A new flavi-like virus identified in populations of wild carrots

Deborah Schönegger, Armelle Marais, Chantal Faure, Thierry Candresse

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*

INTRODUCTION

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

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

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

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

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

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

MATERIALS AND METHODS

Plant samples collection and preparation

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

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

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

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

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

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

Analysis of tlaRNAs genomic sequences and Phylogenetic analyses

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

RESULTS

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

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

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

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

Genome organization of the identified tlaRNAs

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

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

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

To reveal their phylogenetic affinities and taxonomic position, the complete genomes of the various tlaRNAs identified in the carrot virome were aligned with complete tlaRNA genomes retrived from the GenBank database. A recombination analysis using the RDP4 package and this whole genome alignment revealed a single significant recombination event, which was detected by all seven algorithms (corrected probability for the null hypothesis of $5e^{-33}$) and involving a *ca.* 100 nt segment of CtRLVaRNA integrated near the 5' end of CABYVaRNA (genome positions 22-120). This recombination event had been previously identified (Campbell et al., 2020).

The same multiple alignment was then used to reconstruct the neighbour joining phylogenetic tree of Figure 2. As previously shown by Campbell et al. (2020), tlaRNAs cluster into three different clades supported by high bootstrap values and that show a high between-group average nt divergence of 51.3-54.4% +/- 1%. The most diverse clade assembles TuYVaRNA/BWYVaRNA plus CABYVaRNA, PeVYVaRNA, PoPeYVaRNA, TBTDaRNA and TVDVaRNA (Figure 2). The second clade includes CtRLVaRNA and CtRLVaRNA-2, with an average nt divergence between them of 31.1% +/- 0.7% (27.0% +/- 1.3% aa divergence for the viral fusion protein) and the last clade includes only ALVEaRNA, with an average divergence between the partial sequences from Arracacha and those from carrot of only 6.8% +/- 0.5% (4% +/- 0.7% aa average divergence). The CtRLVaRNA variants reported here show an average divergence of 9.2% +/- 0.3% (5.9% +/- 0.5% at protein level) with those already present in GenBank, while the corresponding value for the carrot BWYVaRNA and the

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

Distribution, prevalence and variability of tlaRNAs in wild and cultivated carrot populations

In France, CtRLVaRNA was present in all 16 populations sampled in the Nouvelle-Aquitaine region in 2019 (four cultivated, four off-type and eight wild carrot populations). A similarly high prevalence was observed in the 2020 sampling, with CtRLVaRNA present in 10 cultivated populations, 14 wild populations and three off-type populations, while only one cultivated and one wild population (FR20-27 and FR20-22, Table S1) from northern France (Normandy and Hauts-de-France) had no CtRLVaRNA infection. ALVEaRNA was present in 11/16 of the 2019 sampled populations including four cultivated, four off-type and three wild populations. Among the populations sampled in 2020, eight cultivated, three off-type populations (all located in the Nouvelle-Aquitaine region) and eight wild populations (two of which situated in northern France, in Normandy and Île-de-France, respectively) were infected with ALVEaRNA. Over the two years sampling in France, the new CtRLVaRNA-2 was present in 12 wild populations and only one cultivated population that bordered a wild population in which CtRLVaRNA-2 was present (FR20-25 and FR20-26, respectively, Table S1). All CtRLVaRNA-2 infected populations were from the Nouvelle-Aquitaine region. In contrast, the prevalence of BWYVaRNA was low since it was detected only in two off-type populations and one cultivated population from the Nouvelle-Aquitaine region, as well as from two wild populations from Normandy and Hauts-de-France. Overall rates of detection of the various tlaRNAs in French carrot populations are given in Table 1.

Populations sampled in Spain showed a similarly high prevalence of CtRLVaRNA, which was present in all five cultivated populations but not in the wild one. ALVEaRNA was similarly frequent as it was detected in all six sampled populations. BWYVaRNA and the new CtRLVaRNA-2 tlaRNA were not detected in any of these six Spanish populations.

