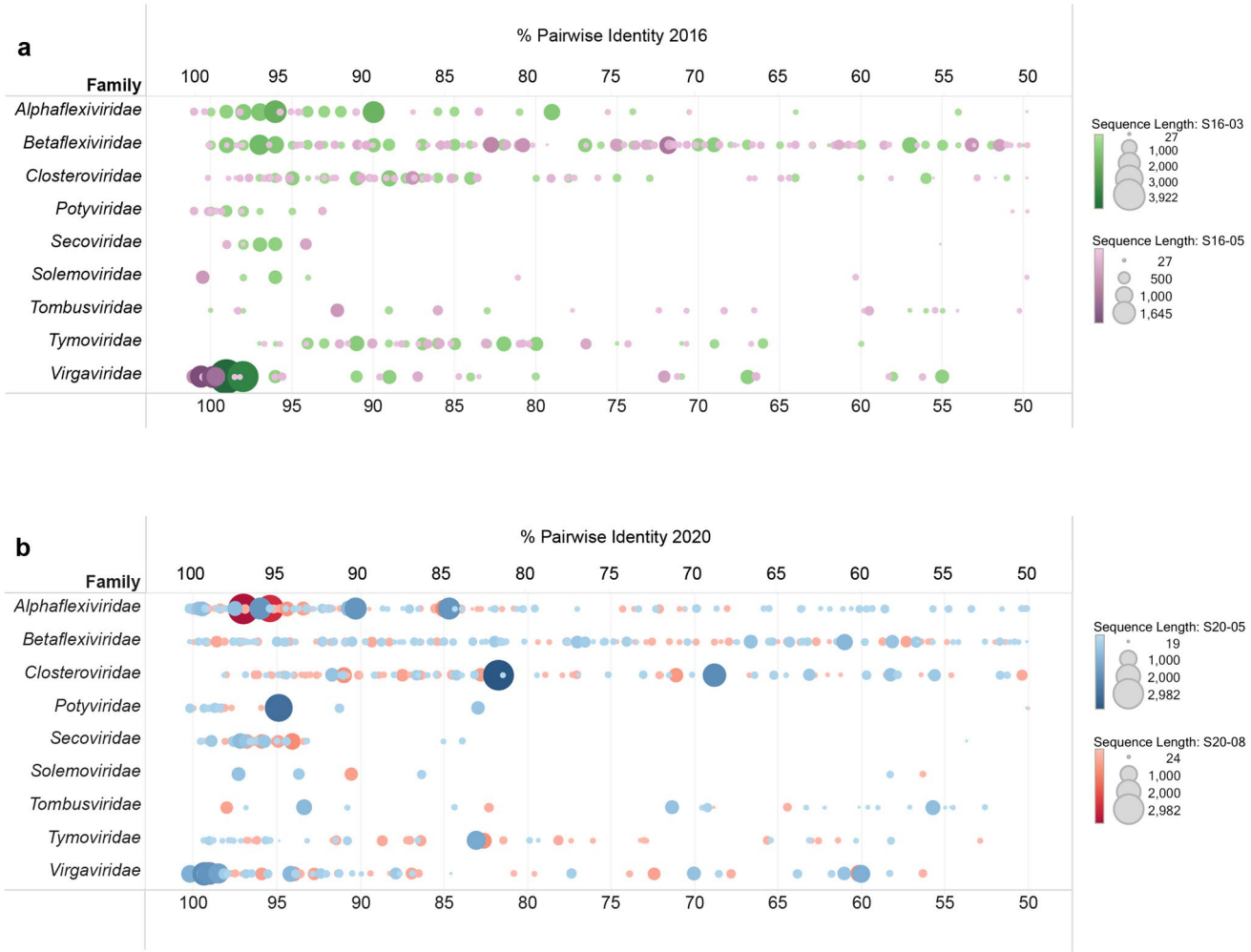

Fig. 2 Amino acid sequence identity distribution of plant virus-related contig sequences from sewage samples collected in 2016 (**a**) and 2020 (**b**). The assembled sequences were compared to sequences in the BLASTx protein database. One spot represents one contig, and the coverage length is indicated by the spot size. Nine families were detected

the other hand, PMMoV had 5 contig hits, but 6,003 reads (mean coverage of 95.3) were mapped.

