

GENOMIC ORGANIZATION OF ODONTOGLOSSUM RINGSPOT VIRUS (Cy-1 STRAIN) RNA AND COMPARISON WITH THAT OF KOREAN STRAIN

Masato Ikegami* and Narinobu Inouye

The complete nucleotide sequence of the genomic RNA of odontoglossum ringspot virus Cy-1 strain (ORSV Cy-1) was determined using cloned cDNA. This sequence is 6611 nucleotides long containing four open reading frames, which correspond to 126 K, 183 K, 31 K and 18 K proteins. The 5' non-coding region of ORSV Cy-1 is 62 nucleotides. The ORFs encoded a 126 K polypeptide and a 183 K read-through product in which helicase-sequence and polymerase-sequence motifs are found. The 5' non-coding region, which extends from bases 1 to 62 has 2G residues and the ribosome binding site (AUU). The 3' non-coding region of ORSV Cy-1 composes 414 nucleotides in length. The genomic organization of ORSV Cy-1 is nearly identical to that of ORSV Korean strain (ORSV-K). However, the ORF encoding 183 K protein overlaps the ORF encoding 31 K protein in ORSV Cy-1, but not in ORSV-K. The 183 K read-through product of ORSV Cy-1 is 16 amino acids longer than that of ORSV-K. The homology of the nucleotide sequences of ORSV Cy-1 and ORSV-K is 96%.

Key words: Tobamovirus, Odontoglossum ringspot virus,
Nucleotide sequence, Genome organization

INTRODUCTION

Odontoglossum ringspot virus (ORSV) was first isolated and characterized from *Odontoglossum grande* orchids showing ringspots in the leaves (Jensen and Gold 1951). This virus also causes diamond mottle, mosaic and flower colour breaking in *Cymbidium*, and flower colour breaking in *Cattleya*

Research Institute for Bioresources, Okayama University, Kurashiki 710, Japan
Received December 18, 1995

*NODAI Research Institute, Tokyo University of Agriculture, Tokyo 156, Japan

(Inouye 1966a and b, Isomura *et al.* 1991). The particles of ORSV are rigid rod and contain single-stranded RNA molecules, approximately 6 kb in length, which is a typical morphology of the tobamoviruses (Jensen and Gold 1951). ORSV is easily distinguishable from tobacco mosaic virus (TMV) and other tobamoviruses based on host range, serology, and the nucleotide sequence of 3'-terminal region (Inouye 1966a, Isomura *et al.* 1990, Isomura *et al.* 1991).

We present here the complete nucleotide sequence of the genomic RNA of ORSV (Cy-1 strain) isolated in Japan and compare it to that of the Korean strain of ORSV (ORSV-K).

MATERIALS AND METHODS

Virus, Virus purification, and RNA isolation

ORSV Cy-1 was originally from *Cymbidium* sp. (Inouye 1966a) and was propagated in *Cymbidium* plants in a greenhouse. The virus was purified from *Cymbidium* plants as described by Inouye (1966a), and RNA was extracted by incubation with 1% SDS, followed by phenol/chloroform extraction and subsequent ethanol precipitation (Ikegami *et al.* 1987).

Complementary DNA synthesis and molecular cloning

ORSV Cy-1 RNA was polyadenylated using poly (A) polymerase (Takara Shuzo, Kyoto, Japan) as described by Sippel (1973). DNA complementary to the RNA was synthesized using a cDNA kit (purchased from Pharmacia, Uppsala, Sweden) according to the manufacturer's specifications. Synthesized double-stranded cDNA was inserted into the *EcoRI* site of pUC119 using *EcoRI* linkers and transformed into *Escherichia coli* JM109 made competent. Transformants carrying plasmids with ORSV Cy-1 cDNA inserts were identified by colony hybridization using single-stranded cDNA synthesized from ORSV Cy-1 RNA.

Nucleotide sequence determination and analysis

Nucleotide (nt) sequence analysis was carried out by the enzymic dideoxynucleotide chain termination sequencing method with a DNA sequencer (model 370A, Applied Biosystems, Foster, USA).

ORSV Cy-1 RNA was decapped with NaIO_4 and aniline (Dasgupta *et al.* 1976), the 5' end was labeled with [γ - ^{32}P] ATP by polynucleotide kinase (PNK), and subjected to direct enzymic sequencing (García-Arenal *et al.* 1987). The sequence of the 5' end of the RNA was further established by the deoxynucleotide chain termination method on the RNA (Gould and Symons 1989) by priming with a 22-base oligonucleotide complementary to nt 81 to 102.

Analysis and comparison of the ORSV Cy-1 RNA sequence, and of the deduced amino acid sequences of the encoded proteins, were carried out using the program DNASIS.