Plants from two populations [one wild (FR19-9) and one cultivated (FR19-7)] of the 2019 sampling in France were analyzed by HTS individually, which revealed a difference in the prevalence of the analysed tlaRNAs. CtRLVaRNA showed a high prevalence in cultivated carrots (86% infection, 46/50) but only infected 20% (10/50) of wild ones. ALVEaRNA had a lower prevalence but still showed preferential infection of cultivated plants with infection detected in 28% (14/50) of cultivated plants but only 2% (1/50) in the wild ones. This trend was recently confirmed by PCR screening of single plants in additional populations (D.S. unpublished results). CtRLVaRNA-2 was only identified in a single plant of the wild population while BWYVaRNA had not been identified from the pools of the two individually analysed populations and was, as expected, not identified in the corresponding single plants.

In an effort to get a broader insight into the variability of tlaRNA isolates in carrot populations all tlaRNA contigs spanning the region corresponding to positions 1135-1845 of the central domain of ORF1b of CtRLVaRNA (AF020617) were retrieved from populations and individual plant assemblies. This region was selected striking a balance between the length of the region (and therefore the amount of phylogenetic information) and the number of isolates for which data could be obtained from HTS contigs. The obtained sequences were used to create a multiple alignment with the corresponding region of GenBank isolates. Figure 3 presents the Neighbor-Joining phylogenetic tree reconstructed from this multiple alignment. The isolates of CtRLVaRNA-2 identified in this study from a highly homogenous ensemble with a low diversity in the analyzed region of 1.6% +/- 0.7% (0.8% +/- 0.7 aa divergence). Similarly, the identified ALVEaRNA isolates found in both French and Spanish carrot populations are very similar with a diversity of 1.5% +/- 0.7% (0.9% +/- 0.3% aa divergence). The sole exception concerns two French isolates (one from cultivated and one from off-type carrots) that form a third cluster within ALVEaRNA, distinct from both the arracacha South-American isolates and the European carrot isolates. In contrast, the variability between of CtRLVaRNA isolates is

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

DISCUSSION

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

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

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

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

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

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

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

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

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

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.

Peng, J., Bu, S., Yin, Y., Hua, M., Zhao, K., Lu, Y. et al. (2021) Biological and Genetic Characterization of Pod Pepper Vein Yellows Virus-Associated RNA From Capsicum frutescens in Wenshan, China. *Frontiers in Microbiology*, 12, 662352.

Sanger, M., Passmore, B., Falk, B.W., Bruening, G., Ding, B. & Lucas, W.J. (1994) Symptom severity of beet western yellows virus strain ST9 is conferred by the ST9-associated RNA and is not associated with virus release from the phloem. *Virology,* 200, 48–55.

Taliansky, M.E., & Robinson, D.J. (2003) Molecular biology of umbraviruses: Phantom warriors. *Journal of General Virology*, 84, 1951–1960.

Tang, J., Quinn, B.D. & Clover, G.R.G. (2009) First report of Carrot red leaf virus-associated RNA co-infecting carrot with Carrot red leaf virus and Carrot mottle mimic virus to cause carrot motley dwarf disease in New Zealand. *Australasian Plant Disease Notes*, 4, 15–16.

Watson, M.T., Tian, T., Estabrook, E. & Falk, B.W. (1998) A Small RNA Resembling the Beet Western Yellows Luteovirus ST9-Associated RNA Is a Component of the California Carrot Motley Dwarf Complex. *Phytopathology*, 88, 164-170.

Yoshida, N. (2020) Biological and genetic characterization of carrot red leaf virus and its associated virus/RNA isolated from carrots in Hokkaido, Japan. *Plant Pathology*, 69, 1379–1389.

Yoshida, N. & Tamada, T. (2019) Host range and molecular analysis of Beet leaf yellowing virus, Beet western yellows virus-JP and Brassica yellows virus in Japan. *Plant Pathology*, 68, 1045–1058.

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

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

Supplementary information

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

Figure S2: Neighbour-joining phylogenetic tree of CtRLVaRNA color coded for the type of carrot population (cultivated, off-type, wild) from which the associated RNA was identified. The tree was 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.

