Nucleotide sequence of tobamovirus Ob which can spread systemically in N gene tobacco

Ryoji Ikeda,1* Eijiro Watanabe,2 Yuichiro Watanabe1 and Yoshimi Okada1

1 Department of Biosciences, Faculty of Science and Engineering, Teikyo University, To'yosatodai 1-1, Utsunomiya, Tochigi 320 and 2 Department of Agricultural Chemistry, Faculty of Agriculture, University of Tokyo, Yayoi, Bunkyo-ku, Tokyo 113, Japan

The genomic RNA sequence of tobamovirus Ob (Ob), which can spread systemically in tobacco carrying the N gene, was determined. It consists of 6507 nucleotides and contains four open reading frames, exactly corresponding to the genomic organization of tobamoviruses known so far, i.e. encoding the 130K, 180K, 30K and coat proteins. There were no nucleotide overlaps between any open reading frames. The Ob nucleic acid sequence, predicted protein sequences and gene organization were compared with those of other tobamoviruses reported previously. This virus was originally reported as a tomato mosaic virus; however, the nucleotide sequence data given here refute this classification. The determinants that allow tobamovirus Ob to overcome the N gene, a feature peculiar to this virus, were not identified apart from sequence data. This virus should be regarded as a new tobamovirus. The determinants interacting with the tentative N gene product have not yet been analysed.

The complete nucleotide sequences of five tobamoviruses, tobacco mosaic virus vulgare (TMV vulgare; Golet et al., 1982), tomato mosaic virus L strain (ToMV-L; Ohno et al., 1984), tobacco mild green mosaic virus (TMGMV; Solis & Garcia-Arenal, 1990), cucumber green mottle mosaic virus SH strain (CGMMV-SH; Ugaki et al., 1991), pepper mild mottle virus S strain (PMMV-S; Alonso et al., 1991) have been reported so far. The tobamovirus RNA encodes at least four proteins: the 130K, 180K, 30K and coat proteins. The 130K and 180K proteins are involved in RNA replication (Ahliquist et al., 1985; Haseloff et al., 1984; Ishikawa et al., 1986; Kamer & Argos, 1984). The 30K protein has a role in cell-to-cell spread of the virus (Deom et al., 1987; Meshi et al., 1987), and the coat protein is involved in long distance spread of the virus as well as in the assembly of virus particle (Osborn et al., 1990; Otsuki et al., 1977).

Tobamovirus Ob (Ob) was first reported as a disease agent in pepper in Hungary (Csillergy et al., 1983). The virus systemically infects peppers containing the L1 allele of the TMV resistance gene (Csillergy et al., 1983). In tobacco plants that carry the N gene, such as Nicotiana tabacum cv. Xanthi nc and N. glutinosa, all tobamoviruses tested so far are confined to small areas situated around their sites of infection, causing necrotic local lesions but limited spread. In contrast to other tobamoviruses, Ob systemically infects tobacco plants carrying the N gene (Csillergy et al., 1983). In order to study the mechanism involved in these biological properties, we cloned an almost full-length sequence of this virus into cDNA, determined the complete nucleotide sequence of the Ob genome and deduced the encoded protein sequences.

Ob was generously provided as purified virus by Drs S. Fraser and G. Csillergy. The infectivity and symptoms caused by this virus were verified on N. tabacum cv. Xanthi nc. Upper leaves were found to be systemically infected and showed clear dark yellow mosaic symptoms. On the inoculated leaves, necrotic local lesions expanded gradually in diameter, and the whole leaf wilted. In N' plants, such as N. tabacum cv. Bright Yellow, Ob caused normal necrotic local lesions on inoculated leaves, similar to those caused by ToMV-L infection.

The viral RNA was extracted from purified virus by conventional SDS-phenol extraction (Meshi et al., 1982). RNAs were polyadenylated using Escherichia coli poly(A) polymerase (Takara Shuzo Company) and cDNAs were obtained using commercial cDNA synthesis kits. We were able to obtain plasmids carrying cDNAs of over 5000 nucleotides (nt) in length. The cDNA clone

* Corresponding author.
Fig. 1. The entire nucleotide sequence of the Ob genome, shown in DNA form, and with deduced amino acid sequences of the ORFs. Termination codons and an amber readthrough codon are indicated by asterisks. The predicted region of the Oa is underlined.
(pOB10) with the longest cDNA insert, containing nt 21 to nt 5565 of the Ob genome, was obtained from cDNAs using the Amersham kit. The cDNA clone pOBKS2, containing nt 5171 to nt 6507 corresponding to the 3' end portion of the genome, was obtained using the Pharmacia kit. Nested deletion clones were generated from these cDNAs using exonuclease III and were subjected to sequence analysis (Henikoff, 1984). The sequencing of the cDNAs was performed by the dideoxynucleotide chain termination method (Sanger et al., 1977) using a DNA sequencer (Pharmacia). Since pOB10 did not contain the 5' end sequence of the genome, the 5'-terminal region of the genome RNA was directly determined using avian myeloblastosis virus reverse transcriptase (Pharmacia) and priming with a synthetic primer (Meshi et al., 1989) (5'CTCTGACGGTGTCAGACAGC-ACCTGCAATGG 3') which was complementary to nt 89 to nt 121 of the Ob genome. Analysis of the Ob RNA sequence and the deduced amino acid sequences of its encoded proteins was carried out using the Gene Works version 2 program (IntelliGenetics Incorporated).

Ob has the unusual feature of being able to infect tobacco plants carrying the N gene. We first analysed whether or not this strain has any distinguishing features in its genome organization and/or any differences in its encoded proteins.

