CLASSIFICATION OF TURNIP MOSAIC VIRUS ISOLATES ACCORDING TO THE 3'-UNTRANSLATED REGION

K. PETRZIK[1], P. LEHMANN[2]

[1]Institute of Plant Molecular Biology, Academy of Sciences of the Czech Republic, Branišovská 31, 370 05 České Budějovice, Czech Republic; [2]Institute of Plant Genetics, Polish Academy of Sciences, Poznan, Poland

Received May 23, 1996

Summary. – The 3'-untranslated region (3'UTR) of five isolates of turnip mosaic virus (TuMV) from United Kingdom, Canada, Greece, the Czech Republic and from Uzbekistan were sequenced and compared with another nine previously sequenced TuMV isolates. All the isolates had 209 nucleotides long 3'UTR, with the exception of the Uzbekistan isolate, which had one-base deletion at nucleotide (nt) position 162. Phylogenetic analysis identified three clusters of related isolates. The clusters correlated with secondary folding of the 3'UTRs, partially with the source host plant of the isolates but not with their geographical origin.

Key words: turnip mosaic virus; potyviridae; 3'-untranslated region; clustering

Plant virus taxonomy is going from initial reliance on morphological, biological and serological properties to assignment based on amino acid and nucleic acid sequences that should represent the ultimate taxonomy criteria today. Amino acid and/or nucleic acid sequences of replication-associated proteins serve as a base for classification of higher taxa (for review see Ward, 1993). The coat protein sequences are widely accepted for discrimination between virus species as well as for classification of isolates, especially in the family *Potyviridae* (Ward and Shukla, 1991). Comparison of 3'UTR of *Potyviridae* revealed similar degrees of homology between distinct viruses and between related strains to those obtained by coat protein comparison (Shukla and Ward, 1989). The 3'UTR is proposed to be an accurate marker of genetic relatedness and therefore could serve in classification of potyviruses (Frenkel *et al.*, 1989).

TuMV is the most important virus of cultivated cruciferous crops as oilseed rape and *Brassica* vegetables. This virus is distributed world-wide and isolates differing in vector transmissibility (Kantrong *et al.*, 1995), host plant, pathogenic and serological properties are known (McDonald and Hiebert, 1975; Provvidenti, 1982; Jenner and Walsh, 1996). A collection of world isolates of TuMV was prepared and techniques for their differentiation were intensively studied (P. Lehmann, unpublished results).

With increasing economical importance of oilseed rape especially for non-food production and increasing area of cropping, the risk of the occurrence of more pathogenic isolates and severe outbreaks of TuMV is rising. In this paper we present classification of previously sequenced TuMV isolates and five new isolates according to their 3'UTR, describe phylogenetic relatedness among them, and discuss relations between this classification, geographical origin and source host plant.

Five isolates of TuMV were obtained from naturally infected plants: the Czech isolate (CZ1) originated from cabbage *(Brassica oleracea capitata)*, the Greek isolate (GK1) from stock *(Matthiola incana)* and the Uzbekistan isolate (UZB1) from oilseed rape *(Brassica napus oleifera)*. Isolates from United Kingdom (UK1) and Canada (CDN1) were found on swede *(Brassica napus rapifera)*.

One µl of sap from mustard *Brassica juncea* cv. Tendergreen infected with the isolates mentioned above was reacted

Abbreviations: MMLV = moloney murine leukemia virus; TuMV = turnip mosaic virus; 3'UTR = 3'-untranslated region

with 50 U of MMLV reverse transcriptase and 2.5 mmol/l oligo(dT)$_{16}$ primer. The coat protein gene and the 3'UTR were amplified with 2.5 U of DNA polymerase (AmpliTaq, Perkin Elmer – Roche, USA) in 35 cycles (1 min at 94°C, 1 min at 55°C and 2 mins at 72°C) with 0.5 mmol/l each of primers 5'ATCAAGCTTCAGGCAATCTTTGAGGATTTAT3' and 5'GGCCACGCGTCGACTAGTACTCGAG(T)$_{17}$3'. Three independent pUC18 clones containing the amplified products were sequenced by Sequenase 2.0 kit (United States Biochemical) as recommended by the manufacturer.

