Metagenomic analysis of virome cross-talk between cultivated Solanum lycopersicum and wild Solanum nigrum

Yuxin Ma, Armelle Marais, Marie Lefebvre, Chantal Faure, Thierry T. Candresse

Metagenomic analysis of virome cross-talk between cultivated *Solanum lycopersicum* and wild *Solanum nigrum*

Yuxin Ma, Armelle Marais, Marie Lefebvre, Chantal Faure, and Thierry Candresse*

UMR 1332 BFP, INRA, Univ. Bordeaux, CS20032, 33882 Villenave d'Ornon cedex, France

Abstract: 148 word

Text: 5029 words (Introduction- Materials and Methods-Results-Discussion-Acknowledgements)

* Corresponding author thierry.candresse@inra.fr

The assembled viral genomic reported here have been deposited in Genbank under accession numbers **MN216346** to **MN216389**. Cleaned virome sequence reads have been deposited on the INRA National Data Portal under the identifier https://doi.org/10.15454/S486RR.

ABSTRACT

Wild plants and weeds growing close to crops constitute a potential reservoir for future epidemies or for the emergence of novel viruses but the frequency and directionality of viral flow between cultivated and wild plants remains poorly documented in many cases. Here, we studied the diversity of viral populations between tomato (*Solanum lycopersicum*) and neighboring european black nightshade (*Solanum nigrum*) using high throughput sequencing (HTS) based metagenomics. A large variability in virome richness with only 17.9% shared Operational Taxonomy Units between tomato and nightshade, but this richness could not be linked to a particular host or to local conditions. A detailed population analysis based on assembled contigs for potato virus Y (PVY), broad wilt bean virus 1 and a new ilarvirus tentatively named Solanum nigrum ilarvirus 1 provides information on the circulation of these viruses between these two *Solanum* species and enriches our knowledge of the tomato virome.

KEYWORDS: metagenomics, virome, double stranded RNA (dsRNA), tomato, spillover

INTRODUCTION

Through the past decade, metagenomics based on high throughput sequencing (HTS) has been widely used in the plant virology field, advancing our knowledge on the diversity of plant viruses. Specifically, metagenomics allowed to discover unknown viruses, explore the intraspecific genetic diversity of known viruses, and study virus ecology and epidemiology (Massart et al., 2014; Roossinck et al., 2015; Villamor et al., 2019). Plant viruses cause epidemics on all major cultures of agronomic importance, representing a serious threat to global food security. As a consequence, virologists have for a long time focused their efforts on economically important crops, often neglecting bordering weeds and wild plants (Wren et al., 2006). However, agro-ecosystems are complex environments in which crop plants sometimes interact with the in-plot and bordering weeds and wild plants. Viruses may be transferred between wild plants/weeds and crops and vice versa by a variety of mechanisms and vectors. Thus wild plants or weeds may constitute “reservoirs” of viruses that may subsequently spread to cultivated plants while crops may constitute a source from which viral infections may spillover to the wild plants/weeds compartment (Power and Mitchell, 2004).

Overall, our understanding of the details of fluxes of viruses from crops to weeds and from weeds to crops remains frequently limited. A role of a weed population as a reservoir or, alternatively, virus spillover from a crop are often assumed. Unfortunately, the techniques most frequently used for viral populations characterization, in particular serological ones, frequently do not provide sufficient intra-specific resolution to provide unambiguousl conclusions on such aspects. It is then difficult to ensure that the co-occurrence of a virus in

crops and weeds reflects the transfer of isolates rather than the existence separate viral populations adapted to the two host populations.

Tomato (*Solanum lycopersicum*) is one of the most popular and extensively consumed vegetable crops. There are at least 136 characterized viral species that are capable of infecting tomato and due to global climate changes and increased international trade, the spread of known viruses to new geographic areas and the emergence of new viruses have been frequently detected in particular in recent years (Brunt, 1996; Hanssen et al., 2010). Torradoviruses (family *Secoviridae*) are an example of a group of recently emerged plant viruses, many of which affect tomato. These include for example tomato torrado virus (ToTV), which was first described from tomato in Mexico (Verbeek et al., 2008) and reported more recently in France (Verdin et al., 2009) and in other host plant species (van der Vlugt et al., 2015), as well as tomato marchitez virus (ToMarV; (Verbeek et al., 2008)) and tomato chocolàte virus (ToChV; (Verbeek et al., 2010)). Another example of recent emergence of a virus in tomato concerns tomato brown rugose fruit virus (ToBRFV), a tobamo like virus which was discovered from tomato in Israel in 2014 (Luria et al., 2017) and that has spread since then to many countries including Jordan, Mexico, the United States (Southern California), Germany, Italy, Turkey, the Netherlands and Saudi Arabia. The source(s) and cause(s) of the emergence of such novel agents is(are) frequently unknown but weed and wild plants are often considered as a major sources of future emerging viruses than may occasionally be transferred to crops (Anderson et al., 2004; Elena et al., 2014; McLeish et al., 2019).

Recently, during a study characterizing the virome of 170 field-grown tomatoes collected in China by small RNAs sequencing, Xu et al. (2017) showed that the tomato viral community is dominated by a few species, most of them being positive-sense ssRNA viruses. Multiple infections were found to be frequent as well as recombination events in viral genomes (Xu et al., 2017).

European black nightshade (*Solanum nigrum*), a wild species in the same botanical genus as tomato is a widespread weed in many countries. However, in southern India it is widely consumed and cultivated on a commercial scale (Jamuna et al., 2017) and sometimes also used as for its medicinal properties (Javed et al., 2011). *S. nigrum* is known to harbor a wide range of viruses such as begomoviruses, orthotospoviruses, potyviruses, tobraviruses under field conditions, and has often been suspected to act as a reservoir host for viruses or for recombinant isolates infecting solanaceous crops (Holm et al., Garcia-Andrès et al., 2006; 1979; Jamuna et al., 2017).

In the present study, using a metagenomics approach, we investigated and compared the virome in tomato samples and in the related *S. nigrum* populations collected either in tomato fields or in various other environments. The comparison of these viromes provides novel insight into the viral fluxes between these two species.

MATERIALS AND METHODS

Study sites and plant samples

Virome richness and composition were analyzed in tomato (*Solanum lycopersicum*) and in European black nightshade (*Solanum nigrum*) that were growing either close to the

sampled tomato crops or in other sites, unmanaged or involving unrelated crops (sunflower, maize, sorghum and alfalfa, Table S1). In total, tomato crops were sampled in seven sites and nightshade in six of the seven tomato sites, plus in five non-tomato sites (Table S1). For each sampled plant populations, leaves from a total of 100 individual plants were collected in summer 2017 or 2018 and assembled in two pools corresponding to fifty individual plants (0.1g of leaf/plant) for nucleic acids extraction. No specific efforts were made to select symptomatic plants, but plants with obvious fungal attack, insect colonization or necrotized parts were excluded.

