**A new *Potyvirus* from hedge mustard (*Sisymbrium officinale* (L.) Scop.) sheds light on the evolutionary history of *Turnip mosaic virus***

**Iason Tsarmpopoulos%, Armelle Marais, Chantal Faure, Sébastien Theil$,and Thierry Candresse\***

Univ. Bordeaux, INRAE, UMR BFP, CS 20032, F-33882 Villenave d'Ornon Cedex, France

**\*** **Author for correspondence:** [thierry.candresse@inrae.fr](file:///C%3A%5CUsers%5Ciason%5CAppData%5CLocal%5CPackages%5Cmicrosoft.windowscommunicationsapps_8wekyb3d8bbwe%5CLocalState%5CFiles%5CS0%5C3%5CAttachments%5Cthierry.candresse%40inrae.fr) Tel : (+33) 557 12 23 89

% Present address: iason.tsarmpopoulos@ucare-watches.com

$ Present address: INRAE, UMRF, Aurillac, France

The nucleotide sequence reported in this work has been deposited in the GenBank database under accession number OP257374

**Abstract**

A novel potyvirus was identified in symptomatic hedge mustards (*Sisymbrium officinale* (L.) Scop.) and wild radish (*Raphanus raphanistrum* L.) in France. The near complete genome sequence of hedge mustard mosaic virus (HMMV) was determined, demonstrating it belongs to a sister species to turnip mosaic virus (TuMV). HMMV readily infected several other *Brassicaceae*, including turnip, shepherd's purse (*Capsella bursa-pastoris*) and Arabidopsis. The identification of HMMV as a *Brassicaceae*-infecting virus closely related to TuMV questions the current scenario of TuMV evolution and suggests a possible alternative one in which transition from a monocot-adapted ancestral lifestyle to a *Brassicaceae*-adapted one could have occurred earlier than previously recognized.

**Keywords:** *Sisymbrium*, Hedge mustard, *Potyvirus*, turnip mosaic virus, host jump, evolution

The genus *Potyvirus* is the largest genus of plant-infecting RNA viruses. Among its members, the *Brassicaceae*-adapted turnip mosaic virus (TuMV) [1] has been the subject of extensive evolutionary studies, allowing the proposal of scenarios for its emergence [2-6] and for its subsequent spread to various areas of the world [7-9]. TuMV is the only dicot-infecting virus in the so-called TuMV phylogenetic group [3,4], a small ensemble of phylogenetically related monocot-infecting potyviruses which includes Japanese yam mosaic virus, scallion mosaic virus, narcissus yellow stripe virus, narcissus late season yellows virus and wild onion symptomless virus [3,5,10,11]. This situation, together with the discovery of four TuMV isolates in wild orchids in Germany that are basal in TuMV phylogeny and poorly adapted to *Brassicaceae* hosts has led to the currently favored evolutionary scenario in which the current *Brassicaceae*-infecting TuMV isolates evolved in Europe about a thousand years ago from orchid-infecting, monocot-adapted ancestors [4]. This early event was then followed about 150 years later by TuMV diversification into several lineages, some of which are characterized by their ability to infect different *Brassicaceae* genera such as *Brassica* and *Raphanus* [4, 6]. Phylogeographic analyses have recently allowed proposal of timed spread scenarios on a global scale for several of these lineages [8, 9]. The development of such evolutionary scenarios relies heavily on phylogenetic analyses and, therefore, on the identification of viral isolates representing divergent, basal groups or closely related species. This situation is well illustrated in the genus *Potyvirus* by the discovery of a divergent potato virus Y (PVY) isolate that allowed reconstruction of an evolutionary history in which adaptation to pepper and potato was modified several times during PVY history [12, 13]. We report here the discovery and characterization of a new *Brassicaceae*-infecting Potyvirus closely related to TuMV which suggests a plausible, alternative evolutionary history scenario for TuMV.

The new virus was identified in southwestern France (Aquitaine region) in hedge mustard (*Sisymbrium officinale* (L.) Scop.), a wild indigenous plant sampled in Villenave d'Ornon in 2010 and 2011 during a study aiming to describe the virome associated with common weeds growing side by side with a range of crops (including tomato, pepper, turnip, radish, lettuce, spinash etc...) in an horticultural site, through the high-throughput sequencing of highly purified double-stranded RNAs (dsRNAs) [14].

