Diversity of polerovirus-associated RNAs in the virome of wild and cultivated carrots

Deborah Schönegger, Bisola Mercy Babalola, Armelle Marais, Chantal Faure, Thierry Candresse

▶ To cite this version:

Deborah Schönegger, Bisola Mercy Babalola, Armelle Marais, Chantal Faure, Thierry Candresse. Diversity of polerovirus-associated RNAs in the virome of wild and cultivated carrots. Plant Pathology, 2022, 71 (9), pp.1892-1900. 10.1111/ppa.13623 . hal-03756847

HAL Id: hal-03756847

https://hal.inrae.fr/hal-03756847

Submitted on 22 Aug 2022

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

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

Diversity of polerovirus-associated RNAs in the virome of wild and cultivated carrots

D. Schönegger[1a], **B.M. Babalola**[2], **A. Marais**[1b], **C. Faure**[1] and **T. Candresse** [1*]

[1]Univ. Bordeaux, INRAE, UMR BFP, CS 20032, 33882 Villenave d'Ornon Cedex, France

[2]Centro de Biotecnología y Genómica de Plantas, (UPM-INIA) and E.T.S.I.A.A.B., Universidad Politécnica de Madrid, Campus de Montegancedo, UPM, 28223, Pozuelo de Alarcón, Madrid, Spain

*Corresponding author: thierry.candresse@inrae.fr ORCID 0000-0001-9757-1835

a: ORCID 0000-0002-7419-5563

b: ORCID 0000-0003-2482-1543

Abstract

The self-replicating polerovirus-associated RNAs have recently been classified as tombusvirus-like associated RNAs (tlaRNAs). In a metagenomic comparison of the virome of wild (*Daucus carota* ssp. *carota*) and cultivated (*D. carota* ssp. *sativus*) carrots, four tlaRNAs were identified: carrot red leaf virus-associated RNA (CtRLVaRNA), beet western yellows virus-associated RNA (BWYVaRNA), and two other tlaRNAs detected for the first time in carrots, arracacha latent virus E-associated RNA (ALVEaRNA) and a new tlaRNA tentatively named carrot red leaf virus-associated RNA 2 (CtRLVaRNA-2). Their genomic sequences show a typical tlaRNA genome organization with a very short 5’ UTR, a long 3' UTR and consecutive ORFs, ORF1a and ORF1b separated by an amber stop codon, leading to the production of ORF1a/ORF1b RdRp fusion protein. The genetic diversity and prevalence of these various agents were analysed in carrot populations in France and Spain. CtRLVaRNA was the most frequently detected in both wild and cultivated carrot populations and was genetically more heterogeneous than ALVEaRNA and CtRLVaRNA-2. ALVEaRNA was detected both in France and Spain and showed preferential association with cultivated carrots. In contrast, CtRLVaRNA-2 was only detected from one region of France and was preferentially associated with wild carrot populations. Analysis of the virome of individual plants showed the presence of CtRLV but not that of the expected helper virus for ALVEaRNA and BWYVaRNA suggesting a degree of flexibility in their associations with helper viruses, which in turn may have biological and epidemiological consequences.

Key words: associated RNA, satellite, high-throughput sequencing, *Polerovirus*, *Daucus carota*

24 **INTRODUCTION**

25 An important group of plant pathogens are satellites, which are subviral agents known to
26 interact with co-infecting viruses and the host cellular machinery (Gnanasekaran &
27 Chakraborty, 2018). A common feature of satellites is their dependence on a helper-virus to
28 complete a full transmission cycle (Gnanasekaran and Chakraborty, 2018; Badar et al., 2021).
29 Satellites are currently divided into satellite viruses, encoding their own capsid protein but
30 relying on the helper virus for replication, and satellite nucleic acids, which vary in their
31 dependence on the helper virus for either encapsidation, cell-to-cell movement and/or
32 replication (Badar et al., 2021).

33 A class of single stranded satellite nucleic acids is often found in association with poleroviruses
34 and displays an autonomous replication capability (Campbell et al., 2020). However, these
35 satellites depend on co-infection partners of the family *Solemoviridae* (genus *Polerovirus* or
36 *Enamovirus*) for encapsidation and systemic movement (Badar et al., 2021; Campbell et al.,
37 2020). Recently, these *Polerovirus*-associated RNAs have been classified as tombusvirus-like
38 associated RNAs (tlaRNA) due to their phylogenetic affinities to viruses of the *Tombusviridae*
39 family (Campbell et al., 2020). Satellite tlaRNAs are single-stranded, positive-sense RNAs of
40 about 3 kb harboring two open reading frames (ORFs), ORF1a and ORF1b. ORF1a encodes a
41 protein of about 30 kDa and ORF1b is expressed by the readthrough of ORF1a termination
42 codon, resulting in a fusion protein encoding the satellite RNA-dependent RNA polymerase
43 (RdRp) (Campbell et al., 2020; Peng et al., 2021). There are few studies addressing the
44 specificity of the association between satellites and their helper virus (Oncino, et al., 1995;
45 Kurath et al. 1993). This may be due to the fact that such disease complexes have been largely
46 analysed under controlled conditions, focusing on a specific virus-satellite association in a
47 single host species (Peng et al., 2021; Mo et al., 2011), leading to a consequential knowledge
48 gap on such complexes in more natural infection conditions. The fact that isolates of beet

49 western yellows virus-associated RNA (BWYVaRNA) and Turnip yellows virus-associated
50 RNA (TuYVaRNA) show over 93% identity over their entire genome and the recent report of
51 the partial sequence of a potato leafroll virus-associated RNA (MF133518, Campbell et al.,
52 2020) showing 99.5% nucleotide identity with carrot red leaf virus-associated RNA
53 (CtRLVaRNA) suggests however some level of promiscuous associations involving tlaRNAs.