Double-stranded RNAs purification, library preparation and Illumina HiSeq sequencing

Double-stranded RNAs were purified from each plant pool by two rounds of CF11 cellulose chromatography and converted to cDNA according to the protocol described by Marais et al. (2018). In parallel, a negative control blank was similarly prepared using only buffer. Whole genome amplifications (WGAs) were performed on each cDNA sample (using the same MID tag for the two pools of each sampling site), the PCR products were purified using the MinElute PCR Purification Kit (Qiagen) and their concentration determined spectrophotometrically (Marais et al., 2018). Equal quantities of the amplification products from the two pools of each sampling site were then regrouped and independent sequencing libraries prepared for each site and sequenced in multiplexed format (2×150 bp) on an Illumina HiSeq 3000 system at the GenoToul platform (INRA Toulouse, France). Cleaned virome sequence reads have been deposited on the INRA National Data Portal under the identifier https://doi.org/10.15454/S486RR.

Bioinformatics analyses: Reads cleaning, contigs assembly and annotation, Operational Taxonomic Units (OTU) clustering

Following demultiplexing, adapters and MID tags were removed with *cutadapt* (Martin, 2011), and reads were quality trimmed (minimum quality score 20, minimum length 70 nucleotides). In order to limit inter-sample cross talk associated with index-hopping (Illumina, 2017; van der Valk et al., 2019), only reads having identical MID tags on both pair members were retained for further analyses (Table S1). Contigs were *de novo* assembled for each library using IDBA-UD (https://academic.oup.com/bioinformatics/article/28/11/1420/266973).

All contigs were annotated using BlastN and BlastX against the NCBI Genbank non redundant nucleotide (nt) or protein databases with a conservative e-value cut-off of 10^{-4}. In this way, contigs were assigned to one of the following categories: virus, eukaryote, bacteria, algae, and unknown. For viral contigs, a family-level annotation was derived from the NCBI taxonomic information for the first Blast hit.

A clustering approach (Lefebvre et al., 2019) was used to define operational taxonomy units, following the strategy highlighted by Simmonds (2015). Briefly, a search of RNA-dependent RNA polymerase (RdRp) conserved protein motifs was performed in all contigs using Reversed Position Specific Blast (RPS-Blast) (Altschul et al., 1997) against the pfam database (Bateman et al., 2018). The contigs encoding a viral RdRp motif (Table S1) were retrieved and aligned with reference sequences and distance matrices computed with the ETE3 toolkit (Huerta-Cepas et al., 2016). These matrices were used to perform a clustering allowing to regroup in a single OTU all contigs differing by less than a set

123 cut-off divergence value (Murtagh and Legendre, 2014). We used a 10% divergence cut-off
124 value, because it has been shown to generate in many viral families OTUs that are a
125 relatively good approximation of taxonomic species (Lefebvre et al., 2019). OTUs were
126 thus defined for each RdRp family, allowing to generate an OTU table indicating for each
127 sampling site the presence/absence and the number of reads integrated in each identified
128 OTU (Table S1).

129 ***Further viral genome assemblies, sequence comparisons and phylogenetic analyses***

130 When needed, contigs were extended by repeated rounds of mapping of quality-trimmed
131 reads using CLC Genomics Workbench 11 (CLC-GWB). For some isolates/viruses,
132 genomic scaffolds were assembled by mapping contigs and/or reads on a reference genome
133 using CLC-GWB. Long contigs or scaffolds showing more than 75% completeness for
134 cucumber mosaic virus (CMV), southern tomato virus (STV), broad wilt bean virus 1
135 (BBWV1, both genomic RNAs), the new ilarvirus (all three genomic RNAs) and potato
136 virus Y (PVY) were used for phylogenetic analyses and have been deposited in Genbank
137 (Accession numbers MN216356 to MN216369 (Table S2).

138 Multiple sequence alignments of contigs/scaffolds obtained from HTS data and of
139 reference isolates retrieved from Genebank (or alignments of deduced encoded proteins)
140 were performed using the ClustalW algorithm (Thompson et al., 1994) as implemented in
141 MEGA 6.0 (Kumar et al., 2008). Phylogenetic trees were reconstructed in MEGA 6.0,
142 using the Maximum Likelyhood (ML) algorithm. Branch support was evaluated by
143 bootstrap analysis (100 replicates).

RESULTS

Comparison of the tomato and nightshade viromes at different sampling sites

A total of 20 viral families were discovered by Blast annotation taking into account the different libraries (18 sampled plant populations) with an average of 4.3±3.3 families per library, but with a very large variability between the sampled plant populations. The tomato sample from the TOM3 site showed the highest number of viral families (13, Figure 1) followed by another tomato sample (TOM7, 9 viral families) and nightshade samples from the TOM2 and NIG3 sites (8 viral families). The *Potyviridae* family was represent in a total of 13 samples including both tomato (six samples) and nightshade (seven samples, of which five were from tomato sites; Figure 1). The family *Totiviridae* was represented in eight samples while at the other extreme the *Tombusviridae* family was represented in a single tomato sample from the TOM3 site. Given the high between-populations variability it was not possible to establish statistically meaningful differences in family-level richness between the tomato and nightshade populations (Figure 1).

Taking into account all sampling sites, a total of 87 unique RNA-dependent RNA polymerase (RdRp) OTUs were detected (Table S1). Similar to the family-level analysis, a very large variability was observed in the number of OTUS detected per site. The richer viromes were found in the TOM7 site tomato population and in the NIG3 nightshade population, with respectively 38 and 27 OTUs, followed by 26 OTUs for the TOM3 site tomato population. In all other samples less than 8 RdRp OTUs were detected (Table S1).

In total, 62 OTUs were identified from tomato samples and 44 from nightshade ones but this difference is largely the consequence of a single tomato sample (TOM7) which is

166 particularly rich in unique mycovirus-like OTUs (Table S1). Nineteen OTUs (21.8% of
167 total) were shared between the two plant species, most of them from the families
168 *Totiviridae*, *Partitiviridae and Chrysoviridae* as well as unclassified mycovirus-like OTUs.
169 RdRP_1-OTU_8 which corresponds to potato virus Y was the most widely shared OTU
170 (Table S1, see below). It explains the wide prevalence of the *Potyviridae* family described
171 above. Twenty-five OTUs were found to be nightshade-specific, among which
172 RdRP_2-OTU_13 corresponds to cucumber mosaic virus (CMV) and RdRP_1-OTU_14 to
173 broad bean wilt virus 1 (BBWV1) (Table S1, see below). Fourty-three OTUs were found to
174 be tomato-specific, some of which have extremely high identity levels with known viruses
175 such as Sclerotinia sclerotiorum hypovirus 1, Sclerotinia sclerotiorum umbra-like virus 2 or
176 Botrytis virus F and very likely correspond to these agents (Table S1).
177 There were overall only very few OTUs shared between tomato and nightshade samples for
178 a given sampling site, with PVY being the most frequent. In five sites out of six, no OTU
179 (sites TOM2 and TOM6) or only one OTU (sites TOM1, TOM4, and TOM5) were shared,
180 whereas in site TOM3, four OTUs were shared (Table S1, Figure S1).