Figure S3: Neighbour-joining phylogenetic tree of ALVEaRNA color coded for the type of carrot population (cultivated, off-type, wild) from which the associated RNA was identified. The tree was 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.

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

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

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

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

Figure 1

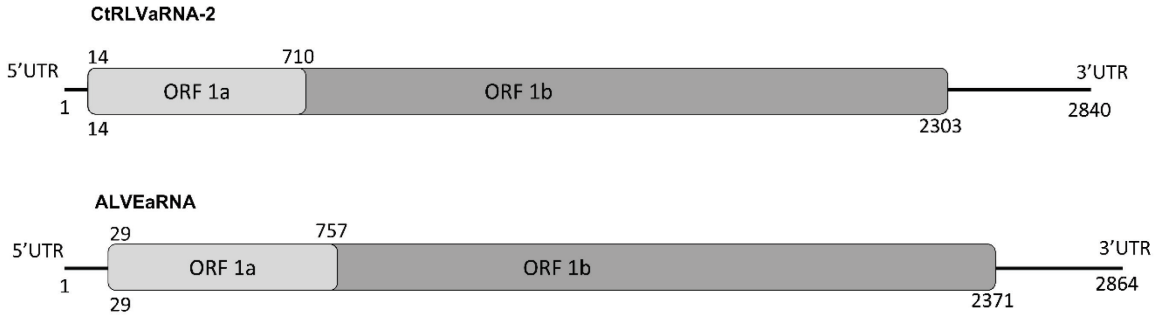
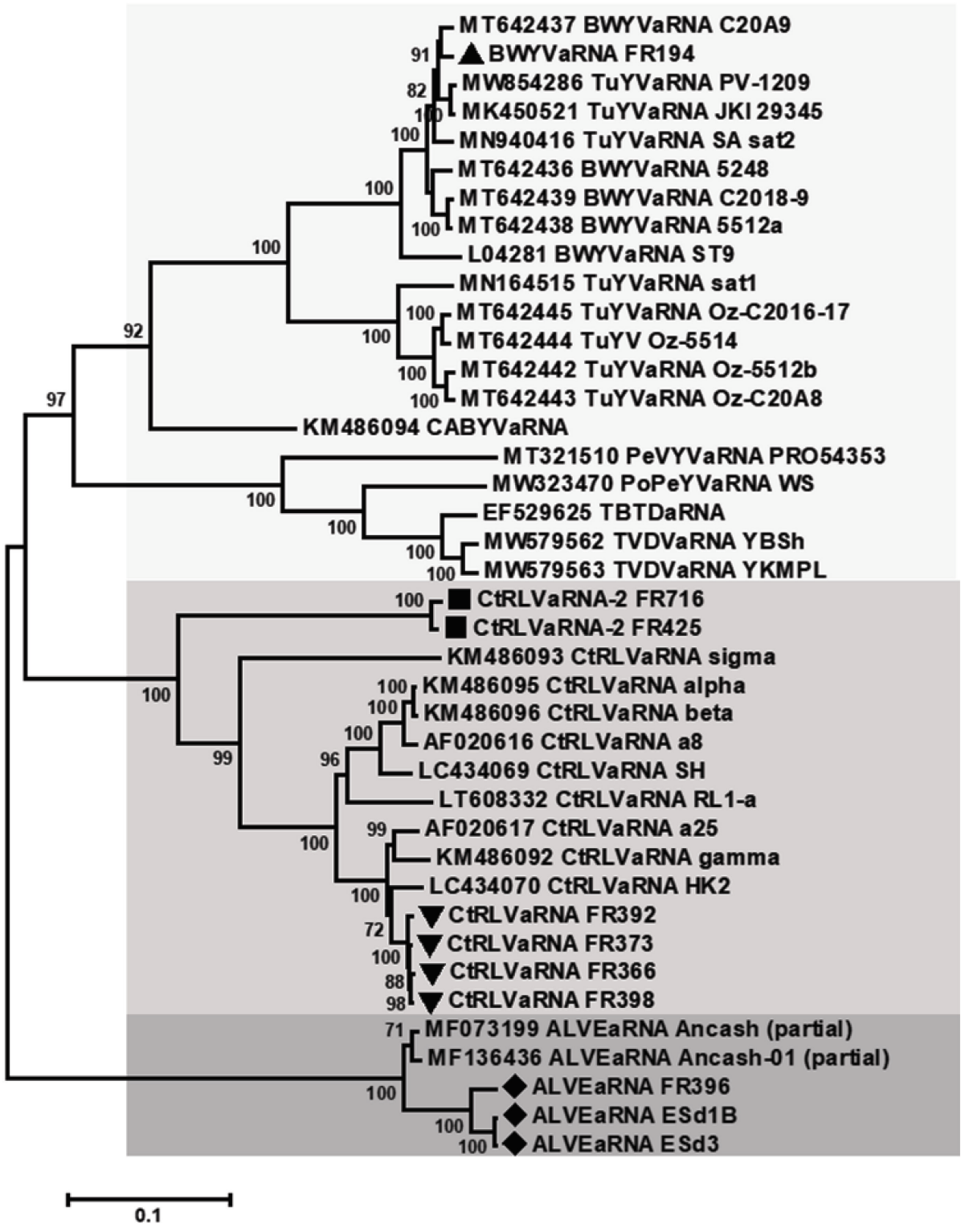
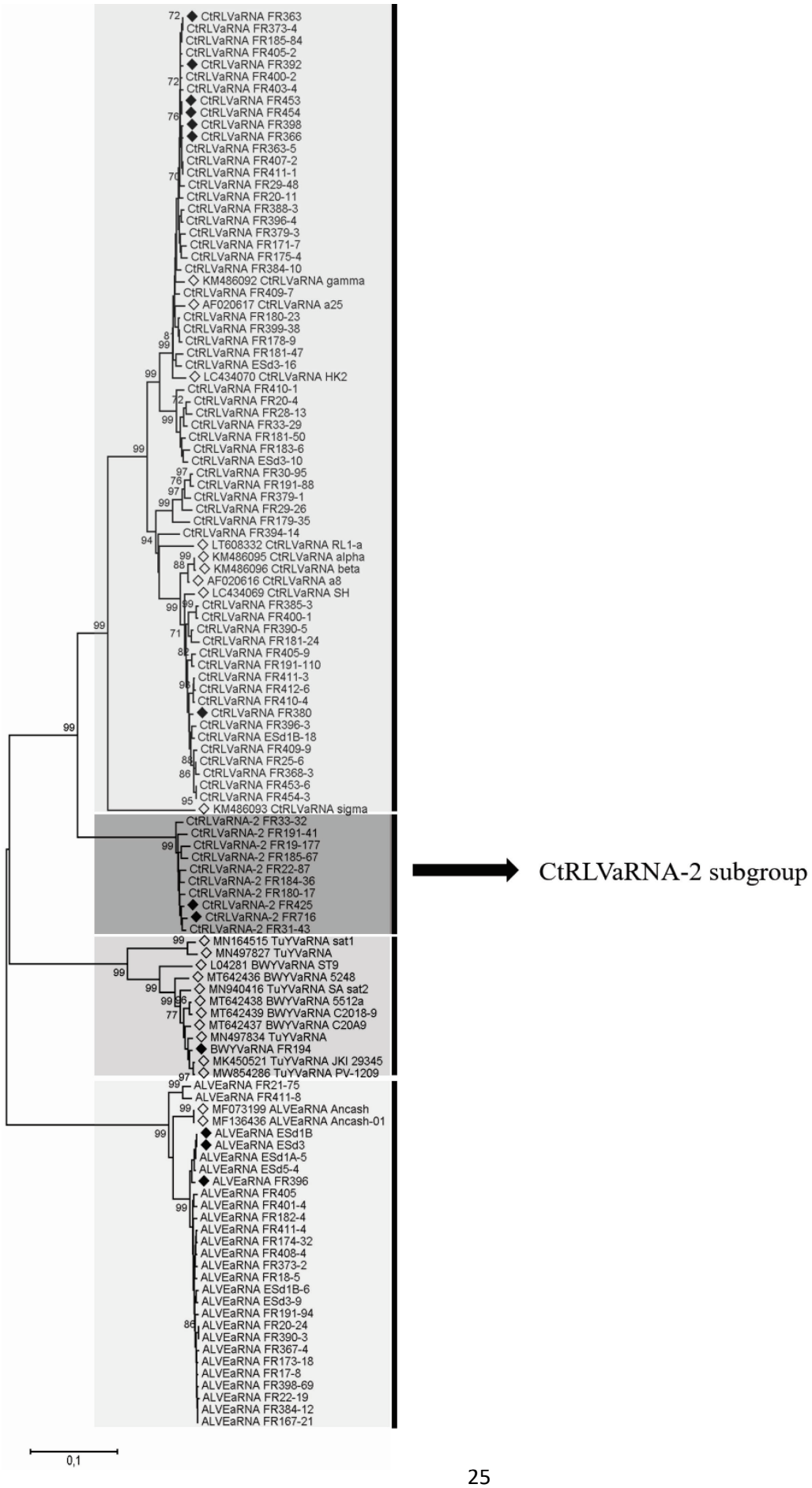