Sequences of the 3'UTR of the Chinese isolate CHI (Kong *et al.*, 1990), of Japanese isolates JPN2 (Sano *et al.*, 1992), JPN1 and JPN31 (Nakashima *et al.*, 1991), of Canadian isolates CAN (Nicolas and Laliberte, 1992) and CDN (Tremblay *et al.*, 1990) were found in the respective literature. Sequences of the Korean isolates KOR1 (AC No. X83968) and KOR2 (X79366), and of the isolate of unknown origin (L12396) were obtained from the EMBL database.

The alignments of sequences were done using the CLUSTAL programme (Higgins and Sharp, 1988). The probable secondary RNA folding structures of all available 3'UTR sequences were computed using the RNAFOLD programme according to the Zuker and Stiegler (1981) algorithm.

All the 3'UTRs of the isolates CZE1, UK1, GK1 and CDN1 were of 209 nucleotides in length. The UZB1 isolate had one-base deletion at position 162 and it is so far the only isolate with 208 nucleotides long 3'UTR (Fig. 1). The TuMV 3'UTR sequences are highly conservative. We observed substitutions and one deletion at 28 different positions that represent about 13% of the length of the 3'UTR.

The 3'UTR sequences of the isolates CAN1, CAN, CDN, UK1, JPN1, JPN31, KOR1, CZE1 and L12396 differed from each other at 1 – 5 positions. Their mutual similarity was above 97%. The Canadian isolate CDN1 had two substitutions: T_{27}->C_{27} and C_{170}->T_{170} in comparison with the previously sequenced isolates CAN and CDN (numbering begins with the first nucleotide after the coat protein termination codon).

The isolates JPN2, KOR2 and CHI differed from the others at 10 – 14 positions. The isolates had identical substitutions: G_{12}->A_{12} (with the exception of the isolate KOR2), T_{27}->G_{27}, T_{97}->C_{97}, A_{148}->T_{148} and T_{200}->A_{200}. However, the most characteristic feature of them was the presence of an A_{167}TACTAT$_{173}$ motif. The function of this motif is not known, but according to the sequence and localization it resembles the second polyadenylation signal of the zucchini yellow mosaic potyvirus (Grumet and Fang, 1990). The newly sequenced isolates GK1 and UZB1 had this motif modified, but nucleotides T_{168} and AT_{173} remained conserved. Both the isolates were more closely related to the three Asian isolates mentioned above than to the European ones.

Phylogenetic analysis of the 14 isolates done by the CLUSTAL programme established three clusters of related isolates (Fig. 2). Cluster 1 contains isolates from three continents: North America (CAN, CDN and CDN1), Europe (UK1 and CZE1), Asia (JPN1, JPN31 and KOR1) and the isolate of unknown origin (L12396). Cluster 2 contains the Asian isolates CHI, JPN2 and KOR2, and the European isolate GK1. The isolate UZB1 is the only member of cluster 3. As the clusters 1 and 2 contain isolates from very distant areas, the correlation between the geographical origin and clustering of the isolates can be excluded.

We have checked the relatedness of the isolates according to their natural host, when known. The isolates of cluster 1 and 3 originated from various *Brassica* species only: the isolates UK1, CAN, CDN, CDN1 and UZB1 from *B. napus*, JPN31 and JPN1 from *B. rapa*, and the isolate CZE1 from *B. oleracea*. The isolates of cluster 2 originated from non-*Brassica* sources – GK1 from *Matthiola incana*, CHI and JPN2 from *Raphanus sativus*. In the case of the isolate KOR2 we record *Brassica campestris* var. *pekinensis* as a susceptible host, not as the original species.

We assume that our classification correlates with the natural host of TuMV (with the exception of the isolate KOR2). Finding and analysis of other non-*Brassica* isolates will be necessary for confirmation of these results.