54 An agriculturally important member of tlaRNAs is CtRLVaRNA, which is involved in the
55 carrot motley dwarf disease (CMD) complex, causing severe damage to commercial carrot
56 crops (Adams, et al., 2014). CtRLVaRNA has a worldwide distribution, being reported from
57 the US (Watson., 1998; Campbell et al., 2020), Japan (Yoshida, 2020), New Zealand (Tang, et
58 al.,2009) and the UK (Adams, et al. 2014). Other viruses involved in the CMD complex are
59 carrot red leaf virus (CtRLV, *Polerovirus*) and carrot mottle virus (CMoV) and/or carrot mottle
60 mimic virus (CMoMV), both of which belong to the genus *Umbravirus*. Umbraviruses (family
61 *Tombusviridae*) are well studied for their dependence for encapsidation/aphid transmission on
62 helper viruses of the family *Solemoviridae* (previously *Luteoviridae*). In addition to their RdRp,
63 umbraviruses encode two additional ORFs, ORF3 and ORF4, involved in cell-to cell and long-
64 distance movement, respectively. In the CMD complex, the capsid protein of CtRLV (helper
65 virus) is used by the coinfecting umbraviruses and satellites (dependent viruses) for
66 transcapsidation of their genome, thus allowing their transmission by the aphid vectors of
67 CtRLV. Satellites generally do not share any sequence similarity with their helper virus but
68 often modulate (exacerbate or attenuate) helper-virus induced symptoms, as well as helper virus
69 accumulation in infected hosts (Gnanasekaran, P. & Chakraborty, S., 2018; Badar, et al., 2021).
70 The accumulation of CtRLV has, for example, been shown to be reduced in plants coinfected
71 by CtRLVaRNA and CMoV (Yoshida, 2020). In contrast, BWYVaRNA, first described as an
72 autonomously replicating subviral RNA (Chin et al., 1993), has been shown to stimulate the
73 accumulation of its helper virus, beet western yellows virus (BWYV, genus *Polerovirus*).

74 Similarly, the recently described pod pepper vein yellows virus-associated RNA has been
75 shown to stimulate the accumulation of its helper virus, leading to increased virulence (Peng et
76 al., 2021). Other examples of such tlaRNAs are tobacco bushy top disease-associated RNA
77 (TBTDaRNA), which is involved in the tobacco bushy top disease complex (Mo et al., 2011)
78 or arracacha latent virus E-associated RNA (ALVEaRNA) which was recently identified in the
79 virome of *Arracacia xanthorrhiza*, characterized by high throughput sequencing (partial
80 genomic sequences MF136436 and MF073199).

81 In a comparative metagenomic study aiming to understand specific virome fingerprints in
82 cultivated and wild relative species, the viromes of different cultivated (*Daucus carota* ssp.
83 *sativus*) and wild carrot (*D. carota* ssp. *carota*) populations were investigated. Four tlaRNAs
84 were thus identified, two of which were detected for the first time in carrots, including a novel
85 tlaRNA distantly related to CtRLVaRNA. The various tlaRNAs showed differences in their
86 distribution and prevalence and, for some of them, preferential association with wild or
87 cultivated carrot populations.

88 **MATERIALS AND METHODS**

89 ***Plant samples collection and preparation***

90 In summer 2019, 16 different populations of cultivated carrots (*Daucus carota* ssp. *sativus*) and
91 wild carrots (*Daucus carota* ssp. *carota*) were sampled locally in the Nouvelle-Aquitaine region
92 of France, including four cultivated populations, four off-type populations growing within the
93 fields (their early bolting phenology and root morphology suggest that they originate from
94 pollen or seed contaminations of the planted commercial varieties and they are therefore
95 referred to here as off-type) and eight wild carrot populations. In summer 2020, a similar
96 sampling was conducted in multiple areas of France, covering a north-south country gradient.
97 In total, 29 carrot populations were collected, comprising 11 cultivated carrots populations,

98 three off-type populations and 15 wild populations. A total of 45 carrot populations were
99 therefore sampled in France over two years. A similar sampling was performed in early summer
100 2021 near Segovia (Central Spain), including five cultivated and one wild carrot populations.
101 The characteristics of all sampled carrot populations are provided in Supplementary Table S1.
102 Plants were sampled independently of the presence of viral symptoms but plants with obvious
103 fungal infestation or necrosis were excluded. Leaf samples were stored desiccated of anhydrous
104 $CaCl_2$ until used. For each population, a pool corresponding to 50 plants was assembled (about
105 0.1g of leaf/plant).

106 ***Double-stranded RNAs purification, Illumina sequencing and virome assembly and***
107 ***annotation***

108 Double-stranded (ds) RNAs were purified from each pool of carrot leaves. In the case of two
109 French carrot populations (one wild and one cultivated), dsRNAs were also extracted from
110 leaves of the 50 individual plants (75 mg per plant). Double-stranded RNAs were purified by
111 two rounds of CF41 cellulose chromatography and nuclease treatment as described (Marais et
112 al., 2018). A negative extraction control was included in each extraction by using only buffer.
113 Purified dsRNAs were converted to cDNA and random amplified while simultaneously adding
114 MID tags (François et al., 2018; Marais et al., 2018). PCR products were purified using the
115 MinElute PCR Purification Kit (Qiagen SAS France, Courtaboeuf, France) and their
116 concentration determined spectrophotometrically (Marais et al., 2018). Amplification products
117 were sequenced in multiplexed format (2×150 bp) on an Illumina NovaSeq 6000 system at the
118 GetPlaGe platform (GenoToul INRAE Toulouse, France).

119 Sequencing reads were imported into CLC Genomics Workbench, version 21.0.3,
120 demultiplexed and trimmed on quality and length using default setting and a minimum read
121 length of 60 nucleotides (nt) and subsequently assembled *de novo* with the following
122 parameters: word size: 50, bubble size: 300, minimal contig length: 250 nt. Viral contigs were

123 identified by a comparison against the Virus REFSEQ GenBank database using BlastX
124 (Altschul et al., 1990). Contigs or scaffold thus identified for tlaRNAs were extended by rounds
125 of mapping of remaining reads. To confirm the HTS sequencing results, specific primers were
126 designed for each of the identified tlaRNAs (Supplementary Table S2) and pools or individual
127 plant samples were tested by two-step RT-PCR (Marais et al., 2011). If necessary, the 5' and
128 3' terminal sequences of individual tlaRNAs were confirmed using specific primers
129 (Supplementary Table S2) and the SMARTer RACE Kit (Takara Bio Europe SAS, Saint-
130 Germain-en-Laye, France).

131 ***Analysis of tlaRNAs genomic sequences and Phylogenetic analyses***

132 Multiple alignments of the fully reconstructed genomes of the known and new tlaRNAs
133 identified in the carrot virome and of tlaRNAs sequences retrieved from GenBank were
134 prepared using ClustalW as implemented in Mega 7 (Kumar et al., 2018). ORFs were identified
135 in tlaRNAs sequences with CLC GW and compared with sequences in GenBank using the
136 BlastP search tool to verify ORF predictions. Recombination events in tlaRNAs were analysed
137 based on a multiple alignment of complete genomic sequences and using the RDP4 package
138 (Martin et al., 2015). Only recombination events detected by at least four out of seven
139 implemented algorithms were considered. Neighbour-joining trees were inferred from
140 alignments of whole genomes or of the RdRp central conserved domain and strict nucleotide
141 (nt) or amino acid (aa) identities calculated using MEGA 7. The significance of branches was
142 estimated with 1000 bootstrap replicates.