181 ***Near complete genome reconstruction for selected viral agents***

182 For several viruses, long, high quality contigs were obtained during the initial trimmed
183 reads assembly. This concerned in particular several single-stranded RNA viruses:
184 cucumber mosaic virus (CMV), broad wilt bean virus 1 (BBWV1, both genomic RNAs),
185 potato virus Y (PVY), and a new ilarvirus (all three genomic RNAs) as well as a
186 double-stranded RNA virus of the *Amalgaviridae* family, southern tomato virus (STV). In a
187 few cases, the viral genome was unambiguously covered by a few contigs that were either

188 non-overlapping or had only a short overlap and which were therefore manually assembled
189 into a scaffold by mapping contigs on a reference genome. All contigs and scaffolds were
190 validated by visual inspection of read mappings at high stringency to ensure the absence of
191 assembly artifacts. The corresponding sequences have been deposited in Genbank
192 (Accession numbers MN216346 to MN216389, Table S2).
193 Multiple alignments and phylogenetic analyses (see below) were used to identify
194 representative contigs for the various phylogenetic clusters of each virus. These
195 representative contigs were in turn used a targets for the mapping of the trimmed reads of
196 all libraries at high stringency. This allowed to evaluate the representation of each
197 virus/variant in the virome of each sampled plant population. The low background of viral
198 reads observed in the negative control, probably resulting from low level experimental
199 contamination or from inter sample cross talk due to index-hopping (Illumina, 2017), was
200 substracted from the mapped reads numbers of each library. The results of this analysis are
201 presented in Table S3.
202 CMV was detected, by high read numbers, at a single sampling site (TOM1), in the
203 nightshade population but not in the corresponding tomato population (Table S3). All three
204 genomic RNAs were assembled into unique long contigs of respectively 3,301 nt (RNA1;
205 ca. 98.2% of the full length molecule), 2,996 nt (RNA2, ca. 98.3% of the full length
206 molecule) and 2,155 nt (RNA3, ca. 97.2% of the full length molecule) but no evidence was
207 found for presence of a CMV satellite. Despite the fact that no specific efforts were made
208 to improve/validate the contigs further, all three genomic RNAs are extremely close to
209 CMV sequences present in Genbank and, in particular to the I17F isolate, a subgroup I

210 isolate characterized from tomato in France at the beginning of the 1980's (Jacquemond
211 and Lot, 1981). Nucleotide identity levels of respectively 99.5%, 99.4% and 99.5% for
212 genomic RNA1, 2 and 3 (respectively 18, 18 and 11 point mutations) are thus observed
213 between the 1981 IF17 isolate and the contigs from HTS data on a 2017 sample,
214 highlighting both the quality of the HTS assemblies and the relative stability of the CMV
215 population over more than 35 years.
216 In the case of southern tomato virus (STV), unique long contigs representing nearly
217 complete genomes were obtained from several plant populations, representing 92.6%-99.4%
218 of the full length genome. Coherent with the low diversity identified so far in this virus,
219 these contigs are nearly identical to each other (<0.4% nucleotide divergence) with the
220 exception of one contig, which diverges by 2.6%-2.8% from the others. Identity levels with
221 isolates present in Genbank range from 100% to 95.9%, again highlighting the quality of
222 the contigs assembled from the HTS data. Overall STV was detected in five of the seven
223 tomato pools, an observation in accordance with the presence of this virus in a wide range
224 of tomato varieties (Sabanadzovic et al., 2009). On the other hand, a surprising result is the
225 detection, with higher reads number than for the tomato pools (Table S3) of STV in a
226 nightshade pool (Nightshade-TOM5), extending the host range of this relatively recently
227 discovered virus. The nightshade STV sequence belongs to the group of closely related
228 isolates and does not present obvious specific molecular properties (data not shown).

229 ***Broad bean wilt virus 1 populations diversity***

230 Broad bean wilt virus 1 (BBWV1) was detected in six of the sampled nightshade
231 populations (out of a total of 11 populations, 55%) and was not detected in any of the

sampled tomato populations (seven populations, Table S3). The assembly of the BBWV1 reads from the various nightshade populations highlighted a complex viral population structure with a total of five RNA1 clusters and three RNA2 clusters identified (Table S3, see below). On average, the reconstructed genomic sequences represented 94.3% +/- 3.9% of the BBWV1 RNA1 (87.8%-99.3%, depending on the contigs) and 87.8% +/- 11.9% of the BBWV1 RNA2 (73.4%-96.6%). For one sample, it was not possible to reconstruct more than 60% of the RNA2 sequence and the corresponding scaffold was therefore not included in further analyses. The average nucleotide divergence between the RNA1 clusters, calculated on representative isolates is 16.6% +/- 0.3% (13.4%-17.6%), explaining the effective separate assembly in cases of mixed infection by isolates belonging to different clusters. For the three RNA2 clusters, the corresponding divergence values are 15.8% +/- 0.6% (10.5%-18.6%). Mapping of reads at high stringency on contigs representative of the various clusters allowed to describe the BBWV1 population present in the various plant populations. Isolates representative of between one and four RNA1 clusters and of one or two RNA2 clusters could thus be detected at individual sampling sites, with some sites providing evidence of only a single RNA1-RNA2 combination, while at the other extreme, one site showed the presence of four RNA1 clusters and a single RNA2 one. Another site showed the presence of a single RNA1 cluster but of two RNA2 ones (Table S3). Taken together, these elements suggest the frequent occurrence of reassortment between BBWV1 genomic segments in the sampled nightshade populations.

Phylogenetic analyses performed on the RNA1 and RNA2 sequences derived from the HTS data and from all full length isolates present in Genbank (Figure 2A and 2B)

254 demonstrate that the BBWV1 isolates present in the nightshade populations sampled here
255 largely expand the known BBWV1 diversity. Indeed, the HTS-derived sequences cluster
256 separately from reference full-length sequences available to date and are, on average,
257 highly divergent from them with an average intergroup distance of 17.0% +/- 0.4% for
258 RNA1 and 19.0% +/- 0.6% for RNA2.

259 ***Presence of a novel ilarvirus in the sampled nightshade and tomato populations***

260 Long, high quality contigs representative of an ilarvirus were identified in several libraries.
261 The contigs corresponding to the three genomic RNAs were further extended and validated
262 for the NIG4 sampling site, allowing to reconstruct near complete molecules. Indeed, a
263 comparison with the genomic RNAs of Parietaria mottle virus (PMoV), the closest
264 characterized ilarvirus (see below) indicated that all five open reading frames (ORFs)
265 [coding respectively for P1 (RNA1), P2 and P2b (RNA2) and the movement (MP) and coat
266 proteins (CP) (RNA3)] were complete, with the exception of ORF2 which misses an
267 estimated 62 nt (21 N-terminal amino acids missing from the P2 protein sequence). The
268 contigs are respectively 3,445, 2,757 and 2,257 nt long for RNA1, RNA2 and RNA3,
269 representing respectively 97.9%, 94.4% and 100.5% of the length of the corresponding
270 genomic RNAs of the reference PMoV isolate (NC_005848, -49 and -54). These genomic
271 sequences have been deposited in Genbank under Accession numbers MN216370 to
272 MN216378. Blast analyses indicated that the virus is most closely related to PMoV and to
273 several other subgroup 1 ilarviruses and this proximity was confirmed by phylogenetic
274 analyses performed on all genome encoded proteins (Figure 3A and 3B, Figure S2).
275 However, these phylogenetic trees demonstrate that the virus is not substantially more

related to PMoV than to any other approved species in that small ensemble. The significant divergence of the virus from PMoV is confirmed by sequence comparisons, the deduced proteins being only 81.8% (P1) to 53.9% (CP) identical with those of PMoV while the genomic RNAs show only 73.2% (RNA1) to 58.6% (RNA3) nucleotide identity (Table S4). Taken together these results suggest that the detected larvirus is a new subgroup 1 member for which the name Solanum nigrum ilarvirus 1 (SnIV1) is proposed.