Interestingly, four viruses were recently first reported in Brazil: garlic latent virus (Da Silva et al. 2019), MNSV (Moura et al. 2018), tobacco mild green mottle virus (Favara et al. 2019), and tomato mottle mosaic virus (Nagai et al. 2018). Since these viruses were detected in the sewage samples of 2016, it implies that these viruses were already circulating before their first reports. Notably for MNSV, it was listed as a quarantine virus by the Ministry of Agriculture, Livestock and Food Supply (MAPA) in Brazil before the first report in 2018. MNSV was detected in both years, 2016 and 2020, suggesting that it might be widespread in the country. Given the importance of this virus, the near-complete genomes were assembled (accession numbers LC745663-LC745666) using reads of the four HTS data and a phylogenetic tree was constructed after alignment with all the 29 full-length genome sequences of MNSV. All four Brazilian MNSV sequences clustered together showing that the same lineage of MNSV was present since 2016. They were closely related with the isolates from North America and Spain (Fig. 3a).

Using the HTS analysis on sewage water, we discovered two other viruses still not reported in Brazil, and named virus A and virus B (Table 1). Due to the Brazilian MAPA regulation, the publication of pests for the first time in Brazil must be preceded by its official communication to the MAPA. This procedure was not done yet, because the host and the location of the incidence are not known. Based on this situation and considering the scope of this study, we report the detection of these viruses omitting their identification. The virus A and virus B are two known viruses, and virus B is a quarantine virus in Brazil. We are currently searching the host plants of these viruses.

Table 1 BLASTx analysis of HTS obtained contigs for virus identification with amino acid sequence identities above 90%

Family	Genus	Virus	% Pairwise identity	E-value	% Query coverage	Number of contigs (Alignment coverage in amino acid)			
						S16-03	S16-05	S20-05	S20-08
Alphaflexiviridae	*Allexivirus*	garlic virus A	90.5–100.0	0 ~ 9.80e-94	38.4–99.8	2 (115–248)	1 (100)	13 (102–690)	8 (107–786)
	Potexvirus	Cactus virus X	90.0–100.0	0 ~ 7.26e-77	70.3–99.9	4 (110–1,544)	4 (118–200)	7 (110–1,543)	1 (1430)
		Cymbidium mosaic virus	92.3–99.1	1.31e-108 ~ 5.36e-68	99.4–99.6	3 (106–174)	0	0	0
		Schlumbergera virus X	98.2–100.0	2.62e-64 ~ 4.43e-72	99.7–100.0	2 (113–200)	0	0	0
		Zygocactus virus X	90.1–100.0	0 ~ 9.79e-81	53.4–100.0	8 (110–1,544)	4 (126–188)	19 (112–1,425)	7 (106–1,546)
Betaflexiviridae	*Carlavirus*	cole latent virus	91.7–99.1	1.52e-119 ~ 9.52e-52	59.4–99.6	2 (138–145)	1 (179)	2 (104–179)	1 (213)
		cowpea mild mottle virus	91.7–94.9	0 ~ 3.92e-96	61.8–82.5	0	1 (288)	1 (188)	1 (137)
		garlic common latent virus	91.8–100.0	0 ~ 9.51e-86	53.9–99.9	14 (101–203)	0	16 (104–291)	7 (108–329)
		garlic latent virus	90.0–99.0	0 ~ 8.03e-72	88.5–99.8	2 (126–220)	2 (106–134)	3 (106–264)	11 (104–331)
		melon yellowing-associated virus	90.0–99.2	0 ~ 6.14e-61	71.4–100.0	6 (100–743)	8 (107–238)	8 (107–238)	2 (162–212)
		potato virus M	90.2–96.1	0 ~ 6.00e-91	99.3–100.0	1 (137)	1 (357)	2 (103–114)	3 (149–220)
Closteroviridae	*Closterovirus*	citrus tristeza virus	90.0–99.2	0 ~ 9.75e-103	52.5–100.0	8 (116–376)	15 (101–331)	29 (101–412)	7 (101–594)
Potyviridae	*Potyvirus*	zucchini yellow mosaic virus	92.3–100.0	0 ~ 8.66e-72	91.3–99.9	9 (108–279)	8 (108–265)	14 (111–265)	7 (103–323)
Secoviridae	*Comovirus*	squash mosaic virus	93.0–99.4	0 ~ 8.81e-71	64.0–100.0	8 (104–379)	2 (268–452)	17 (131–666)	24 (104–547)
Solemoviridae	*Sobemovirus*	papaya lethal yellowing virus	96.5–99.5	0	80.1–99.7	1 (600)	0	0	0
		southern bean mosaic virus	90.5–98.3	0 ~ 9.52e-125	59.0–99.7	2 (116–175)	0	0	1 (420)
Tombusviridae	*Gammacarmovirus*	melon necrotic spot virus	91.4–100.0	0 ~ 5.34e-56	55.8–99.3	1 (101)	2 (193–614)	2 (193–497)	1 (787)
Tymoviridae	*Tymovirus*	tomato blistering mosaic virus	90.8–97.2	0 ~ 9.23e-179	53.0–99.9	6 (107–343)	5 (105–288)	6 (105–304)	0