143 **RESULTS**

144 ***Identification of known and novel tlaRNAs in the virome of wild and cultivated carrot***
145 ***populations***

146 Following *de novo* assembly of cleaned reads from plant pools or from individually sequenced
147 plants of wild and cultivated carrots, a large number of viral contigs were identified via BlastX-
148 based contigs annotation. Among them, multiple contigs showed significant homology to
149 known tlaRNAs. Preliminary analyses revealed that the identified tlaRNAs formed four distinct
150 groups corresponding to isolates closely related to CtRLVaRNA, BWYVaRNA and the partial
151 reported sequence for ALVEaRNA. A fourth group of sequences was more distantly related to
152 CtRLVaRNA. Besides CtRLVaRNA and BWYVaRNA, no other associated RNA had been
153 reported in carrots before. Full genomes of CRLVaRNA and BWYVaRNA were obtained by
154 extension of contigs or scaffolds using CLC Genomics Workbench v21 as explained in
155 Materials and Methods. In the case of some isolates, full genome sequences were directly
156 obtained as judged from comparison with the conserved 3' and 5' untranslated regions (UTR)
157 (Campbell et al., 2020) of GenBank reference sequences, obviating the need to perform RACE
158 experiment. For some others, a few 5' or 3' nucleotides were missing but no effort was made to
159 determine the 5' end 3' genome ends. The sequences of CRLVaRNA and BWYVaRNA isolates
160 thus determined have been deposited in GenBank under accession numbers ON603907 -
161 ON603916.

162 The sequence of an isolate representative of the fourth group only distantly related to
163 CtRLVaRNA was assembled from the sequencing library from an individually extracted wild
164 carrot (FR19-9, Table S1). The assembly covered the entire genome, which was confirmed by
165 sequence comparisons with the conserved 3' and 5' (UTR) of CtRLVaRNAs so that RACE
166 experiments were not needed. The finalized genome (2,840 nt long, GenBank ON603908)
167 included 355,574 reads (~4 % of total reads for that sample) with a high average coverage of
168 14,177x. The name carrot red leaf virus-associated RNA-2 (CtRLVaRNA-2) has been
169 proposed for this novel molecule (see below).

170 The full genome of an isolate representative of the group showing close homology (93% nt
171 identity) with ALVEaRNA was obtained from the assembly derived from the sequencing reads
172 from a single cultivated carrot plant (FR19-7, Table S1). The initial large contig identified was
173 extended through repeated round of residual reads mapping and the 5' and 3' genome ends
174 determined by RACE experiments. The full genome (GenBank ON603911) is 2,864 nt long
175 and integrates 82,202 reads corresponding to 4.51% of total reads for this sample with an
176 average coverage of 3,242x. Detailed information on all full genomes of tlaRNA isolates
177 identified in the present study are provided in Supplementary Table S3.

178 ***Genome organization of the identified tlaRNAs***

179 The genome organization of all identified tlaRNAs isolates shows the typical genome
180 organization for these molecules (Sanger, 1994; Campbell et al., 2020 & Yoshida, 2020) with
181 a very short 5' UTR (Campbell, 2020), two conserved ORFs, ORF1a and ORF1b in the same
182 reading frame and a long 3' UTR (Figure 1). The 5' UTR of the novel CtRLVaRNA-2 (GenBank
183 ON603908) is 14 nt long, while the situation is more complex with ALVEaRNA (GenBank
184 ON603911) since it shows three AUG initiation codons in close succession at its 5' genome end
185 at nt positions 5-7, 8-10 and 29-31. None of these is in a good Kozak initiation context but the
186 most likely candidate appears to be that at positions 29-31, since it leads to a protein starting
187 with the MCAALS sequence which is similar to the MQAALS sequence of CtRLVaRNA-2
188 (Figure 1). The ORF1a of CtRLVaRNA-2 and ALVEaRNA (696 and 757 nt, respectively)
189 encode proteins of about 25 kDa. For all analyzed isolates, this ORF is terminated by an amber
190 stop codon immediately followed in CtRLVaRNA and in other tlaRNAs by a GGL triplet. The
191 sole exception is ALVEaRNA for which the GGL triplet is located upstream of the amber stop
192 codon (Figure S1). Readthrough of the amber stop codon generates the ORF1a-ORF1b RdRp
193 fusion protein. The ORFs of CtRLVaRNA-2 and ALVEaRNA are roughly colinear with
194 respective length and genome positions of 2289 nt (position 15-2303) and 2343 nt (position 29-

195 2371). Characteristically, tlaRNAs have a long 3' UTRs and this is also observed for
196 CtRLVaRNA-2 and ALVEaRNA, with 3' UTRs of 537 and 492 nt respectively (Figure 1).

197 ***Recombination analysis and phylogenetic affinities of the various tlaRNAs identified in the***
198 ***carrot virome***

199 To reveal their phylogenetic affinities and taxonomic position, the complete genomes of the
200 various tlaRNAs identified in the carrot virome were aligned with complete tlaRNA genomes
201 retrived from the GenBank database. A recombination analysis using the RDP4 package and
202 this whole genome alignment revealed a single significant recombination event, which was
203 detected by all seven algorithms (corrected probability for the null hypothesis of $5e^{-33}$) and
204 involving a *ca.* 100 nt segment of CtRLVaRNA integrated near the 5' end of CABYVaRNA
205 (genome positions 22-120). This recombination event had been previously identified (Campbell
206 et al., 2020).
207 The same multiple alignment was then used to reconstruct the neighbour joining phylogenetic
208 tree of Figure 2. As previously shown by Campbell et al. (2020), tlaRNAs cluster into three
209 different clades supported by high bootstrap values and that show a high between-group average
210 nt divergence of 51.3-54.4% +/- 1%. The most diverse clade assembles
211 TuYVaRNA/BWYVaRNA plus CABYVaRNA, PeVYVaRNA, PoPeYVaRNA, TBTDaRNA
212 and TVDVaRNA (Figure 2). The second clade includes CtRLVaRNA and CtRLVaRNA-2,
213 with an average nt divergence between them of 31.1% +/- 0.7% (27.0% +/- 1.3% aa divergence
214 for the viral fusion protein) and the last clade includes only ALVEaRNA, with an average
215 divergence between the partial sequences from Arracacha and those from carrot of only 6.8%
216 +/- 0.5% (4% +/- 0.7% aa average divergence). The CtRLVaRNA variants reported here show
217 an average divergence of 9.2% +/- 0.3% (5.9% +/- 0.5% at protein level) with those already
218 present in GenBank, while the corresponding value for the carrot BWYVaRNA and the