The complete nucleotide sequence of Ob is shown in Fig. 1. Its length (6507 nt) is longer than that of any other tobamovirus reported so far, mainly due to the long 3' non-coding region (268 nt). The 5' non-coding region of this virus which is positioned upstream of the first open reading frame (ORF), contains a G residue (nt 53) as does that of TMGMV (Solis & Garcia-Arenal, 1990), unlike other tobamoviruses whose genomes have been completely sequenced. The first ORF of Ob RNA extends from nt 69 to nt 3416 and is capable of encoding a protein of 1115 amino acid residues (126K). The readthrough of the amber codon directs the synthesis of a 183K protein, which terminates at nt 4919. This protein is composed of 1616 amino acid residues. The third ORF, which initiates at nt 4927 and terminates at nt 5751, encodes a 31K protein of 274 amino acid residues. The fourth ORF encodes an 18K coat protein, ranging from nt 5754 to nt 6239 (161 amino acid residues). The putative products derived from these ORFs were denoted according to rough estimates of Mr, as 130K, 180K, 30K and coat proteins, respectively, to make it easy to compare them with those of other viruses. No genome organization or amino acid sequence feature could be found that distinguished Ob from other tobamoviruses.

The coding regions for the 130/180K protein, the 30K protein and the coat protein do not overlap with each other as is found to be the case for PMMV-S (Alonso et al., 1991). We calculated the percentage sequence identities between each gene of Ob and those of other tobamoviruses for which genomic sequences have been reported, but could not find one closely related to Ob (data not shown).

Ob was previously classified as tomato mosaic virus Ob strain because it causes local necrotic lesions without systemic infection on N. sylvestris plants containing the N' gene. We calculated the percentage sequence identities between each gene of Ob and those of other tobamoviruses reported, but could not find one closely related to Ob (data not shown).

Ob was previously classified as tomato mosaic virus Ob strain because it causes local necrotic lesions without systemic infection on N. sylvestris plants containing the N' gene. The symptoms caused by ToMV-L in N' plants are identical to those caused by Ob. This suggested that Ob is most closely related to ToMV-L; however, this is not the case (the sequence identity between the 180K protein of ToMV-L and that of Ob is only 67.2%). The symptoms in N' plants are controlled by the coat protein gene in ToMV-L (Saito et al., 1989) and we concluded that there is some sequence identity between the coat proteins of Ob and ToMV-L. When computer analysis was employed, we still could not define which strain Ob is most related to.

A ToMV-L mutant that has the coat protein gene deleted still causes necrotic local lesions on N gene plants (Takamatsu et al., 1987). We do not expect that the coat protein sequence of Ob contributes to its ability to overcome the N gene. As described above, although Ob causes necrotic lesions on N gene plants, it also infects systemically. It seems that this is due to either fast replication or efficient cell-to-cell movement, since Ob easily overcomes the barrier of necrotic local lesions on inoculated leaves. It is possible that either the 130K/180K protein or the 30K protein is the determinant for this ability. Investigations to identify the virus factors responsible are now in progress.

Several tobamoviruses have been classified into subgroup 1 (TMV vulgare, TMV OM strain and ToMV-L) and subgroup 2 [TMV cowpea strain (TMV-Cc) and CGMMV W strain] on the basis of assembly origin (Oas) (Matthews, 1991; Takamatsu et al., 1983). The viruses belonging to subgroup 1 have their Oas in the 30K protein cistron and viruses belonging to subgroup 2 have them in the coat protein cistron. The loci of Ob, PMMV-S, TMGMV and CGMMV-SH have not been identified experimentally. These subgroups were arrived at from data reported by Okada (1986) which were used in a search for Oas on these tobamovirus genomes, based on the conserved sequences of the loop of the Oa, constructed Oa structures and their thermodynamic stabilities. We tentatively assigned the Oa of Ob as being in the 30K protein cistron in the genomes of Ob, TMGMV and PMMV-S and assigned the Oa of CGMMV-SH as being in its coat protein cistron. (Further work is needed to make conclusions experimentally as to the Oa loci of the viruses.) According to...
the Oa loci, Ob, TMGMV and PMMV-S were classified as members of subgroup 1 and CGMMV-SH was classified as a member of subgroup 2.

We found two types of genome structures in subgroup 1 viruses. TMV vulgar, ToMV-L and TMGMV have genome structures in which only the 180K and the 30K ORFs overlap; however, there are no overlaps between any adjacent ORFs in the genomes of Ob and PMMV. In the subgroup 2 genome, all adjacent ORFs are overlapping. The genome structures of seven tobamoviruses are compared and summarized in Fig. 2. We further classified subgroup 1 viruses into subgroup 1a (TMV vulgar, ToMV-L and TMGMV) and subgroup 1b (Ob and PMMV). The amino acid sequences of the putative RNA polymerases (180K protein) are the most conserved of any tobamovirus proteins and so we constructed a phylogenetic tree calculated from them (data not shown). The classification suggested is almost totally consistent with the phylogenetic tree. We therefore propose that genome organization and Oa location are appropriate indices for simple tobamovirus classification.

Note added in proof. Analysis of Ob has been independently performed by Padgett and Beachy [Plant Cell 5, 577-586 (1993)]. There are six differences in the nucleotide sequence between the two reports. Differences at nucleotides 527 and 2687 cause amino acid changes, whereas silent changes are found at nucleotides 3065, 3242 and 5631. Our virus has one extra T base at nucleotide 6380.

This work was supported in part by Grants-in-Aid from the Ministry of Education, Science and Culture, Japan.

References


(Received 9 February 1993; Accepted 4 May 1993)