Table 1 (continued)

Family	Genus	Virus	% Pairwise identity	E-value	% Query coverage	Number of contigs (Alignment coverage in amino acid)			
						S16-03	S16-05	S20-05	S20-08
Virgaviridae	*Tobamovirus*	Virus A	91.1–97.9	0 ~ 7.33e-105	99.0–100.0	1 (315)	3 (112–225)	6 (112–498)	3 (237–273)
		pepper mild mottle virus	93.6–100.0	0 ~ 9.21e-148	50.7–99.7	4 (154–1,205)	2 (532–1,168)	5 (128–1,612)	1 (1,612)
		tobacco mild green mosaic virus	94.0–99.7	0 ~ 3.57e-84	62.8–99.8	1 (111)	0	4 (130–697)	2 (785–816)
		tobacco mosaic virus	91.0–100.0	0 ~ 9.50e-67	57.2–100.0	7 (115–576)	0	8 (103–208)	14 (102–365)
		tomato mosaic virus	98.9–99.0	0	75.0–76.3	1 (1,605)	1 (1,605)	1 (1,612)	1 (1,605)
		tomato mottle mosaic virus	95.9–100.0	0 ~ 5.25e-103	54.5–100.0	1 (1,605)	0	4 (108–651)	2 (199–1,290)
	Quarantine virus	Virus B	99.6–100.0	0 ~ 6.42e-98	68.9–99.8	1 (1,648)	1 (1,645)	1 (1,645)	1 (147)

In addition to the 25 viruses, a potential novel virus was found in our datasets. Its genome organization and the phylogenetic relationship suggest that this novel virus is a member of the genus *Tobamovirus*. It was provisionally named as sewage-associated tobamovirus, acronym SaTV, accession number LC745662. The near-complete assembled genome is 6,777 nucleotides-long and presents unique tobamoviral genome organization. Its sequence was aligned with tobamovirus complete genome sequences, and a phylogenetic tree was constructed (Fig. 3b). Maracuja mosaic virus (MarMV) and passion fruit mosaic virus (ParMV) were most related tobamovirus to SaTV. SaTV has a cysteine-rich protein gene (Fig. 3c) which is unique for SaTV, MarMV and ParMV in *Tobamovirus* genus. As the sequence is highly related to passion fruit tomaboviruses, we are searching for the viruses at first in these plants.

In the data sets, other sequences were found that share < 90% amino acid identities with known viruses. They are listed in Supplemental Tables 2 (identity of 80.0–89.9%) and 3 (70.0–79.9%). It implies that these sequences are from potentially new virus species or new variants, according to the taxonomical criteria, and that a thorough analysis is required for their classification.

For validation of the HTS obtained genome sequences, we hypothesized that processed foods would be important virus sources to sewage/wastewater. For testing this hypothesis, two viruses, PMMoV and GarCLV, were chosen and RT-qPCR were performed using RNA extracted from the processed foods and from fresh materials, as well as from the sewage samples collected in 2020. These viruses were chosen because chili and garlic are often consumed as processed foods. Chili dried flakes, chili pepper sauce, paprika powder and fresh malagueta pepper fruits (*Capsicum frutescens*) were used to detect the tobamovirus PMMoV, and garlic dried flakes and fresh cloves were used to detect GarCLV. Remarkably, we detected PMMoV in the sewage sample, and in all types of processed food materials and fresh pepper (Table 2), including the chili pepper sauce which contains vinegar, and hence the sauce has a low pH of 3 ~ 4. This indicated that PMMoV is abundant and stable in the processed foods.