A correlation has been found between clusters and their probable (computed) secondary structures. The isolates of cluster 1 folded in tRNA-like structure with stability of -191 to -208 kJ/mole (Fig. 3a). Isolates of cluster 2 formed quite different secondary structure. Nucleotides at positions 1-50 preferentially bound to nucleotides of the end of UTR and not to themselves as in the cluster 1 (Fig. 3b). The stability of this second type of structure was -182 to -201 kJ/mole. We have found by computer simulations that the motif TAT$_{173}$ of the isolates CHI, JPN2 and KOR2 or the motif AAAT$_{173}$ of the isolate GK1 together with the G_{22} and G_{161} are essential for this type of folding. The isolate UZB1 of the cluster 3 folded in third type of secondary structure (Fig. 3c), which is similar to the cluster 1 folding, but the stability of the UZB1 secondary structure was slightly lower (-186 kJ/mole). It is known that at least in the case of another potyvirus, potato virus Y, the 3'UTR could attenuate expression of symptoms, probably by altering the secondary structures (Van der Vlugt *et al.*, 1993). Similar effects of TuMV in certain host plant were found, too (Jenner and Walsh, 1996).

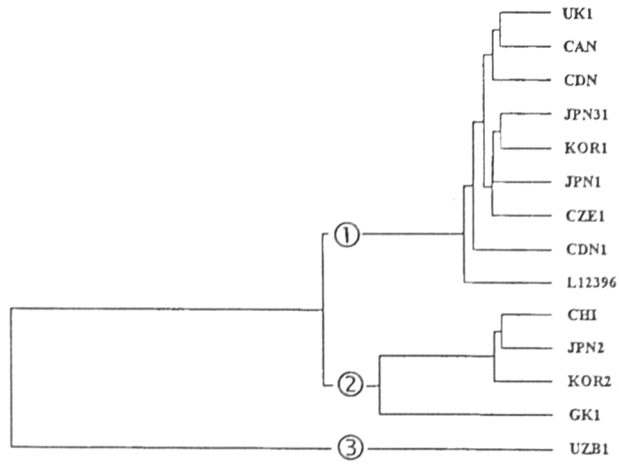
It is very difficult to ascertain why similar mutations occur in isolates originating from very distant regions. TuMV is spread mainly by aphids and the seed transmission was not confirmed. Therefore it is not possible to explain the similarity of the isolates JPN1, JPN31, and KOR1 to the European and North American isolates by seed exchange. The long distance transmission through TuMV-infected aphids or infected green material is of low probability, too. On the other hand, the variability is not so high to exclude the occurrence of an identical mutation in different countries. We have found that

```
UK1      AGTTGTATGCTGGTAGACTATAAGTATTTAAGTTTACTCGTTAGTATTCTCGCTTATGGGAAATATGTAA  70
CAN      ........................................................................  70
CDN      ........................................................................  70
CDN1     ........................C...............................................  70
JPN31    ........................................................................  70
JPN1     ........................................................................  70
KOR1     ........................................................................  70
CZE1     ........................................................................  70
L12396   ......................TT..............................................  70
CHI      ..........A.............G.............................................  70
JPN2     ..........A.............G.............................................  70
KOR2     ........................G.............................................  70
GK1      ......G....A.............G.............................................  70
UZB1     ..........A.............G.A...........................................  70

UK1      GTTTGTTAAAGCAGCCAGTGTGACTTTGTCATGTGTGTTGTTGTTACTTTCTGTATTTTCGCCGAACATT  140
CAN      ........................................................................  140
CDN      ........................................................................  140
CDN1     ........................................................................  140
JPN31    ....................................................A...................  140
JPN1     ............A...........................................A...................  140
KOR1     ....................................................A...................  140
CZE1     ....................................................A...................  140
L12396   ....................................................A...................  140
CHI      ........................C...........................A..............T..  140
JPN2     ........................C...........................A..............T..  140
KOR2     ........................C...........................A..............T..  140
GK1      ........................C...........................A...................  140
UZB1     ....................................................GA........T....TC..  140

UK1      TTATTGGTGTTAGCGCATGTAGTGAGGATCGTCCTCGATTGCCTTAACATTTGATAGGATGCAAGGGAC  209
CAN      ....................G...................................................  209
CDN      ....................G.............................C.....................  209
CDN1     ....................G........T...........................................  209
JPN31    ....................G........T...........................................  209
JPN1     ....................G...................................................  209
KOR1     ....................G...................................................  209
CZE1     ........................................................................  209
L12396   ....................G...................................................  209
CHI      ............T......G.....ATA.TAT.........................A.............  209
JPN2     ....................G.....ATA.TAT.........................A.............  209
KOR2     ....................G.....ATA.TAT.........................A.............  209
GK1      ....................G.....TTAAAAT.........................................  209
UZB1     C...................G-....AT...AT.........................G..T..A...     208
```

Fig. 1
Multiple alignment of the 3'UTR of thirteen TuMV isolates
Positions with substitutions occured are marked only.



five nucleotides only markedly influenced the secondary folding of 3'UTRs. Therefore isolates with modified pathogenic properties could arise from present isolates very easily.