CtRLVaRNA-2
5'UTR
14
710
ORF 1a
ORF 1b
3'UTR
1
14
2303
2840
ALVEaRNA
29
757
5'UTR
ORF 1a
ORF 1b
3'UTR
1
29
2371
2864

Figure 2


MT642437 BWYVaRNA C20A9
BWYVaRNA FR194
MW854286 TuYVaRNA PV-1209
MK450521 TuYVaRNA JKI 29345
MN940416 TuYVaRNA SA sat2
MT642436 BWYVaRNA 5248
MT642439 BWYVaRNA C2018-9
MT642438 BWYVaRNA 5512a
L04281 BWYVaRNA ST9
MN164515 TuYVaRNA sat1
MT642445 TuYVaRNA Oz-C2016-17
MT642444 TuYV Oz-5514
MT642442 TuYVaRNA Oz-5512b
MT642443 TuYVaRNA Oz-C20A8
KM486094 CABYVaRNA
MT321510 PeVYVaRNA PRO54353
MW323470 PoPeYVaRNA WS
EF529625 TBTDaRNA
MW579562 TVDVaRNA YBSh
MW579563 TVDVaRNA YKMPL
CtRLVaRNA-2 FR716
CtRLVaRNA-2 FR425
KM486093 CtRLVaRNA sigma
KM486095 CtRLVaRNA alpha
KM486096 CtRLVaRNA beta
AF020616 CtRLVaRNA a8
LC434069 CtRLVaRNA SH
LT608332 CtRLVaRNA RL1-a
AF020617 CtRLVaRNA a25
KM486092 CtRLVaRNA gamma
LC434070 CtRLVaRNA HK2
CtRLVaRNA FR392
CtRLVaRNA FR373
CtRLVaRNA FR366
CtRLVaRNA FR398
MF073199 ALVEaRNA Ancash (partial)
MF136436 ALVEaRNA Ancash-01 (partial)
ALVEaRNA FR396
ALVEaRNA ESd1B
ALVEaRNA ESd3
0.1