The presence of GarCLV was confirmed in the sewage, but in a low amount, and only in one of the two samples (Table 2). In contrast, it was found in higher amount in dried flakes. In the fresh garlic sample, the virus was not detected. In general, GarCLV was less abundant than PMMoV.

Discussion

In our study, viruses of the plant virus families *Alphaflexiviridae*, *Betaflexiviridae*, *Closteroviridae*, *Secoviridae*, *Solemoviridae*, *Tombusviridae*, *Tymoviridae*, and *Virgaviridae* were the most commonly found in sewage samples, confirming previous studies (Cantalupo et al. 2011; Ng et al. 2012; Fernandez-Cassi et al. 2018; Martínez-Puchol et al. 2020). We believe, though, that the abundance of these viruses in sewage samples are not completely correlated with

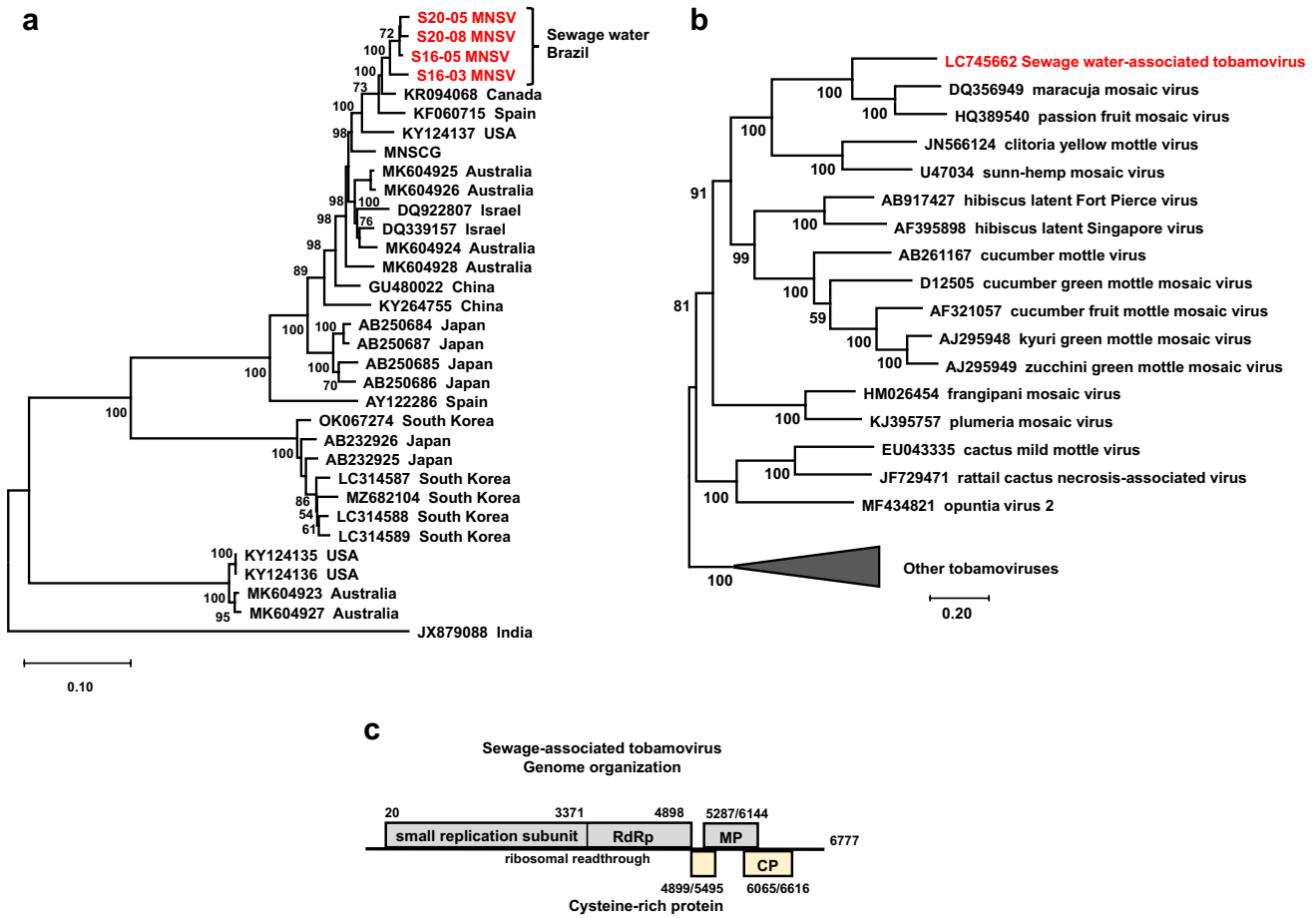

Fig. 3 Analyses of the melon necrotic spot virus (MNSV) and sewage-associated tobamovirus sequences obtained by HTS. Phylogenetic trees were constructed based on the maximum-likelihood method using MEGA11, for MNSV sequences **(a);** and for sewage-associated tobamovirus sequence aligned with tobamoviruses **(b)**. **c**. Genome organization of sewage-associated tobamovirus with position of each ORF (above or below) and gene name. Accession numbers are indicated. For MNSV sequences, the isolated country is also indicated

their incidence in the fields. At least two aspects should be considered for this analysis. (1) The first aspect is related to virus stability. Tospoviruses, which are very common pathogens of tomato and zucchini plants in Brazil, were rarely identified in sewage samples (Fig. 1). The viruses of this group have virus particles with a lipid membrane, and thus are likely fragile in the wastewater environmental condition and in the human digestive tract, resulting in rapid degradation. Therefore, viruses with high stability are expected to be preserved for a longer time in the sewage water condition. On the other hand, the property of particle stability may not be a general characteristic of a genus or a family. For example, zucchini yellow mosaic virus (ZYMV) was a potyvirus frequently detected in our sewage samples (Table 1). However, other potyviruses that infect cucurbit plants, such as watermelon mosaic virus and papaya ringspot virus-W (Supplementary Table 2), two other important potyviruses for cucurbits in Brazil, were rarely detected. Furthermore, other common potyviruses, such as