219 BWYVaRNA/TuYVaRNA isolates already in GenBank is 8.7% +/- 0.3% (6.2% +/-0.5% at
220 protein level).

221 ***Distribution, prevalence and variability of tlaRNAs in wild and cultivated carrot populations***
222 In France, CtRLVaRNA was present in all 16 populations sampled in the Nouvelle-Aquitaine
223 region in 2019 (four cultivated, four off-type and eight wild carrot populations). A similarly
224 high prevalence was observed in the 2020 sampling, with CtRLVaRNA present in 10 cultivated
225 populations, 14 wild populations and three off-type populations, while only one cultivated and
226 one wild population (FR20-27 and FR20-22, Table S1) from northern France (Normandy and
227 Hauts-de-France) had no CtRLVaRNA infection. ALVEaRNA was present in 11/16 of the 2019
228 sampled populations including four cultivated, four off-type and three wild populations. Among
229 the populations sampled in 2020, eight cultivated, three off-type populations (all located in the
230 Nouvelle-Aquitaine region) and eight wild populations (two of which situated in northern
231 France, in Normandy and Île-de-France, respectively) were infected with ALVEaRNA. Over
232 the two years sampling in France, the new CtRLVaRNA-2 was present in 12 wild populations
233 and only one cultivated population that bordered a wild population in which CtRLVaRNA-2
234 was present (FR20-25 and FR20-26, respectively, Table S1). All CtRLVaRNA-2 infected
235 populations were from the Nouvelle-Aquitaine region. In contrast, the prevalence of
236 BWYVaRNA was low since it was detected only in two off-type populations and one cultivated
237 population from the Nouvelle-Aquitaine region, as well as from two wild populations from
238 Normandy and Hauts-de-France. Overall rates of detection of the various tlaRNAs in French
239 carrot populations are given in Table 1.
240 Populations sampled in Spain showed a similarly high prevalence of CtRLVaRNA, which was
241 present in all five cultivated populations but not in the wild one. ALVEaRNA was similarly
242 frequent as it was detected in all six sampled populations. BWYVaRNA and the new
243 CtRLVaRNA-2 tlaRNA were not detected in any of these six Spanish populations.

244 Plants from two populations [one wild (FR19-9) and one cultivated (FR19-7)] of the 2019
245 sampling in France were analyzed by HTS individually, which revealed a difference in the
246 prevalence of the analysed tlaRNAs. CtRLVaRNA showed a high prevalence in cultivated
247 carrots (86% infection, 46/50) but only infected 20% (10/50) of wild ones. ALVEaRNA had a
248 lower prevalence but still showed preferential infection of cultivated plants with infection
249 detected in 28% (14/50) of cultivated plants but only 2% (1/50) in the wild ones. This trend was
250 recently confirmed by PCR screening of single plants in additional populations (D.S.
251 unpublished results). CtRLVaRNA-2 was only identified in a single plant of the wild population
252 while BWYVaRNA had not been identified from the pools of the two individually analysed
253 populations and was, as expected, not identified in the corresponding single plants.
254 In an effort to get a broader insight into the variability of tlaRNA isolates in carrot populations
255 all tlaRNA contigs spanning the region corresponding to positions 1135-1845 of the central
256 domain of ORF1b of CtRLVaRNA (AF020617) were retrieved from populations and individual
257 plant assemblies. This region was selected striking a balance between the length of the region
258 (and therefore the amount of phylogenetic information) and the number of isolates for which
259 data could be obtained from HTS contigs. The obtained sequences were used to create a
260 multiple alignment with the corresponding region of GenBank isolates. Figure 3 presents the
261 Neighbor-Joining phylogenetic tree reconstructed from this multiple alignment. The isolates of
262 CtRLVaRNA-2 identified in this study from a highly homogenous ensemble with a low
263 diversity in the analyzed region of 1.6% +/- 0.7% (0.8% +/- 0.7 aa divergence). Similarly, the
264 identified ALVEaRNA isolates found in both French and Spanish carrot populations are very
265 similar with a diversity of 1.5% +/- 0.7% (0.9% +/- 0.3% aa divergence). The sole exception
266 concerns two French isolates (one from cultivated and one from off-type carrots) that form a
267 third cluster within ALVEaRNA, distinct from both the arracacha South-American isolates and
268 the European carrot isolates. In contrast, the variability between of CtRLVaRNA isolates is

269 significant, with an average nt diversity of 6.3% +/- 0.5% (1.6% +/- 0.7% aa diversity). Indeed,
270 the CtRLVaRNA isolates cluster in two ensembles with strong bootstrap support and a highly
271 divergent sigma isolate (KM486093). In turn the two clusters each separate into subclusters,
272 some of which had not been identified from the full-length genomes available in GenBank or
273 reported here (Figure 3). No obvious correlation between the type of the carrot hosts and the
274 phylogeny of CtRLVaRNA or ALVEaRNA isolates could be identified (Supplementary
275 Figures S2 and S3) suggesting that the populations of these tlaRNAs in wild and cultivated
276 carrots largely function as single metapopulations.

277 **DISCUSSION**

278 CtRLVaRNA and BWYVaRNA were the only tlaRNA detected in carrots so far. In the present
279 study, two additional tlaRNAs were identified in the virome of cultivated and wild carrots. All
280 identified tlaRNAs show a typical genome organization for this group of molecules, with a very
281 short 5' UTR, a long 3' UTR and ORF1a and ORF1b separated by an amber stop codon, leading
282 to the production of a ORF1a/ORF1b RdRp fusion protein of ~90kDa. Similar to umbraviruses,
283 tlaRNAs are dependent on helper viruses of the genus *Polerovirus* for encapsidation and
284 transmission. However, unlike the RdRp of tlaRNAs, which is expressed through a readthrough
285 event, umbraviruses express their RdRp by a -1 frameshift. They also possess two additional
286 ORFs for cell-to-cell and long-distance movement (Liu et al., 2021). The role of umbraviruses
287 in some specific disease complexes is well documented (Taliansky & Robinson, 2003; Murant
288 et al., 1969), while less is known about the contribution of tlaRNAs to such disease complexes.
289 The individually analysed plants in this study were frequently coinfected with CtRLVaRNA,
290 CMoV, and/or CMoMV but interestingly, of the 10 individually analysed wild plants showing
291 tlaRNAs-umbravirus coinfection, about a third (4/10) did not show evidence of the presence a
292 a helper virus of the *Polerovirus* or *Enamovirus* genera. Eventhough, it is not possible to draw

293 a firm conclusion from this observation, tlaRNAs are known to be self-replicating (Campbell
294 et al., 2020), so it could be hypothesized that on occasion they could be transmitted by vectors
295 without a concomitant transmission of the helper virus, in particular if presence of the tlaRNA
296 or of umbravirus(es) had a negative impact on helper virus accumulation (Yoshida, 2020). In
297 such a scenario, there exists the possibility of an assistance for movement of tlaRNAs by
298 coinfecting umbraviruses but, in the absence of encapsidation and of the aphid transmission it
299 permits, this situation could represent an epidemiological dead end.