Double-stranded RNAs were purified, converted to cDNA and randomly amplified and tagged [15] before being submitted to 454 pyrosequencing. After demultiplexing and quality trimming, reads were *de novo* assembled using CLC Genomics Workbench and annotated using BLASTn and BLASTx against the GenBank database. This allowed the identification from hedge mustard plants of contigs with homology with members of the *Potyvirus* genus. In order to obtain a complete genome for the virus, Illumina resequencing (2x100 nt) of dsRNAs purified from a single turnip plant that had been inoculated using an infected hedge mustard (isolate 1079, see below) yielded 384,497 high quality trimmed reads. The contigs *de novo* assembled from those reads were then assembled into a scaffold spanning a complete potyvirus genome and integrating 65,656 reads (16.6% of total reads, average coverage 476x). The genome ends were determined using 5' RACE and 3' LD-PCR [16] and a low coverage internal region resequenced using the primers described in Supplementary Table 1. The completed viral genome is 9,741 nucleotides (nt) long without the polyA tail and has been deposited in GenBank (OP257374) under the hedge mustard mosaic virus (HMMV) name (see below). The genomic RNA harbors a single long open reading frame (ORF) of 9,405 nt which encodes a 3,135 amino acids (aa) long polyprotein. The 5’ and 3’ non-coding regions (NCRs) are 108 and 219 nt long, respectively, but comparison with the 5' NCR of TuMV suggests that the HMMV 5' genome end may not have been reached and that 14 nucleotides are probably missing at the 5' end since nt 1-12 of the HMMV sequence are identical to nt 15-26 of the TuMV one. The genome and the encoded polyprotein present all the expected hallmarks of a potyvirus genomic organization, including presence of a PIPO ORF [17] and all the expected conserved motifs (KITC and PTK like motifs in HCPro; G2A6 nucleotide motif upstream of PIPO; helicase signature in CI; GDD in replicase, catalytic triads in proteases (P1, HCPro, NIa-Pro); DAG in CP...) and cleavage sites in the viral polyprotein.

Blast analysis shows that the virus closest to the new virus is turnip mosaic virus, with an overall complete genome nt identity of 64.5-65.8% (species cut off for this parameter 76%) [18]. This close relationship is confirmed by nt and aa comparisons of the CP gene and CP protein, which show respectively 70.8-73.5% nt identity (species cut off 76%) and 74.7-77.9% aa identity (species cut off 80%). Taken together, these results show that the new agent is clearly distinct from TuMV and represents a novel species for which the name hedge mustard mosaic virus (HMMV) is therefore proposed. This conclusion is also supported by differences observed between TuMV and HMMV in the various cleavage sites of their polyproteins (Table 1).

A recombination analysis performed using RDP4 [19] failed to identify any recombination event between HMMV and selected TuMV isolates. A phylogenetic analysis was then performed using complete polyprotein sequences for representative members of the various TuMV phylogroups, including TuMV isolates identified in monocot hosts, and for the monocot-infecting viruses identified as forming the TuMV lineage. This analysis unambiguously places HMMV in that lineage and identifies it as the closest relative of TuMV (Figure 1). Together with TuMV, HMMV is therefore the second virus from the TuMV lineage to be identified naturally infecting *Brassicaceae* hosts.