potato virus Y and pepper yellow mosaic virus in solanaceous plants were not detected in our sewage samples. Thus, it is possible that ZYMV may have unique feature in particle stability among potyviruses. (2) The second condition is the productivity of the infected plants. Tospoviruses cause severe symptoms and, hence, a drastic reduction in leaf and fruit production (hence lower circulation of infected crop in market), while PMMoV is known to cause mild symptoms in pepper plants. It implies that viruses that cause mild or no symptoms will stay more in food materials, then released more in the sink or ingested/dispersed in the sewage.

Our results open important questions about the potential use of sewage water analysis for plant virus surveys, as discussed in Bačnik et al. (2020). For example, the presence of MNSV in sewage was detected since 2014, when we started analyzing sewage virome by HTS (data not shown). In the present study, we found MNSV genomic fragments

Table 2 Amount expressed in copy numbers of PMMoV and GarCLV detected by RT-qPCR

Sample type	Sample	Number of copies in total RNA from 1 mL (sewage water) or 1 ng (plant material)	
		PMMoV detection	GarCLV detection
Sewage water	S20-08	11,120,384	28
	S20-05	1,213,928	0
Fresh chili pepper	*Capsicum frutescens* (Malagueta)	2,580	nt
Chili pepper flakes (dried)	Supplier L	47,456	nt
	Supplier KS	67,138	nt
Paprika powder	Supplier L	350,598	nt
	Supplier K	4,293,756	nt
Chili pepper sauce	Supplier KN	418	nt
	Supplier C	2,002	nt
Garlic flakes (dried)	Supplier L	*nt	4,156
	Supplier K	nt	2,114
Fresh garlic	Supplier A	nt	0
	Supplier B	nt	0

*nt = not tested. Single sample was used for each supplier

in the samples of 2016 and 2020. The natural occurrence of MNSV in Brazil was firstly reported only in 2018 (Moura et al. 2018), and thus removed from the list of quarantine virus in Brazil. When a new virus is found in a sewage sample, it suggests that the virus may be present in the food, and consequently in commercial fields. Therefore, we are now trying to find the virus A and the virus B in plants to confirm their natural occurrence in Brazil. The virus B is currently listed as a quarantine virus in Brazil, and it has been found in our sewage samples of 2016 and 2020, and also in another sample collected in 2014 (data not shown). However, we should be aware about the possibility that this virus' genome fragments are present in imported (processed) foods.