300 Currently, there are no molecular criteria established by the International Committee on
301 Taxonomy of Viruses (ICTV) for satellites species demarcation. CtRLVaRNA-2 forms a well
302 separated cluster from CtRLVaRNA isolates with on average less than 70% nucleotide identity
303 (average nt divergence 31.1% +/- 0.7%). Its RdRp is similarly divergent, with on average 27.0%
304 +/- 1.3% as compared to that of CtRLVaRNA isolates (Figure 3). At both nucleotide and protein
305 levels, these values are close to those separating CABYVaRNA from
306 TuYVaRNA/BWYVaRNA (nt: 33.6% +/- 1.8%; aa: 29.7% +/- 0.8%). Based on these elements,
307 the CtRLVaRNA-2 isolates reported here are proposed to belong to a new, distinct species.

308 Until the present work ALVEaRNA was only known through two partial sequences submitted
309 in GenBank and identified during virome studies of *Arracacia xanthorrhiza* (Arracacha, family
310 *Apiaceae*) in Peru. ALVEaRNA isolates were predominantly obtained from field-grown carrots
311 accounting for 89.3% (25/28) of the identified isolates (Supplementary Table S4), while the
312 remaining three isolates come from wild carrots (10.7%; 3/28). These isolates form a separate
313 cluster from the partial sequences from Peru (Figure 3) but show a high degree of identity with
314 them (6.8% +/- 0.5% average nt divergence) and there is little doubt that they should be
315 considered as belonging to the same satellite species. Arracacha latent virus E (ALVE,
316 GenBank MF136435, MF073198), the helper virus of ALVEaRNA in Arracacha, is reported
317 in GenBank as an unclassified *Luteoviridae,* but its RdRp phylogenetic affinities and genomic

318 organization place it in the *Enamovirus* genus. No evidence for the presence of ALVE was
319 found in any of the individual carrot plants or carrot populations analyzed in the present study,
320 indicating that (an) other virus(es) fullfil the role of helper virus for ALVEaRNA in Europe.
321 CtRLV is a strong candidate because the ALVEaRNA-infected plants (15 of the total 100
322 individual plants examined) were always coinfected with CtRLV and only one of these plants
323 was infected with a possible other helper virus in the form of a novel *Enamovirus*. These
324 observations reinforce the notion that at least some tlaRNAs present a degree of promiscuity
325 and can be assisted by different helper viruses. The near identity of tlaRNAs associated with
326 BWYV and TuYV (Figure 2) point in the same direction and so does the partial sequence of a
327 tlaRNA associated with potato leafroll virus (MF133518; Campbell et al., 2020) which is 99.5%
328 identical to CtRLVaRNA isolates. Similarly, another tlaRNA identified during the present
329 study is very closely related to BWYVaRNA isolates or TuYVaRNA isolates, with an average
330 8.7% nt divergence (6.2% aa divergence). Remarkably, similar to the present finding,
331 TuYVaRNA was detected in British carrots in the absence of BWYV or TuYV but in the
332 presence of CtRLV (Adams et al. 2014), suggesting once again promiscuity of tlaRNAs and an
333 ability of TuYVaRNA to be assisted by CtRLV.

334 CtRLVaRNA was detected in 43 out of 45 populations from four different French regions (Île-
335 de-France, Normandy, Occitanie and Nouvelle-Aquitaine) and in five out of six carrot
336 populations sampled in central Spain, confirming its wide distribution. At plant population level
337 it did not show preferential detection in one or another carrot type. However, when analyzing
338 single plants, it was much higher in cultivated than in wild carrots (86% vs 20%) a situation
339 that has been confirmed by the PCR testing of individual plants from additional populations
340 (not shown). For their part ALVEaRNA and CtRLVaRNA-2 showed preferential association
341 to one type of carrot population. ALVEaRNA was detected more frequently in cultivated carrot
342 populations than in wild ones (79.9% vs 47.8%) and had a higher prevalence in cultivated single

343 plants than in wild one (28% vs 2%). The situation is reversed for CtRLVaRNA-2 which was
344 detected much more frequently in wild (52.2%) than in cultivated (6.7%) populations (Table
345 1). Interestingly, the single cultivated population in which it was detected was bordered by a
346 CtRLVaRNA-2 positive wild population, and in a 500 nt shared region the contigs from the
347 two populations showed only 3 single nucleotide polymorphisms. Taken together these results
348 suggest a flow from the wild carrots to the cultivated ones. CtRLVaRNA-2 also showed a
349 limited geographical distribution being detected only in the Nouvelle-Aquitaine region of
350 France.

351 The off-type populations differed phenologically and morphologically from cultivated carrots,
352 showing early bolting and not or very poorly tuberized roots, both traits linking them to wild
353 carrots. However, in a number of cases their roots had the orange color typical of cultivated
354 carrots, suggesting they might represent pollen or seed contamination during commercial seed
355 production resulting in hybrid genotypes. Even though their phenotype was evidently closer to
356 that of wild carrots, off-type populations showed tlaRNAs detection frequencies close to those
357 of cultivated carrots growing side by side rather than to those of wild carrots (Table 1),
358 suggesting that whatever the underlying mechanism(s), differences in tlaRNAs prevalence are
359 likely more influenced by growth conditions than by plant genotype.

360 The results reported here extended our knowledge of the host range and diversity of tlaRNAs
361 and provide further insights in the variability of associations they are able to establish with their
362 helper viruses, demonstrating the flexibility of such systems of assistance and
363 complementation. With four different tlaRNAs identified, the carrot virome has proven
364 particularly rich in these unusual agents, some of which like CtRLVaRNA and ALVEaRNA
365 appear to show a wide distribution and a high prevalence. Given these elements, it will be of
366 much interest to analyse the specificity of associations and the complex interplay between co-

367 dependent viruses (poleroviruses, umbraviruses and satellites of various types) in the carrot
368 pathosystem and their epidemiological consequences.

369 **Acknowledgments:** This study was funded by the European Union through the Horizon 2020
370 Marie Skłodowska-Curie Actions Innovative Training Network (H2020 MSCA- 60 ITN)
371 project "INEXTVIR" (GA 813542). The authors wish to thank the Platform GetPlaGe
372 (GenoToul, INRAE, Toulouse, France) for the Illumina sequencing of the French and Spanish
373 samples, F. Garcia-Arenal, A. Fraile and L. Svanella-Dumas (UMR BFP, INRAE, Bordeaux)
374 for help in collecting carrot samples and S. Plas (Invenio), S. Dandin (Ctifl Sileban) and F.
375 Villeneuve (Ctifl Lanxade) for arranging access to carrot crops. The authors declare that they
376 have no conflict of interest.