The nucleotide sequences data reported in this paper have been submitted to the EMBL nucleotide sequence database and have been assigned the AC No. X96541 – for the CZE1 isolate, X96542 – GK1, X96543 – UK1, X96544 – CDN1, and X96545 – UZB1.

This work was supported by the EC grant No. ERBCIPACT923007.

Fig. 2
Phylogenetic relationship between the isolates of TuMV based on the homology in the 3'UTR

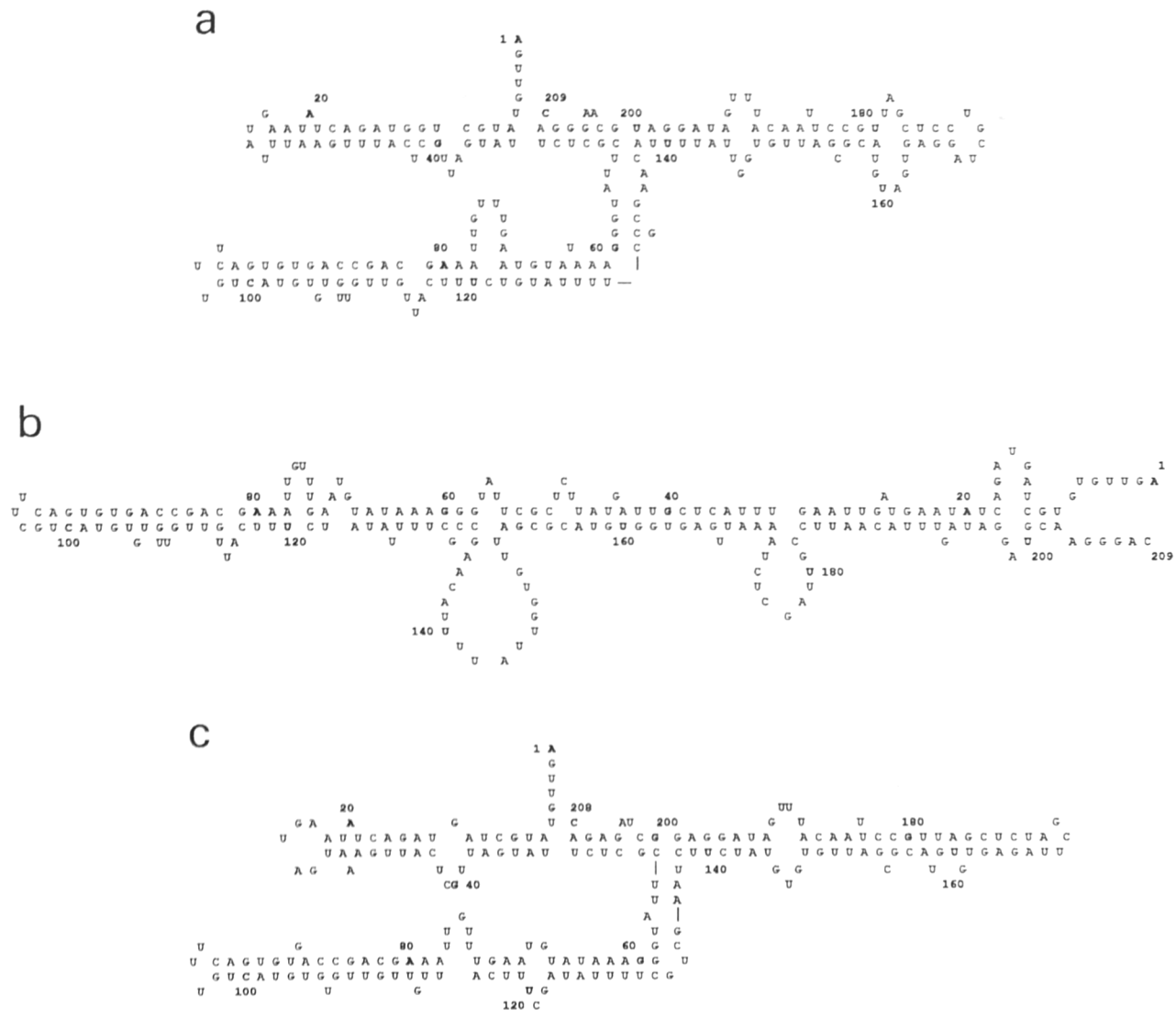


Fig. 3
Potential secondary structure of 3'UTR of TuMV isolates UK1 (a), GK1 (b) and UZB1 (c)
Each 20th nucleotide is in bold.

References

Frenkel MJ, Ward CW, Shukla DD (1989): The use of 3'-noncoding nucleotide sequence in the taxonomy of potyviruses: application to watermelon mosaic virus 2 and soybean mosaic virus. *J. Gen. Virol.* **70,** 2775–2783.

Grumet R, Fang G (1990): cDNA cloning and sequence analysis of the 3'-terminal region of zucchini yellow mosaic virus. *J. Gen. Virol.* **71,** 1619–1622.

Higgins DG, Sharp PM (1988): Clustal - A package for performing multiple sequence alignment on a microcomputer. *Gene* **73,** 237–244.

Jenner C, Walsh JA (1996): Pathotypic variation in turnip mosaic virus with special reference to European isolates. *Plant Pathology* (in press).

Kantrong S, Saunal H, Briand JP, Sako N (1995): A single amino acid substitution at N-terminal region of coat protein of turnip mosaic virus alters antigenicity and aphid transmissibility. *Arch. Virol.* **140,** 453–467.

Kong LJ, Fang RX, Chen ZH, Mang KQ (1990): Molecular cloning and nucleotide sequence of coat protein gene of turnip mosaic virus. *Nucleic Acids Res.* **18,** 5555.