In order to demonstrate the efficiency of the method for identifying circulating viruses, we selected two viruses, PMMoV and GarCLV, for detection and quantification by RT-qPCR in the processed foods, as test cases. Interestingly, the amount of PMMoV found in the processed foods as well as in sewage samples was extremely high, implying in the high stability of the particles in *in vitro* conditions. The virus was detected in all samples, chili pepper (*C. frutescens*), flakes (usually made of Dedo-de-moça type chili, *C. baccatum*), paprika (*C. annuum*), and chili sauce (usually made of chili *C. chinense* and *C. frutescens*), indicating that PMMoV occurs widely in many pepper types. Surprisingly, we were able to detect PMMoV in the chili pepper sauce, which is very acid (pH 3 ~ 4). Although high amounts of PMMoV was found in sewage water, we do not believe that PMMoV is waterborne virus. We tried to inoculate the semi-purified virus solution from sewage water in *Nicotiana benthamiana* plants (highly susceptible host of PMMoV), but the plants were not infected (data not shown).

Bačnik et al. (2020) reported that PMMoV, tomato mosaic virus, and tobacco mild green mosaic virus, all belonging to the *Tobamovirus* genus, were successfully recovered via mechanical inoculation of concentrated wastewater fractions in *N. benthamiana* and *N. occidentalis* plants. In contrast, inoculation with non-concentrated wastewater did not result in infections. Although these tobamoviruses exhibit infectivity, the viruses present in sewage are unlikely to serve as a source of new plant infections. Nevertheless, further surveys are necessary to explore the possibility of tobamoviruses being water-borne.

One of the great advantages of the plant virus study by using the sewage water source is the sample collection process. Visiting the production fields is laborious, and costly, especially in a continental country like Brazil. The collection may not represent the overall diversity of plants and localities. The use of sewage water, on the other hand, may provide the collection of plant samples from multiple sources that were produced in many different areas.

Epidemiological studies based on sewage samples have a potential to predict the onset of human viral diseases (Xagoraraki and O'Brien 2020), including SARS-CoV2 that caused an important pandemics in 2020–2021 (Bonanno Ferraro et al. 2022). Here, we demonstrate the validity of the use of sewage samples to predict or confirm outbreaks or emergence of plant viral diseases. Furthermore, the sewage water analysis may contribute to a phytosanitary alert by monitoring the presence of pathogenic plant viruses, for

those viruses with stable particles in environmental water samples. It is worth mentioning that there are limitations on the use of plant virus metagenomic data in sewage samples for plant virus survey and extreme care is need for interpretation of the results.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40858-023-00575-8.

Acknowledgements This study was funded by FAP-DF and CAPES

Author contribution Study conception and design: MF Duarte and T Nagata. Acquisition of data: MF Duarte, IA Andrade, AM Machado, and T Nagata. Data analysis and interpretation: MF Duarte, JMF Silva, FL Melo, AK Inoue-Nagata, and T Nagata. Manuscript drafting: MD Duarte, JMF Silva, AK Inoue-Nagata, and T Nagata. Critical revision: MD Duarte, JMF Silva, AK Inoue-Nagata, and T Nagata.

Data availability Data will be made available on reasonable request.

Declarations

Conflict of interest All the authors declare that they have not conflict of interest.

References

Bačnik K, Kutnjak D, Pecman A, Mehle N, Žnidarič MT, Aguirre IG, Ravnikar M (2020) Viromics and infectivity analysis reveal the release of infective plant viruses from wastewater into the environment. Water Research 177:115628

Bonanno Ferraro G, Suffredini E, Mancini P, Veneri C, Iaconelli M, Bonadonna L, Montagna MT, De Giglio O, La Rosa G (2021) Pepper mild mottle virus as indicator of pollution: Assessment of prevalence and concentration in different water environments in Italy. Food Environ Virol 13:117–125

Bonanno Ferraro G, Veneri C, Mancini P, Iaconelli M, Suffredini E, Bonadonna L, Lucentini L, Bowo-Ngandji A, Kengne-Nde C, Mbaga DS, Mahamat G, Tazokong HR, Ebogo-Belobo JT, Njouom R, Kenmoe S, La Rosa G (2022) A state-of-the-art scoping review on SARS-CoV-2 in sewage focusing on the potential of wastewater surveillance for the monitoring of the COVID-19 pandemic. Food Environ Virol 14:315–354. https://doi.org/10.1007/s12560-021-09498-6

Brownell MJ, Harwood VJ, Kurz RC, McQuaig SM, Lukasik J, Scott TM (2007) Confirmation of putative stormwater impact on water quality at a Florida beach by microbial source tracking methods and structure of indicator organism populations. Water Research 41:3747–3757