Mapping of the reads from each plant population on the SnIV1 genomic RNAs showed that this virus was present in eight of them, corresponding to 6/11 nightshade populations (54.5%) and, represented by relatively low read numbers, to 2/7 tomato populations (28%) (Table S3).

Analysis of PVY populations in the sampled nightshade and tomato populations

As for the other viruses, long, high quality contigs were obtained in most cases for PVY. In a few cases, probably resulting from low reads numbers or from the simultaneous presence of closely related isolates in the sampled plant populations, only short PVY contigs were obtained for some isolates. However, from all plants populations with high PVY read numbers, one to three long contigs could be assembled presenting on average 95.4% +/- 4.8% of the full length PVY genome (85.3%-99.9%). A phylogenetic analysis of these contigs, together with representative reference isolates retrieved from Genbank (Figure 4) shows a very contrasted situation, with on the one hand, a large number of sequences forming a very tight cluster corresponding to PVY-NTN and, on the other, a much more diverse second cluster corresponding to PVY-C. No isolates representative of the PVY-O and PVY-N strains were observed nor some of their frequent recombinants such as PVY-Wilga

298 (Figure 4). In total, 10 contigs were obtained for PVY-NTN (five from tomato and five
299 from nightshade) and four for PVY-C (three from tomato and one from nightshade).
300 The reads from all plant populations were then mapped on selected contigs representative
301 of PVY-NTN and of the three PVY-C variants identified, using stringent parameters so as
302 to limit cross-mapping between isolates. Under these conditions, from one to four PVY
303 variants could be detected in the analyzed plant populations. Some populations showed
304 extremely low read numbers (<90), which is suggestive of an absence or a very low
305 prevalence of PVY in the corresponding plant populations. Remarkably, this situation
306 corresponds to 2/7 (28.6%) tomato populations, to 2/6 (33.3%) nightshade populations
307 growing side by side with tomato but to 4/5 (80%) of the nightshade population growing
308 away from tomato.
309 As judged from the mapping results, the two most frequent PVY strains were PVY-NTN,
310 which was detected in all tomato and nightshade populations in which PVY was detected,
311 and isolates with mapping affinities with isolate TOM7-C, which clusters together with the
312 French PVY-C1 SON41 pepper isolate (Table S3 and Figure 4). By contrast, isolates
313 corresponding to the two other PVY-C mapping references used were only detected in one
314 to three of the sampled plant populations. The frequency of detection of the various clades
315 does not seem to differ much between tomato and nightshade (Table S3).

316 **DISCUSSION**

317 The viromes characterized in the present work vary greatly between the sampled plant
318 populations and, for some of them, showed only a limited number of OTUs or of viral

319 families despite the size of the composite plant samples analyzed. This might reflect the
320 impact of fungicide treatments in the sampled crops which might have reduced fungal
321 diversity and in turn the ability to detect mycoviral communities associated with the
322 sampled plants. It should however be stressed that the OTU-based analysis provides a
323 lower bound estimate of viral diversity since viruses for which the genome region encoding
324 the conserved RdRp motif is not represented in the assembled contigs will not be identified
325 by a corresponding OTU. On the other hand, competition between the sequence of different
326 viruses for representation in the sequencing reads is unlikely to have adversely impacted
327 the richness of the identified viromes since the three richest viromes were identified in
328 plant populations for which the percentage of mapped viral reads was not obviously higher
329 (or lower) than that observed in samples with a much lower viral diversity (Tables S1 and
330 S2).
331 In contrast to a recent virome study of 170 tomato samples which indicated that diverse
332 ssRNA viruses represented 77% of the identified viruses (Xu et al., 2017), they represented
333 only 12.6% of the viral OTUs identified here (Table S1). The corresponding value for
334 dsRNA viruses is 26.4% while unassigned or unannotated agents accounted for a
335 cumulated 60.9%. Whether this difference is a consequence of differences in the
336 methodology used or actually reflects differences in the analyzed viromes cannot easily be
337 ascertained. However, some frequent viruses of tomato such as PVY, CMV or STV were
338 detected in both studies (Xu et al., 2017).
339 Despite the use of complex plant pools composed of 100 individual plants, we were able to
340 assemble long, high quality contigs for some viruses (PVY, BBWV1, STV, and the new

341 SnIV1), covering a very high proportion of the genome of these agents. In a few cases,
342 such long contigs could not be assembled, possibly as a consequence of too low coverage
343 and read numbers, or because mixed infection involving closely related variants created
344 problems during contig assembly. Indeed, there is some evidence that at least one
345 additional clade of PVY existed in some tomato samples as judged by the detection of
346 some partial contigs diverging from the fully assembled genomes (data not shown).
347 For read mapping, stringent parameters were used so that there is no or extremely limited
348 cross talk between isolates of different clades, as seen by reads numbers in the case of
349 BBWV1 and PVY (Table S2). At the same time, it is difficult to know precisely how to
350 interpret the samples with a very low number of reads mapped. Even if the background
351 observed in the negative control was substracted, this cross-talk background likely due to
352 index hoping (Illumina, 2017; van der Valk et al., 2019) may not be completely uniform
353 from sample to sample. These low read numbers may therefore either reflect an absence of
354 the virus but a low, slightly uneven cross-talk with other samples or a true, very low
355 prevalence of the virus in the sampled population. It is not possible to decide between these
356 two options here.
357 A very large and unexpected BBWV1 diversity was identified in the sampled nightshade
358 populations. The analysis of BBWV1 populations suggests the existence of frequent
359 reassortment between RNA1 and RNA2 variants, an observation in line with the results of
360 (Ferriol et al., 2014). BBWV1 is a Fabavirus with a relatively wide host range and which is
361 pathogenic on a range of crops including broad bean, pea, spinach, lettuce, pepper and,
362 occasionally, tomato (Blancard, 2012; Carpino et al., 2019; Taylor and Stubbs, 1972). It is

363 therefore surprising that this aphid-transmitted virus was only detected from nightshade
364 samples in this study. This observation suggest the existence of a biological or
365 epidemiological barrier limiting the spread of BBWV1 from nightshade to tomato. In this
366 respect, it is noteworthy that during a recent comparison of BBWV1 isolates, infection
367 rates in tomato following artificial inoculations ranged only from 40% to 60% for four
368 genetically different BBWV1 isolates (Carpino et al., 2019).
369 The novel ilarvirus here named Solanum nigrum ilarvirus 1 (SnIV1) was detected in both
370 tomato and nightshade samples. However, both the prevalence and, with one exception, the
371 read numbers of SnIV1 appear to be higher in the nightshade populations than in the
372 tomato ones. On the other hand, the presence of SnIV1 in nightshade samples does not
373 seem to be affected by whether they were growing side by side with tomato or not
374 (respectively 3/6 and 3/5 cases, Table S2). Interestingly, reanalysis of metagenomics data
375 showed that this virus was already present in 2011 at the TOM3 site,in *S. villosum* (hairy
376 nightshade) a close relative of *S. nigrum*. Whether this novel ilarvirus is pathogenic to
377 tomato or whether it has the potential to emerge at some point as a tomato pathogen in the
378 same fashion as its close relatives Parietaria mottle virus (Roggero et al., 2000) and tomato
379 necrotic spot virus (Batuman et al., 2011) remains to be evaluated.
380 The main PVY strains identified in this study were PVY-NTN and PVY-C1. PVY-C1
381 isolates were mainly detected from tomato, with one isolate shared between tomato and
382 nightshade in the TOM3 site (Figure 4). On the other hand, PVY-NTN isolates were found
383 in both tomato (5/7 samples) and nightshade samples (6/11 samples) from a total of seven
384 of the 12 sampling sites. Interestingly, PVY populations at the TOM3 site had been studied