RESULTS AND DISCUSSION

Fig. 1 shows the complete sequence of ORSV Cy-1 RNA as determined from the cDNA clones. The genomic organization of ORSV Cy-1 was compared to that of ORSV-K (GenBANK, accession number ; X82130). The genomic RNA of ORSV Cy-1 is 6611 nt long, 7 nt shorter than that of ORSV-K. The homology of the nucleotide sequences of ORSV Cy-1 and ORSV-K is 96%. Computer analysis of the complete sequence of ORSV Cy-1 reveals four open reading frames (ORF) in the (+) strand. The first AUG initiation codon was found at residues 63 to 65, which starts an ORF encoding a protein composed of 1111 amino acids (126 K). This ORF terminates at an amber codon, UAG, positioned at residues 3399 to 3401. The ORF encoding the read-through protein composed of 1611 amino acids (183 K) terminates at residues 3999 to 4901. The terminal 23 bases of the ORF encoding the 183 K protein overlap the ORF encoding the 31 K protein, which terminates with UGA. The ORF encoding the 18 K protein initiates 2 bases upstream from the terminal codon for the 31 K protein. This terminal ORF is the coat protein gene, as determined by expression in *E. coli* (Isomura *et al.* 1991). The coat protein gene terminates at residues 6195 to 6197. The coat protein is 157 amino acids. Such genomic organization is similar to that of ORSV-K (Fig. 2). However, the ORF encoding 183 K protein does not overlap the ORF encoding 31 K protein in ORSV-K. The 5' non-coding region of ORSV Cy-1 has 62 nt, 7 nt shorter than that of ORSV-K.

The nucleotide sequence of 5' non-coding region of ORSV Cy-1 has 85 to 90% similarity to that of ORSV-K. The 5' non-coding regions of both strains of ORSV, which extends from base 1 to 62, have 2 G residues, at positions 6 and 10, but the 5' non-coding region of TMV-V is characteristically free of the G residue (Mandele 1968). The 5' non-coding regions of TMV-L (Ohno *et al.* 1984), pepper mild mottle tobamovirus (PMMV) (Avilá-Rincon *et al.* 1989), and CGMMV (Ugaki *et al.* 1991) are devoid of the G residue. The deletion of the 5'-terminal 8 nt (GUAUUUUU) of TMV-L caused the complete loss of infectivity and the other sequence of the region seemed not to be crucial, although a large deletion had a large effect (Tomenius *et al.* 1987). This region is also conserved in ORSV Cy-1 and ORSV-K, although it has a G residue at position 6. However, the role of the sequence of 8 nt is not known. The ribosome-binding site is thought to be AUU (Tyc *et al.*

ORSV genome sequence

↗→126k, 183k

```

GUUUUUGUCGUAUUCACAAUUAACAACAAACAAUUAACAACAAUUUCACAAUUCACAUCAACAUGGCACACUCCCAACAAACAAUGAACAAACAAAGUUAUUGAGGCGUGUAUGGGCAGAA 120
          A H F Q Q T M N N K V I E A G M G R 18
ACAGCCUCAUUAACGAUUUGCCUCAAGACGUGUUACGAUACCGCUGUCGAGAGUUAACACCCGUCUAGACGCCAAAGGUAACUUUCCAAGGUAUUUCAGGAACAGAUUA 240
N S L I N D L A Q R R V Y D N A V E E L N H R S R R P K V N F S K V I S Q E Q I 58
UUCAAAGCUACUAAUGCGUACCUCGUAUUUGAAUACCCUUUUAACAACUCACCGCUGUCACUAGCGGGGAGGUCUUCAGGACUCAGGCUAGAUUUUGAUGCAAUCC 360
I Q A T N A Y P E F E I T F Y N T Q L A V H S M A G G L R A L E L E Y L M M Q I 98
CAUUUGGUUUAUUAUUGAUGACUCCGGGAAACUUUUCAGCCACUUUAUAAAGGACGAGUUAACGUCCUUAUGGCUAUGCCUAAUUGAUAUUGCGGAGUGGCGUCGUAUCA 480
P F G S I T Y D I G G N F S A H L Y K G R D Y V H C C M P N L D I R D V A R H I 138
AUCACAAAGAUACUGUUUCUUAUCUGGCUAGGCGGAGAGAAAGAGGCGUCCGCUUUUCAGCAUUCAGCUUUUUAUAAUUAUGAUAUGACCUGAUGCUGUGCGUCG 600
N Q Q D T V S T Y L A R L E R S K R G L P V F Q Q S A F N K Y M N D P D A V C C 178
AUAAAAGGUUUAACAGACUGUUCUUAUCUCCGUAGUUCUCCGGAAACAAUACCGCUGUGCCGUUCUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 720
D K R F Q D C S Y S V D L P G K T Y A V A L H S I Y D I P A D E F G A A L L R K 218
AUGUUCACAUUAUUAUUGGCUUUUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 840
D V H I C Y A A F H F S E N L L L E T T S A P L D E I G A T F Y K S G D R L S F 258
UCUUUCAAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 960
F F Q N E S T L N Y E H S Y K N V I K Y V C K T F F P A S N R F V Y H K E F M C 298
CCAGAUUAACACUAGGUUUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 1080
T R V N T W F C K F T K V D T Y F L F R G V Y T R G E D S E Q F Y T A M D E A W 338
AGUAUUAAGAAACUUGGCAUUGCUUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 1200
E Y K K T L A N L K C E R T I F R D R A A V N F W F P K V K D M V I V P L F D G 378
CCGUCACUUCAGGAAAGUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 1320
S V T S G K M K R S E V M V N K D F V Y T V L N H I R T Y Q D K A L T Y K N V L 418
CCUUUUGCGAGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 1440
S F U V E S I R S R V I I N G V S A R S E W D V D K S V L Q A L S M T F L L Q T K 458
UGGCUAGAACAAAGUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 1560
L A E A K D Q V V L K K F Q K F D D T V T N L F W K Q I S D A V G D L F P S I K 498
AAACGCUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 1680
E T L I S G G F V K V A E Q S L Q I K T P D E Y I T F A D K L V M E Y K A T E E 538
UUCAACUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 1800
L Q H L D I S K P L E R A E K Y Y N A L S E L S V L K E C D E F D I T Q F K N L 578
GUGAAGAAAGGACUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 1920
C E E K D I D P D V V A K V I V A I M K N E L T L P F K N P T P E A L S D A L S 618
CGUUAACCAAGACUUGGAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 2040
P L P K D L D M R F D L L K L S T C A P F P S V K T L D S G L L P K Q S Y G D E 658
GACAAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 2160
R Q F E S Q S V S V S D F H L K S V S V E S V K M K S M S S A V Y T G P L K V Q Q 698
UGAAGAACUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 2280
M K N Y M D Y L S A S I S A T V S N L C K V L K D V Y G A D P E S A E K S G V Y 738
AUGUAGUGAAGGCAAUGGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 2400
D V V K G K W L L K P K G K C H A W G V A E L N N G E K V I V L L E W A D G F P 778
UUUGUGGUUAUUGGACGAGUGGCGGUCUCUUCAGAUUUCCUUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 2520
I C G D . W R R V A V S S D S L I Y S D M G K L Q T L L S C L K D G E P V P S D A 818
AAGUCACGUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 2640
K V T L V D G V P G C G K T K E I L E T V N F D E D L I L V P G K E A C K M I I 858
    
```