Bushnell B, Rood J, Singer E (2017) BBMerge – Accurate paired shotgun read merging via overlap. PLoS ONE 12:e0185056

Cantalupo PG, Calgua B, Zhao G, Hundesa A, Wier AD, Katz JP, Grabe M, Hendrix RW, Girones R, Wang D, Pipas JM (2011) Raw sewage harbors diverse viral populations. Mbio 2:e00180-e211

Da Silva LA, Oliveira AS, Melo FL, Ardisson-Araújo DMP, Resende FV, Resende RO, Ribeiro BM (2019) A new virus found in garlic virus complex is a member of possible novel genus of the family Betaflexiviridae (order Tymovirales) PeerJ 7:e6285

Favara GM, Camelo-García VM, Nagata T, Silva JMF, Saito M, Rezende JAM, Salaroli RB, Kitajima EW (2019) Tobacco mild green mosaic virus found naturally infecting Nicotiana glauca in Brazil. Australasian Plant Disease Notes 14:1–2

Fernandez-Cassi X, Timoneda N, Martínez-Puchol S, Rusiñol M, Rodriguez-Manzano J, Figuerola N, Bofill-Mas S, Abril JF, Girones R (2018) Metagenomics for the study of viruses in urban sewage as a tool for public health surveillance. Science of the Total Environment 618:870–880

Gyawali P, Croucher D, Ahmed W, Devane M, Hewitt J (2019) Evaluation of pepper mild mottle virus as an indicator of human faecal pollution in shellfish and growing waters. Water Research 154:370–376

Kitajima M, Sassi HP, Torrey JR (2018) Pepper mild mottle virus as a water quality indicator. NPJ Clean Water 1:1–9

Martínez-Puchol S, Rusiñol M, Fernández-Cassi X, Timoneda N, Itarte M, Andrés C, Antón A, Abril JF, Girones R, Bofill-Mas S (2020) Characterisation of the sewage virome: comparison of NGS tools and occurrence of significant pathogens. Science of the Total Environment 713:136604

Menzel P, Ng KL, Krogh A (2016) Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nature Communications 7:1–9

Moura MCF, Holanda ISA, Sales Júnior R, Queiroz APO, Araújo EOA, Oliveira GDC, Nunes GHS, Nagata T, Negreiros AMP (2018) First Report of melon necrotic spot virus in melon plantations in Brazil. Plant Dis 102:1048–1048

Nagai A, Duarte LML, Chaves ALR, Alexandre MAV, Ramos-González PL, Chabi-Jesus C, Harakava R, dos Santos DYAC (2018) First complete genome sequence of an isolate of tomato mottle mosaic virus infecting plants of *Solanum lycopersicum* in South America. Genome Announc 6:e00427–e518

Nakamura K, Iwai M, Zhang J, Obara M, Horimoto E, Hasegawa S, Kurata T, Takizawa T (2009) Detection of a novel recombinant Norovirus from sewage water in Toyama prefecture, Japan. Jpn J Infect Dis 62:394–398

Ng TFF, Marine R, Wang C, Simmonds P, Kapusinszky B, Bodhidatta L, Oderinde BS, Wommack KE, Delwart E (2012) High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. J Virol 86:12161–12175

Rosario K, Symonds EM, Sinigalliano C, Stewart J, Breitbart M (2009) Pepper mild mottle virus as an indicator of fecal pollution. Applied and Environmental Microbiology 75:7261–7267

Symonds EM, Nguyen KH, Harwood VJ, Breitbart M (2018) Pepper mild mottle virus: A plant pathogen with a greater purpose in (waste)water treatment development and public health management. Water Research 144:1–12

Tamura K, Stecher G, Kumar S (2021) MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol 38:3022–3027

Xagoraraki I, O'Brien E (2020) Wastewater-based epidemiology for early detection of viral outbreaks. In: O'Bannon DJ (ed) Women in water quality: Investigations by prominent female engineers. Springer International Publishing, Cham, pp 75–97. https://doi.org/10.1007/978-3-030-17819-2_5

Zhang T, Breitbart M, Lee WH, Run JQ, Wei CL, Soh SWL, Hibberd ML, Liu ET, Rohwer F, Ruan Y (2005) RNA Viral community in human feces: Prevalence of plant pathogenic viruses. PLoS Biol 4:e3