385 2011-2012 using specific RT-PCR assays (Moury et al., 2017). At the time, PVY-C1 and
386 recombinant isolates likely to represent PVY-NTN were detected in tomato, while a more
387 diverse population involving PVY-O, PVY-NTN, PVY-N and PVY-C1 was detected in
388 nightshade and in the related *S. villosum* (Moury et al., 2017). The results reported here
389 therefore suggest a simplification of the PVY nightshade population at that site, with the
390 loss of PVY-O and PVY-N, possibly as a consequence of the competition with PVY-NTN
391 and C1.
392 A noteworthy observation concerns PVY prevalence in nightshade populations at tomato
393 sites (4/6 sites, 66.6%) and at non-tomato sites (1/5 sites, 20%). This suggests that infection
394 in nightshade is greatly increased by the presence of tomato, reflecting a likely spillover
395 effect from tomato crops to the wild nightshade population (Power and Mitchell, 2004).
396 Taken together the results reported here provide evidence for viral exchanges between
397 tomato and nightshade populations growing side by side (such as the extremely closely
398 related tomato and nightshade PVY isolates shared at the TOM3 site or the low detection of
399 the new ilarvirus in tomato only at sites were it is also present in nightshade). At the same
400 time, our results also highlight situations where an expected transfer is not observed, likely
401 as a consequence of unforeseen biological or ecological barriers. This concerns in
402 particular BBWV1 only found in nightshade when there are numerous indications that this
403 virus should be able to infect tomato (Carpino et al., 2019). These results also highlight the
404 power of metagenomics to analyze viral exchanges in complex plant populations, from the
405 overall virome structure down to the intra-specific variability level, revealing unknown
406 novel agents but also unforeseen biological processes.

ACKNOWLEDGMENTS

The authors would like to thank Eric Sclaunich (INVENIO, Sainte Livrade sur Lot, France) for assistance in accessing some of the tomato and nightshade sampling sites and the Genotoul Platform (INRA, Toulouse, France) for the Illumina sequencing. We would like to thank Laurence Svanella-Dumas and other colleagues in the INRA virology team for help in sample processing. Yuxin Ma was supported by a China Scholarship Council PhD grant.

414 **REFERENCES**

415 Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J.,
416 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search
417 programs. Nucleic Acids Res. 25, 3389-3402.

418 Anderson, P.K., Cunningham, A.A., Patel, N.G., Morales, F.J., Epstein, P.R., Daszak, P.,
419 2004. Emerging infectious diseases of plants: pathogen pollution, climate change and
420 agrotechnology drivers. Trends Ecol. Evol. 19, 535-544.

421 Bateman, A., Smart, A., Luciani, A., Salazar, G.A., Mistry, J., Richardson, L.J., Qureshi, M.,
422 El-Gebali, S., Potter, S.C., Finn, R.D., Eddy, S.R., Sonnhammer, E.L L., Piovesan, D.,
423 Paladin, L., Tosatto, S.C E., Hirsh, L., 2018. The Pfam protein families database in
424 2019. Nucleic Acids Res. 47, D427-D432.

425 Batuman, O., Chen, L., Gilbertson, R., 2011. Characterization of Tomato necrotic spot
426 virus (ToNSV), a new ilarvirus species infecting processing tomatoes in the Central
427 Valley of California, Phytopathology 103, 1391-1396.

428 Blancard, D., 2012. Tomato diseases: identification, biology and control: A Colour
429 Handbook. CRC Press, USA.

430 Brunt, A., 1996. Plant Viruses Online: Descriptions and Lists from the VIDE Database. Ver.
431 20. http://biology.anu.edu.au/Groups/MES/vide/.

432 Carpino, C., Elvira-González, L., Rubio, L., Peri, E., Davino, S., Galipienso, L., 2019. A
433 comparative study of viral infectivity, accumulation and symptoms induced by broad
434 bean wilt virus 1 isolates. J. Plant Pathol. 101, 275-281.

435 Elena, S.F., Fraile, A., Garcia-Arenal, F., 2014. Evolution and emergence of plant viruses.
436 Adv. Virus Res. 88, 161-191.

437 Ferriol, I., Ferrer, R.M., Luis-Arteaga, M., Guerri, J., Moreno, P., Rubio, L., 2014. Genetic
438 variability and evolution of broad bean wilt virus 1: role of recombination, selection
439 and gene flow. Arch. Virol. 159, 779-784.

440

García-Andrés, S., Monci, F., Navas-Castillo, J., Moriones, E., 2006. Begomovirus genetic diversity in the native plant reservoir Solanum nigrum: evidence for the presence of a new virus species of recombinant nature. Virology 350, 433-442.

Hanssen, I.M., Lapidot, M., Thomma, B.P., 2010. Emerging viral diseases of tomato crops. Mol. Plant. Microbe Interact. 23, 539-548.

Holm, L., Pancho, J.V., Herberger, J.P., Plucknett, D.L., 1979. A geographical atlas of world weeds. John Wiley and Sons, New York, USA.

Huerta-Cepas, J., Bork, P., Serra, F., 2016. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. Mol. Biol. Evol. 33, 1635-1638.

Illumina, 2017. Effects of index misassignment on multiplexing and downstream analysis. https://www.illumina.com/content/dam/illumina-marketing/documents/products/whitepapers/index-hopping-white-paper-770-2017-004.pdf.

Jacquemond, M., Lot, H., 1981. L'ARN satellite du virus de la mosaïque du concombre I. - Comparaison de l'aptitude à induire la nécrose de la tomate d'ARN satellites isolés de plusieurs souches du virus. Agronomie 1, 927-932.

Jamuna, S., Rajendran, L., Haokip, B.D., Nagendran, K., Karthikeyan, G., Manoranjitham, S.K., 2017. First Report of Natural Infection of Solanum nigrum with Tomato mosaic virus in India. Plant Dis. 102, 1044-1044.

Javed, T., Ashfaq, U.A., Riaz, S., Rehman, S., Riazuddin, S., 2011. In-vitro antiviral activity of Solanum nigrum against Hepatitis C Virus. Virol J. 8, 26-26.

Kumar, S., Nei, M., Dudley, J., Tamura, K., 2008. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. Briefings in Bioinformatics 9, 299-306.

465 Lefebvre, M., Theil, S., Ma, Y., Candresse, T., 2019. The VirAnnot pipeline: a resource for
466 automated viral diversity estimation and operational taxonomy units (OTU)
467 assignation for virome sequencing data. Phytobiomes Journal, online first,
468 DOI:10.1094/PBIOMES-07-19-0037-A.