The host range of HMMV was further evaluated by mechanical inoculation of plants of various species, using the 1079 isolate passaged in turnip that has been shown to be free of contamination by TuMV or by any other virus by Illumina HTS analysis. The results are presented in Table 2. Systemic, symptomatic infection of the inoculated plants was identified in several *Brassicaceae* species including turnip (*Brassica rapa*), hedge mustard, sheperd's purse (*Capsella bursa-pastoris*), *Arabidopsis thaliana* and wild radish (*Raphanus raphanistrum*) and was validated using a Potyvirus genus-specific ELISA assay (Agdia, Elkhart, Indiana, USA) and an HMMV-specific RT-PCR assay using primers 1079-potyF2/1079-potyR2 (Supplementary Table 1). In particular, turnip developed typical leaf mosaic symptoms while hedge mustard and arabidopsis showed leaf deformation and drying. On the other hand, no infection was detected in radish (cv. Rose de Chine) or cabbage (broccoli, white cabbage or cauliflower types), contrary to the results obtained with a TuMV isolate in parallel experiments. Systemic, symptomatic infection was also observed in a few non-*Brassicaceae* hosts including *Nicotiana benthamiana*, *N. clevelandii* and garden pea (*Pisum sativum*), while no evidence for systemic infection was obtained in all other inoculated species (Table 2). In *N. benthamiana* symptoms of stunting and chlorosis were observed. In parallel, sampling in fall 2013 and spring 2014 of wild radish populations at the site where the original infected hedge mustards had been collected in 2010-2011 did not reveal any infection in 2013 (0/48 sampled plants) but showed a 5.1% prevalence in spring 2014 (9/177 plants). Taken together these results confirm HMMV as a *Brassicaceae*-infecting virus under natural conditions, suggesting that it is well adapted to this host family. They also provide evidence for differences in host range between TuMV and HMMV, in particular the failure of HMMV to infect lettuce and *Chenopodium quinoa*, two known systemic hosts of TuMV. However, it should be considered that these results rest on the testing of a single HMMV isolate leaving open the possibility of some biological variability between HMMV isolates.

The identification of HMMV as both a *Brassicaceae*-infecting virus and the closest relative to TuMV questions the currently proposed evolutionary scenario that has ancestral TuMV isolates as monocot-adapted, orchid-infecting viruses which jumped hosts in Europe about a thousand years ago [4] to become the widely successful *Brassicaceae*-infecting virus we know today [6, 9]. Indeed, since HMMV and TuMV are today *Brassicaceae*-adapted viruses, an alternative scenario is possible, in which their common ancestor was already a *Brassicaceae*-adapted virus and with the monocots-to-dicots host jump having occurred at some earlier point, prior to the separation of TuMV and HMMV. In this scenario, the recently discovered orchid TuMV isolates [4] would be seen as having reverted to the ancestral monocot-infecting lifestyle of the TuMV lineage. It should be noted that the two scenarios are equally parsimonious, both requiring two host jump events, and are in both cases based on a limited number of viral isolates with a basal position in TuMV phylogeny. In the current scenario [4], the two host jumps correspond to two independent jumps for HMMV and TuMV, transitioning from a monocot-adapted lifestyle to a *Brassicaceae*-adapted one. In the alternate scenario suggested here, the two host changes correspond to an early adaptation to *Brassicaceae*, predating the separation of HMMV and TuMV, and to a more recent reversion of the orchid TuMV isolates to a monocot-adapted lifestyle. Although equally parsimonious and therefore of comparable probability, it should be stressed that only 4 orchid-isolates of TuMV have ever been recorded (TuMV OM isolates) and that these were all identified in a single collection of wild orchids [4]. Also, these isolates failed to infect some monocot hosts on which they were assayed [*Narcissus tazetta* (*Amaryllidaceae*), *Dioscorea japonica* (*Dioscoreaceae*), *Allium fistulosum* (*Liliaceae*), *Orchis graminifolia* (*Orchidaceae*)] but infected two *Brassicaceae* species, *Camelina sativa* and, for two of three assayed isolates, *Eruca sativa* [4]. Thus, although it is clear that TuMV orchid isolates host range widely differs from that of typical TuMV isolates, their interpretation as ancestral, monocot-adapted isolates is weakened by the artificialized set-up in which they were identified and by the possibility that their adaptation might be recent and somehow linked to this set-up. It should also be noted that although a range of TuMV isolates have been recorded from monocot hosts, a phylogenetic analysis places them distally in the World-B or Basal-B phylogroups (Figure 1) so that they do not bear on the two discussed evolutionary scenarios discussed here. As a consequence, we suggest that the alternate scenario identified here, with a much older transition to a dicot-adapted lifestyle should be given consideration.