504 **Figure 3**


CtRLVaRNA FR363
CtRLVaRNA FR373-4
CtRLVaRNA FR185-84
CtRLVaRNA FR405-2
CtRLVaRNA FR392
CtRLVaRNA FR400-2
CtRLVaRNA FR403-4
CtRLVaRNA FR453
CtRLVaRNA FR454
CtRLVaRNA FR398
CtRLVaRNA FR366
CtRLVaRNA FR363-5
CtRLVaRNA FR407-2
CtRLVaRNA FR411-1
CtRLVaRNA FR29-48
CtRLVaRNA FR20-11
CtRLVaRNA FR388-3
CtRLVaRNA FR396-4
CtRLVaRNA FR379-3
CtRLVaRNA FR171-7
CtRLVaRNA FR175-4
CtRLVaRNA FR384-10
KM486092 CtRLVaRNA gamma
CtRLVaRNA FR409-7
AF020617 CtRLVaRNA a25
CtRLVaRNA FR180-23
CtRLVaRNA FR399-38
CtRLVaRNA FR178-9
CtRLVaRNA FR181-47
CtRLVaRNA ESd3-16
LC434070 CtRLVaRNA HK2
CtRLVaRNA FR410-1
CtRLVaRNA FR20-4
CtRLVaRNA FR28-13
CtRLVaRNA FR33-29
CtRLVaRNA FR181-50
CtRLVaRNA FR183-6
CtRLVaRNA ESd3-10
CtRLVaRNA FR30-95
CtRLVaRNA FR191-88
CtRLVaRNA FR379-1
CtRLVaRNA FR29-26
CtRLVaRNA FR179-35
CtRLVaRNA FR394-14
LT608332 CtRLVaRNA RL1-a
KM486095 CtRLVaRNA alpha
KM486096 CtRLVaRNA beta
AF020616 CtRLVaRNA a8
LC434069 CtRLVaRNA SH
CtRLVaRNA FR385-3
CtRLVaRNA FR400-1
CtRLVaRNA FR390-5
CtRLVaRNA FR181-24
CtRLVaRNA FR405-9
CtRLVaRNA FR191-110
CtRLVaRNA FR411-3
CtRLVaRNA FR412-6
CtRLVaRNA FR410-4
CtRLVaRNA FR380
CtRLVaRNA FR396-3
CtRLVaRNA ESd1B-18
CtRLVaRNA FR409-9
CtRLVaRNA FR25-6
CtRLVaRNA FR368-3
CtRLVaRNA FR453-6
CtRLVaRNA FR454-3
KM486093 CtRLVaRNA sigma
CtRLVaRNA-2 FR33-32
CtRLVaRNA-2 FR191-41
CtRLVaRNA-2 FR19-177
CtRLVaRNA-2 FR185-67
CtRLVaRNA-2 FR22-87
CtRLVaRNA-2 FR184-36
CtRLVaRNA-2 FR180-17
CtRLVaRNA-2 FR425
CtRLVaRNA-2 FR716
CtRLVaRNA-2 FR31-43
MN164515 TuYVaRNA sat1
MN497827 TuYVaRNA
L04281 BWYVaRNA ST9
MT642436 BWYVaRNA 5248
MN940416 TuYVaRNA SA sat2
MT642438 BWYVaRNA 5512a
MT642439 BWYVaRNA C2018-9
MT642437 BWYVaRNA C20A9
MN497834 TuYVaRNA
BWYVaRNA FR194
MK450521 TuYVaRNA JKI 29345
MW854286 TuYVaRNA PV-1209
ALVEaRNA FR21-75
ALVEaRNA FR411-8
MF073199 ALVEaRNA Ancash
MF136436 ALVEaRNA Ancash-01
ALVEaRNA ESd1B
ALVEaRNA ESd3
ALVEaRNA ESd1A-5
ALVEaRNA ESd5-4
ALVEaRNA FR396
ALVEaRNA FR405
ALVEaRNA FR401-4
ALVEaRNA FR182-4
ALVEaRNA FR411-4
ALVEaRNA FR174-32
ALVEaRNA FR408-4
ALVEaRNA FR373-2
ALVEaRNA FR18-5
ALVEaRNA ESd1B-6
ALVEaRNA ESd3-9
ALVEaRNA FR191-94
ALVEaRNA FR20-24
ALVEaRNA FR390-3
ALVEaRNA FR367-4
ALVEaRNA FR173-18
ALVEaRNA FR17-8
ALVEaRNA FR398-69
ALVEaRNA FR22-19
ALVEaRNA FR384-12
ALVEaRNA FR167-21
0,1
CtRLVaRNA-2 subgroup

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 %)

Table 1

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