469 Luria, N., Smith, E., Reingold, V., Bekelman, I., Lapidot, M., Levin, I., Elad, N., Tam, Y.,
470 Sela, N., Abu-Ras, A., 2017. A new Israeli Tobamovirus isolate infects tomato plants
471 harboring Tm-22 resistance genes. PLoS One 12, e0170429.

472 Marais, A., Faure, C., Bergey, B., Candresse, T., 2018. Viral Double-Stranded RNAs
473 (dsRNAs) from Plants: Alternative Nucleic Acid Substrates for High-Throughput
474 Sequencing. Methods Mol. Biol. 1746, 45-53.

475 Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing
476 reads. EMBnet J 17, 10-12.

477 Massart, S., Olmos, A., Jijakli, H., Candresse, T., 2014. Current impact and future
478 directions of high throughput sequencing in plant virus diagnostics. Virus Res. 188,
479 90-96.

480 McLeish, M.J., Fraile, A., García-Arenal, F., 2019. Evolution of plant–virus interactions:
481 host range and virus emergence. Curr. Opin. Virol. 34, 50-55.

482 Moury, B., Simon, V., Faure, C., Svanella-Dumas, L., Marais, A., Candresse, T., 2017. Host
483 groups of Potato virus Y: Vanishing barriers. In Potato virus Y: biodiversity,
484 pathogenicity, epidemiology and management. C. Lacomme, L. Glais, D.U. Bellstedt,
485 B. Dupuis, A.V. Karasev &, E. Jacquot, Eds. dir., *Potato virus Y: biodiversity,*
486 *pathogenicity, epidemiology and management* (p. 243-261). Springer, Cham. pp.
487 243-261.

488 Murtagh, F., Legendre, P., 2014. Ward's Hierarchical Agglomerative Clustering Method:
489 Which Algorithms Implement Ward's Criterion? J Classif 31, 274-295.

490 Power, Alison G., Mitchell, Charles E., 2004. Pathogen Spillover in Disease Epidemics.
491 The American Naturalist 164, S79-S89.

492

493 Roggero, P., Ciuffo, M., Katis, N., Alioto, D., Crescenzi, A., Parrella, G., Gallitelli, D.,
494 2000. Necrotic disease in tomatoes in Greece and southern Italy caused by the tomato
495 strain of Parietaria mottle virus. J. Plant Pathol. 82, 159.

496 Roossinck, M.J., Martin, D.P., Roumagnac, P., 2015. Plant Virus Metagenomics: Advances
497 in Virus Discovery. Phytopathology 105, 716-727.

498 Sabanadzovic, S., Valverde, R.A., Brown, J.K., Martin, R.R., Tzanetakis, I.E., 2009.
499 Southern tomato virus: The link between the families Totiviridae and Partitiviridae.
500 Virus Res. 140, 130-137.

501 Simmonds, P., 2015. Methods for virus classification and the challenge of incorporating
502 metagenomic sequence data. J Gen Virol 96, 1193-1206.

503 Taylor, R.H., Stubbs, L.L., 1972. Broad bean wilt virus 1.CMI-AAB Description of plant
504 viruses, 81, http://www.dpvweb.net/dpv/showadpv.php?dpvno=81

505 Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the
506 sensitivity of progressive multiple sequence alignment through sequence weighting,
507 position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22,
508 4673-4680.

509 van der Valk, T., Vezzi, F., Ormestad, M., Dalén, L., Guschanski, K., 2019. Index hopping
510 on the Illumina HiseqX platform and its consequences for ancient DNA studies. Mol.
511 Ecol. Resour. 2019; 00:1-11. https://doi.org/10.1111/1755-0998.13009.

512 van der Vlugt, R.A., Verbeek, M., Dullemans, A.M., Wintermantel, W.M., Cuellar, W.J.,
513 Fox, A., Thompson, J.R., 2015. Torradoviruses. Annu. Rev. Phytopathol. 53,
514 485-512.

515 Verbeek, M., Dullemans, A., van den Heuvel, H., Maris, P., van der Vlugt, R., 2010.
516 Tomato chocolate virus: a new plant virus infecting tomato and a proposed member
517 of the genus Torradovirus. Arch. Virol. 155, 751-755.

518 Verbeek, M., Dullemans, A.M., van den Heuvel, J.F., Maris, P.C., van der Vlugt, R.A.,

519 2008. Tomato marchitez virus, a new plant picorna-like virus from tomato related to
520 tomato torrado virus. Arch. Virol. 153, 127-134.
521

522 Verdin, E., Gognalons, P., Wipf-Scheibel, C., Bornard, I., Ridray, G., Schoen, L., Lecoq, H.,
523 2009. First report of Tomato torrado virus in tomato crops in France. Plant Dis. 93,
524 1352-1352.

525 Villamor, D.E.V., Ho, T., Al Rwahnih, M., Martin, R.R., Tzanetakis, I.E., 2019. High
526 Throughput Sequencing For Plant Virus Detection and Discovery. Phytopathology
527 109, 716-725.

528 Wren, J.D., Roossinck, M.J., Nelson, R.S., Scheets, K., Palmer, M.W., Melcher, U., 2006.
529 Plant Virus Biodiversity and Ecology. PLoS Biol. 4, e80.

530 Xu, C., Sun, X., Taylor, A., Jiao, C., Xu, Y., Cai, X., Wang, X., Ge, C., Pan, G., Wang, Q.,
531 Fei, Z., Wang, Q., 2017. Diversity, Distribution, and Evolution of Tomato Viruses in
532 China Uncovered by Small RNA Sequencing. J. Virol. 91, e00173-00117.

LEGENDS TO THE FIGURES

Figure 1. Barplot illustrating the presence/absence data based on Blast annotation for identified viral families in each sampled plant population.

Figure 2. Maximum Likelyhood trees reconstructed from the alignment of near complete nucleotide sequences of RNA1 (A) and RNA2 (B) of broad bean wilt virus 1 (BBWV1) isolates and other *Fabavirus* members. Statistical significance of the branches was evaluated by bootstrap analysis (100 replicates) and only bootstrap values higher than 70% are indicated. The scale bars represent 0.1 substitutions per site. Sequences of BBWV1 determined in this work are indicated by a black diamond. The abbreviations followed by the accession numbers are: BBWV2: broad bean wilt virus 2; GeMV: gentian mosaic virus; LLMV: Lamium mild mosaic virus; PeLaV: peach leaf pitting-associated virus; PrVF: Prunus virus F; ChVF: cherry virus F; GFabV: grapevine fabavirus.

Figure 3. Maximum Likelyhood trees reconstructed from the alignment of amino acid sequences of the P1 protein (A) and coat protein (B) of representative members of the genus *Ilarvirus*. Statistical significance of branches was evaluated by bootstrap analysis (100 replicates) and only bootstrap values higher than 70% are indicated. The scale bars represent 0.2 substitutions per site. Solanum nigrum ilarvirus 1 (SnIV1) characterized in this study is indicated by a black diamond.

Figure 4. Maximum Likelyhood tree of the near complete nucleotide genome sequences of potato virus Y (PVY) isolates determined in this study (indicated by diamonds) and reference sequences. PVY isolates from tomato samples are indicated by red diamonds and those from nightshade samples by blue diamonds. The tree was constructed by the Maximum Likelyhood method and the statistical significance of branches was evaluated by bootstrap analysis (100 replicates). Only bootstrap values higher than 70% are indicated. The scale bar represents 0.05 substitutions per site.