Whatever the evolutionary history of TuMV and HMMV, the discovery of HMMV as a *Brassicaceae*-infecting sister species to TuMV provides a cautionary note on developing evolutionary scenarios that can be compromised by the discovery of a single new virus or of a few isolates with distinct properties [12, 13]. It also illustrates how the exploration of the vast diversity of the virome of wild plants might allow us to reconstruct more accurately and over longer evolutionary times the history of viral lineages like that of TuMV. Outside of these evolutionary considerations, the finding of HMMV with a significant prevalence in indigenous *Brassicaceae* weeds and its ability to cause symptoms in at least some crops such as turnip or pea suggest that it might have pathological relevance under at least some circumstances.

**Acknowledgements**

The authors wish to thank the Platform GetPlaGe (GenoToul, INRAE, Toulouse, France) for the 454 pyrosequencing, the Platform GenomEast (IGBMC, CNRS, Strasbourg, France) for the Illumina sequencing, Laurence Svanella-Dumas (UMR BFP, INRAE) for help with field samples collection and Thierry Mauduit (UMR BFP, INRAE) for taking care of all experimental plants.

**Author statements – Compliance with ethical standards**

**Conflict of interest.** All authors declare they have no conflict of interest.

**Ethical approval.** This article does not contain any studies with human participants or animals performed by any of the authors.