McDonald JG, Hiebert E (1975): Characterization of the capsid and cylindrical inclusion proteins of three strains of turnip mosaic virus. *Virology* **63,** 295–303.

Nakashima H, Sako N, Joh K, Nonaka F (1991): Nucleotide sequences of the coat protein genes of aphid transmissible and non-transmissible isolates of turnip mosaic virus. *Ann. Phytopathol. Soc. Japan* **57,** 549–557.

Nicolas O, Laliberte JF (1992): The complete nucleotide sequence of turnip mosaic potyvirus RNA. *J. Gen. Virol.* **73,** 2785–2793.

Provvidenti R (1982): A destructive disease of garden balsam caused by a strain of turnip mosaic virus. *Plant Disease* **66,** 1076–1077.

Sano Y, Van der Vlugt R, de Haan P, Takahashi A, Kawakami M, Goldbach R, Kojima M (1992): On the variability of the 3' terminal sequence of the turnip mosaic virus genome. *Arch. Virol.* **126,** 231–238.

Shukla DD, Ward CW (1989): Structure of potyvirus coat proteins and its application in the taxonomy of the potyvirus group. *Adv. Vir. Res.* **36,** 273–314.

Tremblay MF, Nicolas O, Sinha RC, Lazure C, Laliberté JF (1990): Sequence of the 3'-terminal region of turnip mosaic virus RNA and the capsid protein gene. *J. Gen. Virol.* **71,** 2769–2772.

Van der Vlugt RAA, Leunissen J, Goldbach R. (1993): Taxonomic relationships between distinct potato virus Y isolates based on detailed comparisons of the viral coat proteins and 3'-nontranslated regions. *Arch. Virol.* **131,** 361–375.

Ward CW (1993): Progress towards a higher taxonomy of viruses. *Res. Virol.* **144**, 419–453.

Ward CW, Shukla DD (1991): Taxonomy of potyviruses: Current problems and some solutions. *Intervirology* **32,** 269–296.

Zuker M, Stiegler P (1981): Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* **9,** 133–148.