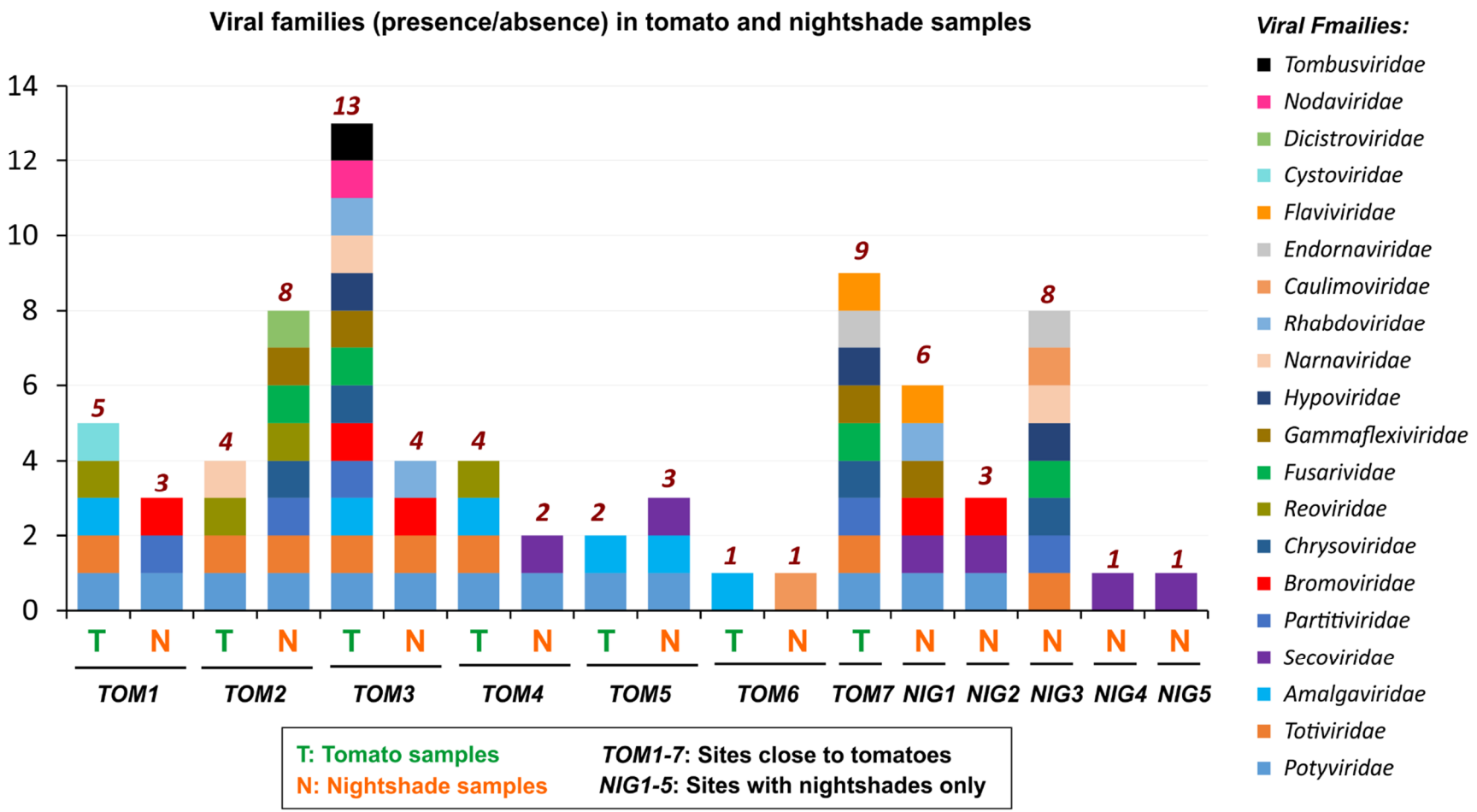
558 **Legends to Supplementary Figures**

559 **Figure S1. Between samples virome cross-talk at OTU level.** The sample/library and
560 identified number of OTUs are indicated at bottom-left; the interactions between different
561 viromes were shown in the matrix layout at the bottom-right, the aggregates based on the
562 groupings and the corresponding numbers of OTUs were plotted and shown in the upper
563 part.

564 **Figure S2. Maximum Likelyhood trees reconstructed from the alignment of amino**
565 **acid sequences of the P2a, P2b and movement (MP) proteins of representative**
566 **members of the genus *Ilarvirus*.** Statistical significance of branches was evaluated by
567 bootstrap analysis (100 replicates). Only bootstrap values higher than 70% are indicated.
568 The scale bars represent 0.2 substitutions per site. Solanum nigrum ilarvirus 1 (SnIV1)
569 characterized in this study is indicated by a black diamond.