**References**

1. Walsh JA, Jenner CE (2002) Turnip mosaic virus and the quest for durable resistance. Mol Plant Pathol 3:289–300. <https://doi.org/10.1046/j.1364-3703.2002.00132.x>
2. Tomimura K, Gibbs AJ, Jenner CE, Walsh JA, Ohshima K (2003) The phylogeny of Turnip mosaic virus; comparisons of 38 genomic sequences reveal a Eurasian origin and a recent ‘emergence’ in east Asia. Mol Ecol 12:2099–2111. <https://doi.org/10.1046/j.1365-294x.2003.01881.x>
3. Gibbs AJ, Ohshima K (2010) Potyviruses and the digital revolution. Annu Rev Phytopathol 48:205–223. <https://doi.org/10.1146/annurev-phyto-073009-114404>
4. Nguyen HD, Tomitaka Y, Ho SY, Duchêne S, Vetten HJ, Lesemann D, Walsh JA, Gibbs AJ, Ohshima K (2013) Turnip mosaic potyvirus probably first spread to Eurasian brassica crops from wild orchids about 1000 years ago. PLoS One 8:e55336. <https://doi.org/10.1371/journal.pone.0055336>
5. Gibbs AJ, Nguyen HD, Ohshima K (2015) The 'emergence' of turnip mosaic virus was probably a 'gene-for-quasi-gene' event. Curr Opin Virol 10:20–26. <https://doi.org/10.1016/j.coviro.2014.12.004>
6. Yasaka R, Fukagawa H, Ikematsu M, Soda H, Korkmaz S, Golnaraghi A, Katis N, Ho SYW, Gibbs AJ, Ohshima K (2017) The Timescale of emergence and spread of turnip mosaic potyvirus. Sci Rep 7:4240. <https://doi.org/10.1038/s41598-017-01934-7>
7. Tomitaka Y, Ohshima K (2006) A phylogeographical study of the Turnip mosaic virus population in East Asia reveals an ‘emergent’ lineage in Japan. Mol Ecol 15:4437–4457. <https://doi.org/10.1111/j.1365-294X.2006.03094.x>
8. Yasaka R, Ohba K, Schwinghamer MW, Fletcher J, Ochoa-Corona FM, Thomas JE, Ho SYW, Gibbs AJ, Ohshima K (2015) Phylodynamic evidence of the migration of turnip mosaic potyvirus from Europe to Australia and New Zealand. J Gen Virol 96:701–713. <https://doi.org/10.1099/jgv.0.000007>
9. Kawakubo S, Gao F, Li S, Tan Z, Huang YK, Adkar-Purushothama CR, Gurikar C, Maneechoat P, Chiemsombat P, Aye SS, Furuya N, Shevchenko O, Špak J, Škorić D, Ho SYW, Ohshima K (2021) Genomic analysis of the *Brassica* pathogen turnip mosaic potyvirus reveals its spread along the former trade routes of the Silk Road. Proc Natl Acad Sci USA 118:e2021221118. <https://doi.org/10.1073/pnas.2021221118>
10. Ohshima K, Korkmaz S, Mitoma S, Nomiyama R, Honda Y (2016) First genome sequence of wild onion symptomless virus, a novel member of *Potyvirus* in the turnip mosaic virus phylogenetic group. Genome Announc 4:e00851-16. <https://doi.org/10.1128/genomeA.00851-16>
11. Ohshima K, Mitoma S, Gibbs AJ (2018) The genetic diversity of narcissus viruses related to turnip mosaic virus blur arbitrary boundaries used to discriminate potyvirus species. PLoS One 13:e0190511. <https://doi.org/10.1371/journal.pone.0190511>
12. Moury B (2010) A new lineage sheds light on the evolutionary history of potato virus Y. Mol Plant Pathol 11:161–168. <https://doi.org/10.1111/j.1364-3703.2009.00573.x>
13. Moury B, Desbiez C (2020) Host range evolution of potyviruses: a global phylogenetic analysis. Viruses 12:111. <https://doi.org/10.3390/v12010111>
14. Svanella-Dumas L, Τsarmpopoulos Ι, Marais A, Theil S, Faure C, Gaudin J, Candresse T (2018) Complete genome sequence of lettuce chordovirus 1 isolated from cultivated lettuce in France. Arch Virol 163:2543–2545. <https://doi.org/10.1007/s00705-018-3858-y>
15. Marais A, Faure C, Bergey B, Candresse T (2018) Viral double-stranded RNAs (dsRNAs) from plants: alternative nucleic acid substrates for high-throughput sequencing. In: Pantaleo V, Chiumenti M (editors). Viral metagenomics: methods in molecular biology. New York, NY, Humana Press, pp. 45–53. <https://doi.org/10.1007/978-1-4939-7683-6_4>
16. Youssef F, Marais A, Faure C, Gentit P, Candresse T (2011) Strategies to facilitate the development of uncloned or cloned infectious full-length viral cDNAs: Apple chlorotic leaf spot virus as a case study. Virology J 8:488–500. <https://doi.org/10.1186/1743-422X-8-488>
17. Chung BY, Miller WA, Atkins JF, Firth AE (2008) An overlapping essential gene in the *Potyviridae*. Proc Natl Acad Sci USA 105:5897–5902. <https://doi.org/10.1073/pnas.0800468105>
18. Inoue-Nagata AK, Jordan R, Kreuze J, Li F, López-Moya JJ, Mäkinen K, Ohshima K, Wylie SJ (2022) ICTV Virus Taxonomy Profile: *Potyviridae* 2022. J Gen Virol 103:001738. <https://doi.org/10.1099/jgv.0.001738>
19. Martin DP, Murrell B, Khoosal A, Muhire B (2017) Detecting and Analyzing Genetic Recombination Using RDP4. Methods Mol Biol 1525:433–460. <https://doi.org/10.1007/978-1-4939-6622-6_17>

**Table 1. Comparisons of the amino acids at the polyprotein cleavage sites of turnip mosaic virus (TuMV) and hedge mustard mosaic virus (HMMV).**

|  |  |  |
| --- | --- | --- |
| **Cleavage site\*** | **TuMV** | **HMMV** |
| P1 / HC-Proa | I(I-V)H(F-Y)/S | MVHY/S |
| HC-Pro / P3 | YRVG/G | YRVG/G |
| P3 / 6K1 | Q/A | Q/A |
| 6K1 / CI | Q/T-Q/A | Q/S |
| CI / 6K2 | Q/S-Q/N | Q/S |
| 6K2 / VPg | E/A | E/G |
| VPg / NIa-Pro | E/S | E/S |
| NIa-Pro / NIb | Q/T | Q/A |
| NIb / CP | Q/A | Q/A |

\*: P1: P1 protein; HC-Pro: helper component-proteinase; P3: P3 protein 3; 6K1: 6K1 protein; CI: cylindrical inclusion protein; 6K2: 6K2 protein; VPg: genome-linked protein; NIa-Pro: nuclear inclusion a-proteinase; NIb; nuclear inclusion b polymerase; CP: capsid protein.