377

378 **Data availability:** The genome sequences for the tlaRNAs reported here have been deposited
379 in GenBank. The raw sequence datasets are available on request from the authors

380

381

382 **References**

383 Adams, I.P., Skelton, A., Macarthur, R., Hodges, T., Hinds, H., Flint, L., Deb Nath, P.,
384 Boonham, N., & Fox, A. (2014) Carrot yellow leaf virus Is Associated with Carrot Internal
385 Necrosis. *PLoS ONE*, 9, e109125.

386 Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) Basic local
387 alignment search tool. *Journal of Molecular Biology*, 215, 403–410.

388 Badar, U., Venkataraman, S., AbouHaidar, M. & Hefferon, K. (2021) Molecular interactions
389 of plant viral satellites. *Virus Genes* 57, 1–22.

390 Campbell, A.J., Erickson, A., Pellerin, E., Salem, N., Mo, X., Falk, B.W. & Ferriol, I. (2020)
391 Phylogenetic classification of a group of self-replicating RNAs that are common in co-
392 infections with poleroviruses. *Virus Research*, 276, 197831.

393 Chin, L.S., Foster, J.L., Falk, B.W. (1993) The beet western yellows virus ST9-associated
394 RNA shares structural and nucleotide sequence homology with carmo-like viruses.
395 *Virology*, 192, 473-482.

396 François, S., Filloux, D., Fernandez, E., Ogliastro, M., & Roumagnac, P. (2018) Viral
397 Metagenomics Approaches for High-Resolution Screening of Multiplexed Arthropod and
398 Plant Viral Communities. In: Pantaleo, V. & Chiumenti, M. (Eds.) Viral Metagenomics:
399 Methods and Protocols. New York, NY: Humana Press, pp 77-95.

400 Gnanasekaran, P. & Chakraborty, S. (2018) Biology of viral satellites and their role in
401 pathogenesis. *Current Opinion in Virology*, 33, 96–105.

402 Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018) MEGA X: molecular
403 evolutionary genetics analysis across computing platforms. *Molecular biology and*
404 *evolution*, 35, 1547.

405 Kurath, G., Rey, M.E.C. & Dodds, J.A. (1993) Tobamovirus helper specificity of satellite
406 tobacco mosaic virus involves a domain near the 5 ’end of the satellite genome. *Journal of*
407 *general virology*, 74, 1233–1243.
408 Liu, J., Carino, E., Bera, S., Gao, F., May, J.P., & Simon, A.E. (2021) Structural analysis and
409 whole genome mapping of a new type of plant virus subviral rna: Umbravirus-like
410 associated RNAs, *Viruses*, 13, 646.
411 Marais, A., Svanella-Dumas, L., Barone, M., Gentit, P., Faure, C., Charlot, G. et al. (2011)
412 Development of a polyvalent RT-PCR detection assay covering the genetic diversity of
413 Cherry capillovirus A (CVA), *Plant Pathology*. 61, 195–204
414 Marais, A., Faure, C., Bergey, B., Candresse, T. (2018) Viral double-stranded RNAs
415 (dsRNAs) from Plants: Alternative nucleic acid substrates for high-throughput sequencing.
416 In: Pantaleo, V. & Chiumenti, M. (Eds.) Viral Metagenomics: Methods in Molecular
417 Biology. New York, NY: Humana Press, pp 45–53.
418 Mo, X.H., Chen, Z.B., & Chen, J.P. (2011) Molecular identification and phylogenetic analysis
419 of a viral RNA associated with the Chinese tobacco bushy top disease complex. *Annals of*
420 *Applied Biology*, 158, 188–193.
421 Murant, A.F., Goold, R.A., Roberts, I.M., & Cathro, J. (1969) Carrot Mottle--a Persistent
422 Aphid-borne Virus with Unusual Properties and Particles. *Journal of General Virology*. 4,
423 329–341.
424 Oncino, C., Hemmer, O. & Fritsch, C. (1995) Specificity in the Association of Tomato Black
425 Ring Virus Satellite RNA with Helper Virus. *Virology*, 213, 87-96.
426 Passmore, B.K., Sanger, M., Chin, L.S., Falk, B.W. & Bruening, G. (1993) Beet western
427 yellows virus-associated RNA: An independently replicating RNA that stimulates virus
428 accumulation. *Proceedings of the National Academy of Sciences of the United States of*
429 *America,* 90, 10168–10172.

430 Peng, J., Bu, S., Yin, Y., Hua, M., Zhao, K., Lu, Y. et al. (2021) Biological and Genetic
431 Characterization of Pod Pepper Vein Yellows Virus-Associated RNA From Capsicum
432 frutescens in Wenshan, China. *Frontiers in Microbiology*, 12, 662352.
433 Sanger, M., Passmore, B., Falk, B.W., Bruening, G., Ding, B. & Lucas, W.J. (1994) Symptom
434 severity of beet western yellows virus strain ST9 is conferred by the ST9-associated RNA
435 and is not associated with virus release from the phloem. *Virology,* 200, 48–55.
436 Taliansky, M.E., & Robinson, D.J. (2003) Molecular biology of umbraviruses: Phantom
437 warriors. *Journal of General Virology*, 84, 1951–1960.
438 Tang, J., Quinn, B.D. & Clover, G.R.G. (2009) First report of Carrot red leaf virus-associated
439 RNA co-infecting carrot with Carrot red leaf virus and Carrot mottle mimic virus to cause
440 carrot motley dwarf disease in New Zealand. *Australasian Plant Disease Notes*, 4, 15–16.
441 Watson, M.T., Tian, T., Estabrook, E. & Falk, B.W. (1998) A Small RNA Resembling the
442 Beet Western Yellows Luteovirus ST9-Associated RNA Is a Component of the California
443 Carrot Motley Dwarf Complex. *Phytopathology*, 88, 164-170.
444 Yoshida, N. (2020) Biological and genetic characterization of carrot red leaf virus and its
445 associated virus/RNA isolated from carrots in Hokkaido, Japan. *Plant Pathology*, 69,
446 1379–1389.
447 Yoshida, N. & Tamada, T. (2019) Host range and molecular analysis of Beet leaf yellowing
448 virus, Beet western yellows virus-JP and Brassica yellows virus in Japan. *Plant Pathology*,
449 68, 1045–1058.
450

List of tables and figures

Table 1: Rate of detection of the different tlaRNAs in carrot populations sampled in France

Figure 1: Genome organization of CtRLVaRNA-2 and ALVEaRNA. ORF1a of CtRLVaRNA-2 and ALVEaRNA codes for a putative protein of 25.8kDa and 27.8 kDa, respectively, and ends in an amber stop codon (UAG). ORF1b is expressed by readthrough of ORF1a and encodes the viral RNA-dependent RNA polymerase.