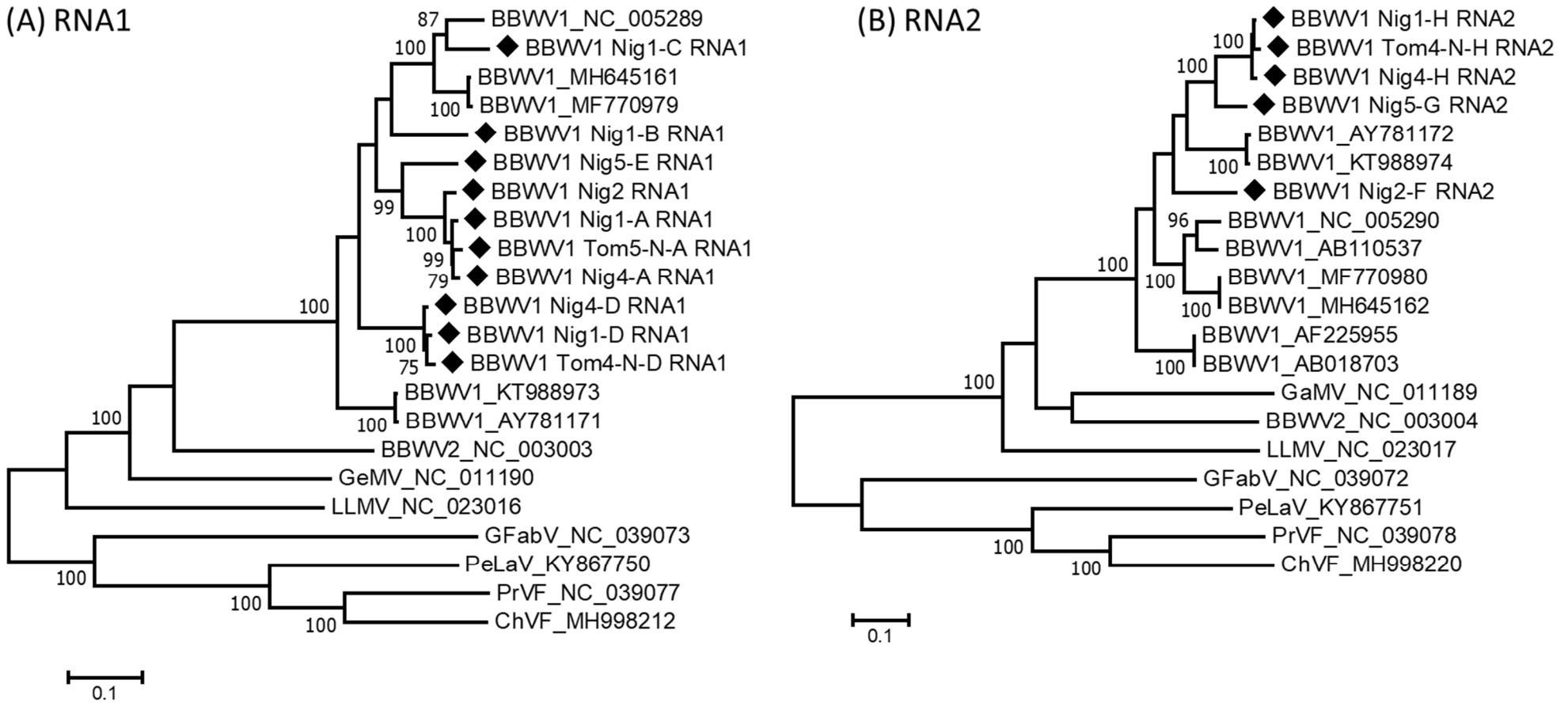
570

571 **Figure 1**



572
573

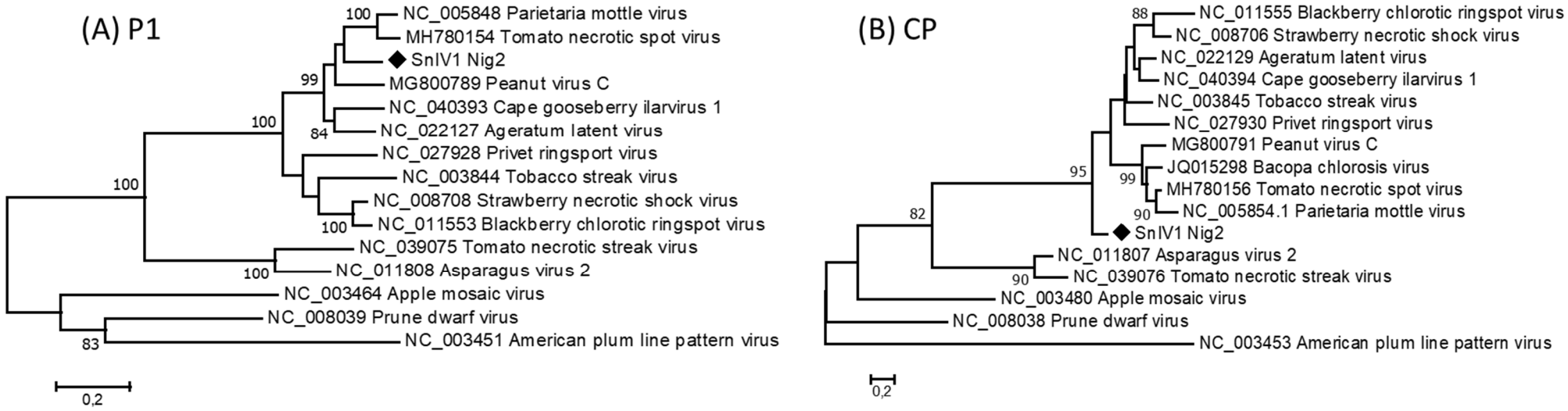
574 **Figure 2.**



575

576

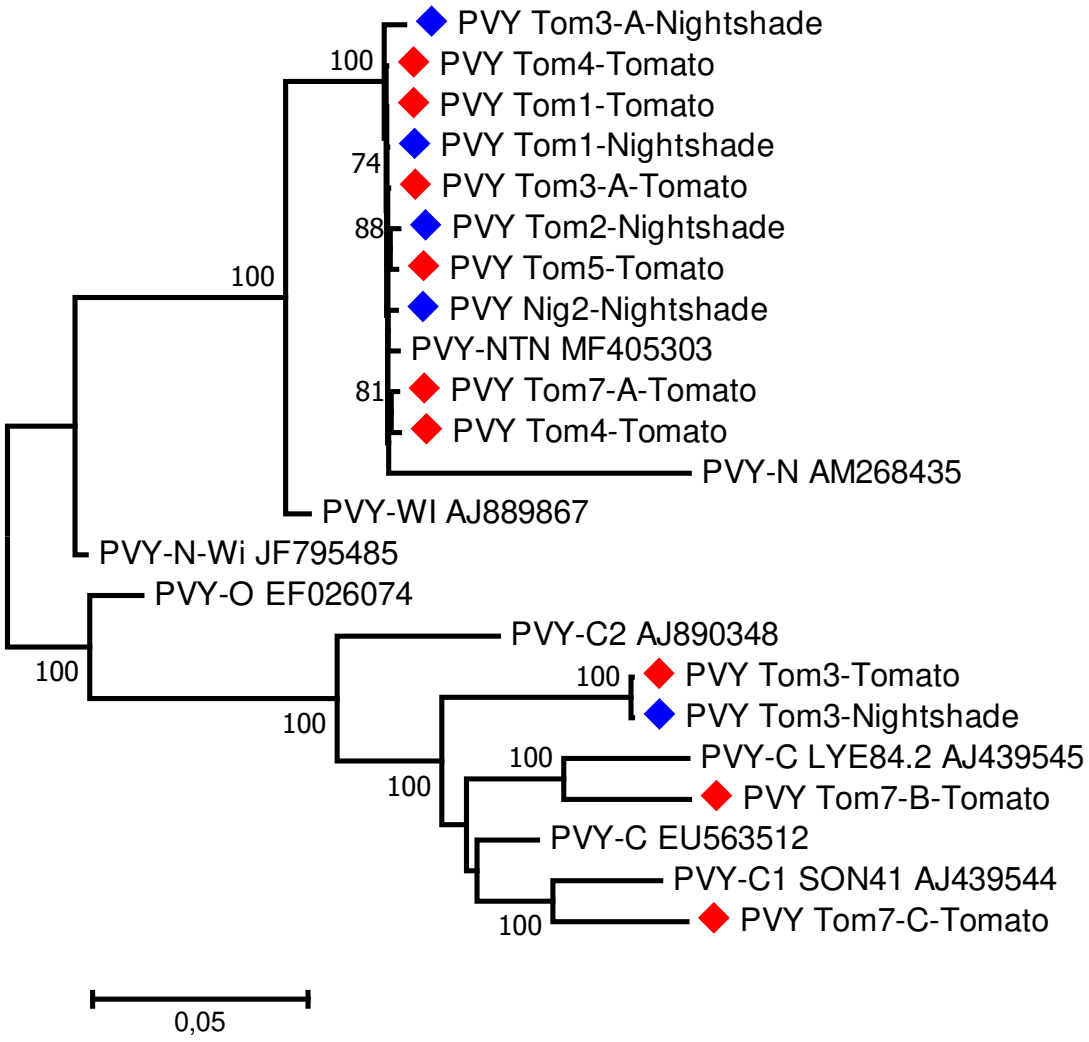
577 **Figure 3.**



578

579

580 **Figure 4.**



581

582