**Table 2**. **Experimental analysis of the host-range of hedge mustard mosaic virus (HMMV) following mechanical transmission**. The number of systemically infected plants over the number of inoculated plants is given.

|  |  |
| --- | --- |
| **Plant species** | **Infected/Inoculated** |
| *Sisymbrium officinale* (hedge mustard) | 12/12 |
| *Capsella bursa-pastoris* (sheperd's purse) | 12/12 |
| *Arabidopsis thaliana* | 24/24 |
| *Raphanus raphanistrum* (wild radish) | 10/11 |
| *Brassica rapa var rapa* (turnip) | 24/24 |
| *Pisum sativum* (garden pea) | 8/10 |
| *Nicotiana benthamiana* | 12/12 |
| *Nicotiana clevelandii* | 12/12 |
| *Raphanus sativus* (radish, cv Rose de chine) | 0/56 |
| *Brassica oleracea var. italica* (broccoli) | 0/12 |
| *Brassica oleracea var botrytis* (cauliflower cv. Merveille de toutes saisons) | 0/12 |
| *Brassica oleracea var botrytis* (cauliflower cv. Flora blanca) | 0/12 |
| *Brassica oleracea var capitata* (white cabbage) | 0/12 |
| *Nicotiana tabacum* (tabacco cv. Xanthi) | 0/12 |
| *Nicotiana tabacum* (tobacco cv. Samsun) | 0/24 |
| *Nicotiana occidentalis* | 0/12 |
| *Nicotiana rustica* | 0/12 |
| *Nicotiana debneyi* | 0/24 |
| *Lactuca sativa* (lettuce cv. Trocadero) | 0/12 |
| *Chenopodium. quinoa* | 0/12 |
| *Chenopodium amaranticolor* | 0/12 |
| *Chenopodium foetidum* | 0/24 |
| *Chenopodium murale* | 0/24 |

**Figure 1. Maximum likelihood phylogenetic tree reconstructed using the amino acid sequences of the complete polyprotein of hedge mustard mosaic virus (HMMV), of members of the TuMV phylogenetic group and of selected TuMV isolates**. Statistical significance of branches was evaluated by bootstrap analysis (1,000 replicates). Only values higher than 70% are indicated. The scale bar represents 5% amino acid divergence. GenBank accession numbers are indicated for each sequence, together with virus acronym and, in the case of TuMV, isolate name. TuMV isolates from monocot hosts are marked by grey triangles and the host is indicated. HMMV is marked by a black square. The various groups identified with TuMV diversity are indicated on the right of the tree. JYMV: Japanese yam mosaic virus; ScaMV: scallion mosaic virus; NYSV: narcissus yellow stripe virus; NLSYV: narcissus late season yellows virus; WOSV: wild onion symptomless virus.

**Figure 1**.

****

**Supplementary Table 1**. Name and sequence of the oligonucleotide primers used in this study, together with the size of the amplicons generated and the annealing temperature used in RT-PCR amplification.

|  |  |  |  |
| --- | --- | --- | --- |
| **Primer name** | **Primer sequence (5'-3')** | **Amplicon (nt)** | **Annealing temperature** |
| 1079-potyF1 | GCGTATGGAATGGTTCGTTGAAA | 630 | 66°C |
| 1079-PotyR1 | ACACTATGCGTTCCTCCTCCA |
| 1079-potyF2 | CGCTTGGTGAAGGGAATTTACA | 451 | 64°C |
| 1079-potyR2 | CGCATCTTCTAAGGTCAGAAGA |
| 1079-Poty-Race1a | ATGAGCTTCCTTGAATGCTACCCGTGTC | 717 | 70°C |
| 1079-Poty-NRace1a | GTACTGTTGCCATTCGTGTGTTTCAGG | 182 | 70°C |
| 1079-Poty-LD-Sys3 | GAACATGCATACACTCATGG | 253 | 68°C |
| LD-prim | CACTGGCGGCCGCTCGAGCATGTACT |

a: primers used in conjunction with the universal primer provided by the 5’ RACE kit