Figure 2: Neighbour-joining phylogenetic tree reconstructed from the alignment of the full genome sequence of tlaRNAs using a strict nucleotide identity distance. Bootstrap values greater than 70% (1,000 repeats) are shown. Novel tlaRNA sequences reported here are indicated by black shapes of different shapes for each tlaRNA. The accession numbers of reference sequences retrieved from GenBank are indicated. The three clusters of tlaRNAs are indicated with different shadings. The scale bar corresponds to 10% nucleotide divergence. ALVEaRNA: arrachacha latent virus E-associated RNA; TuYVaRNA: turnip yellows virus-associated RNA; BWYVaRNA: beet western yellows virus-associated RNA; PeVYVaRNA: pepper vein yellows-associated RNA; PoPeYVaRNA: pod pepper yellows virus-associated RNA; TBTDaRNA: tobacco bunchy top disease-associated RNA; TVDVaRNA: tobacco vein distortion virus-associated RNA; CABYVaRNA: cucurbit aphid borne virus-associated RNA; ctRLVaRNA: carrot red leaf virus-associated RNA; ctRLVaRNA-2: carrot red leaf virus-associated RNA 2

Figure 3: Neighbour-joining phylogenetic tree of tlaRNAs reconstructed from the alignment of nucleotide sequence of the central part of ORF1b corresponding to positions 1135-1845 of CtRLVaRNA AF020617. Bootstrap values greater than 70% (1,000 repeats) are shown. Sequences derived from the novel complete genomes shown in Figure 2 are indicated by black

474 diamonds, those from reference sequences obtained from GenBank are indicated white
475 diamonds, together with the relevant accession number. The four tlaRNAs are indicated by
476 different shadings, and that corresponding to the novel CtRLVaRNA-2 indicated on the right
477 side. The scale bar corresponds to 10% nucleotide divergence.

478 **Supplementary information**

479 **Figure S1:** multiple alignments of of the readthrough region of tlaRNAs-encoded P1-P2
480 fusion protein with the GLL motif highlighted

481 **Figure S2:** Neighbour-joining phylogenetic tree of CtRLVaRNA color coded for the type of
482 carrot population (cultivated, off-type, wild) from which the associated RNA was identified.
483 The tree was reconstructed from the alignment of nucleotide sequence of the central part of
484 ORF1b corresponding to positions 1135-1845 of CtRLVaRNA AF020617. Bootstrap values
485 greater than 70% (1,000 repeats) are shown.

486 **Figure S3:** Neighbour-joining phylogenetic tree of ALVEaRNA color coded for the type of
487 carrot population (cultivated, off-type, wild) from which the associated RNA was identified.
488 The tree was reconstructed from the alignment of nucleotide sequence of the central part of
489 ORF1b corresponding to positions 1135-1845 of CtRLVaRNA AF020617. Bootstrap values
490 greater than 70% (1,000 repeats) are shown.

491 **Table S1**: characteristics of all carrot populations sampled in the present study

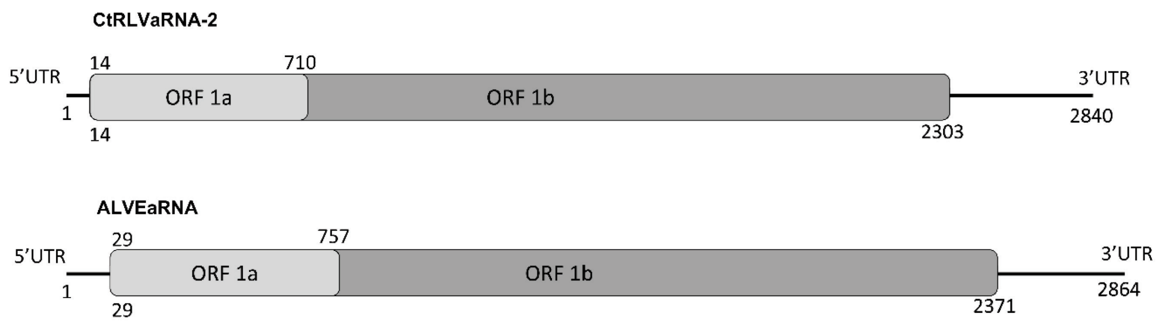
492 **Table S2**: specific detection and RACE primers used in the present study

493 **Table S3**: origin, sequence assembly parameters and accession numbers for the full length
494 genomes of tlaRNA isolates reconstructed during the present study

495 **Table S4:** origin and accession numbers for the partial ORF1b sequences of tlaRNA isolates
496 analyzed in the present study

497

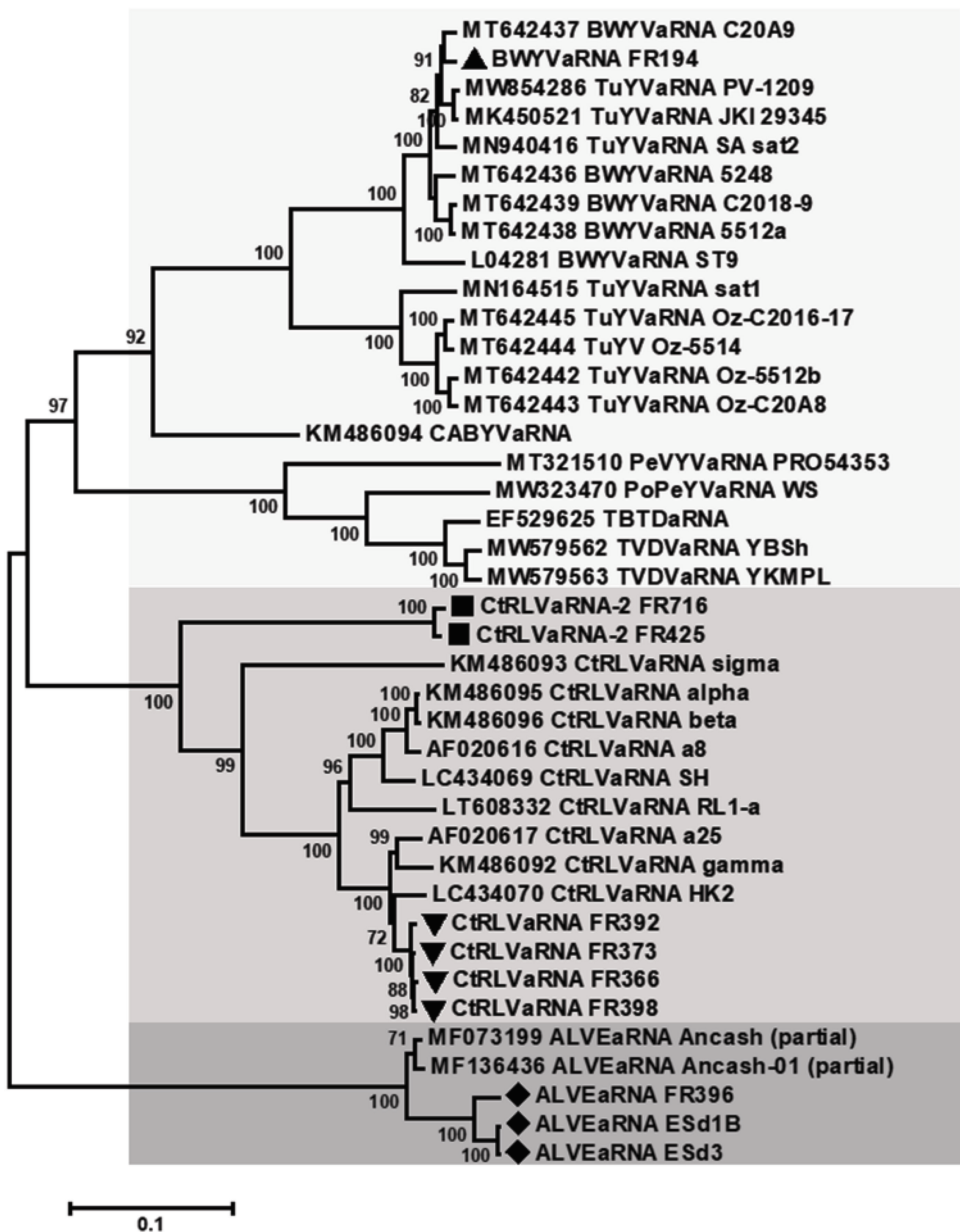
498 **Figure 1**



499

500

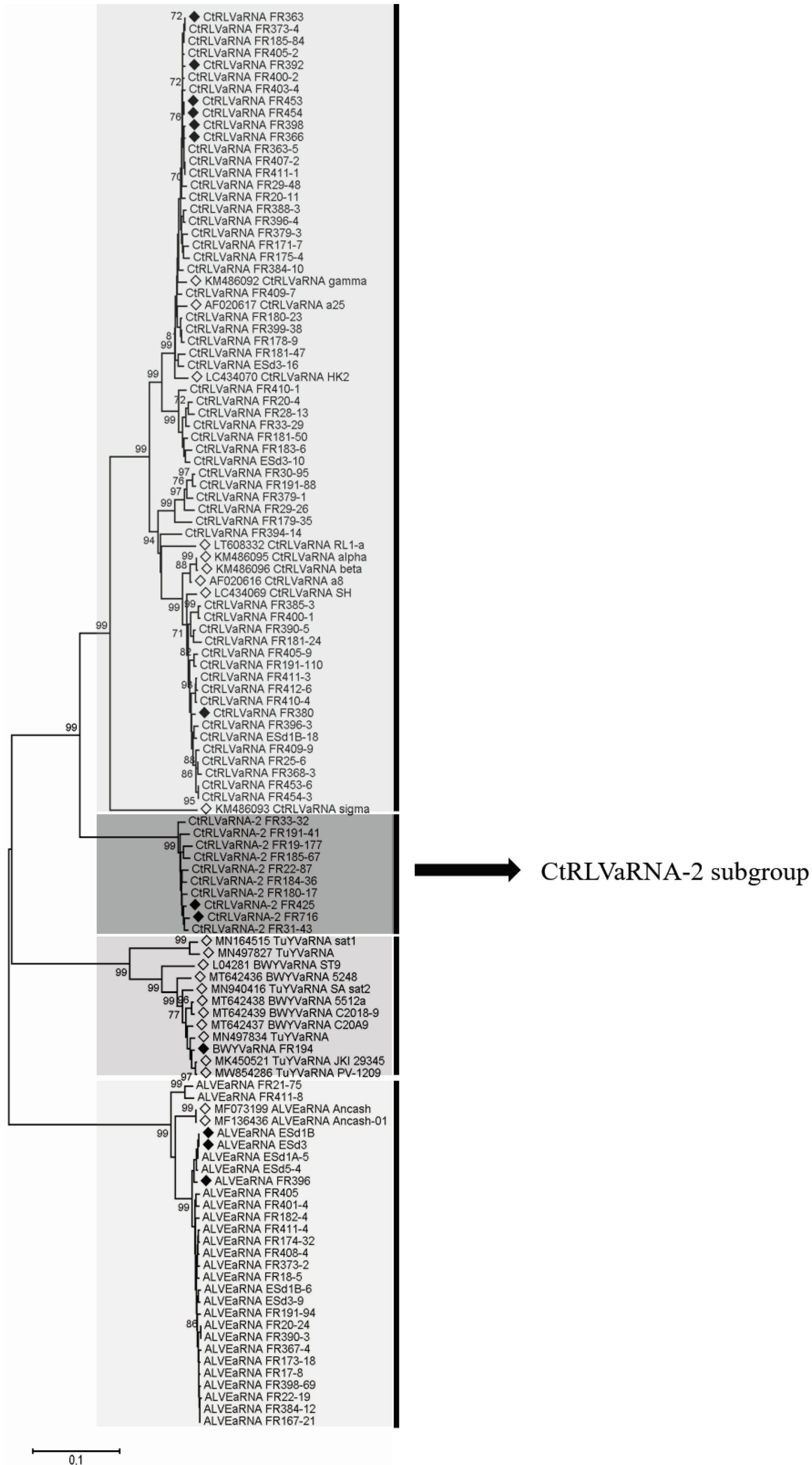
501 **Figure 2**



502

503

504 **Figure 3**



505

tlaRNA	Cultivated carrot populations	Wild carrots populations	Off-type[a] carrot populations
CtRLVaRNA	14/15 (93.3%)	22/23 (93.3%)	7/7 (100%)
ALVEaRNA	12/15 (79,9%)	11/23 (47,8%)	7/7 (100%)
CtRLVaRNA-2	1/15 (6.7%)	12/23 (52.2%)	0/7 (0%)
BWYVaRNA	1/15 (6.7%)	2/23 (8.7%)	2/7 (28,6 %)

506 **Table 1**

507 [a]carrot plants that were phenologically and morphologically more similar to wild carrots but
508 grew within cultivated carrot fields