Fig. 1-1.

IKEGAMI and INOUE

AGAGAGCUAAUAAGUCUGGACAUGAAGCGCAAAAGGAUAUGUAAGAACUGGGUUCGUUUUAUUGCAUUGAAGCCUAAAGACUUUAACAAACUGUUUACGAUGGGGUUGA 2760
K R A N K S G H V R A T K D N V R T V D S F L M H L K P K T Y N K L F I D E G L 898

UGUUAACUACUGGUUGUGUUAACUUUCUGAUGCAUUGCCCAUUGUCGUGAAGCAUUGUUUUUGGUGAUAUGAACAACUUCCUUUAUAACAGAGUGGCAAAUUCUUUACCGGA 2880
M L H T G C V N F L I A L S H C R E A M V F G D T E Q I P F I N R V A N F P Y P 938

AACAUUUUGCCACACUUGUUUAUGAUCUAGGGAGGUUAGGGUUUAUCUUAAGAUGCCCGCAGUGUUUACCAUUUUUAUGAUCUCCAAUAUUGCGGGAAGUUUAUGUACGAAUG 3000
K H F A T L V Y D H R E V R R L S L R C P A D V T H F M N S K Y D G K V L C T N 978

AUGUGAUUGCCUCAGUUGAUGUGAUGGUGUAGGGGAAAGGGUUUUUAACCCGAAAAGCAAAACCGCUUAAAGGAAAUAUACCUUCAUCUAAAUGACAAAGCUGAACUGAAGG 3120
D V I R S V D A E V V R G K G V F N P K S K P L K G K I I T F T Q S D K A E L K 1018

AGCGUGUUAUGAAGAAGUUUACUUUUGUGAAUUAACCCGUACACGAAUCCAAGGUGAGACCUUUGAAGACGUUUGUUGUUGACACCAACUCUUGGAGCUUUAU 3240
E R G Y E E V S T F G E I N T V H E I Q G E T F E D V S V V R L T P T P L E L I 1058

CAAAGAGUCCCGCAGCUUUGUGCAUUGCAAGACAUAUAUUAACCUUGUUCUUGAUCUUCUGCUGCAAGUUUUGCUGAUUUUAAGCAAGGUGUCAGACU 3360
S K S S P H V L V A L T R H T K S F K Y Y S V V L D P L V K V C S D L S K V S D 1098

UUUAUCUGGAUUGUUAAGGUGAUGCCCGGAUCUUUAAGCAUUAAGCAUUAAGGAGUUAUCUUAAGCGAAAUAUUGUUAUUGCAUUGCUUUAUUGCUUUAUUGCUUUAUUGC 3480
F I L D M Y K V D A G I L * Q L Q V G S I F K G E N L F V P C P K S G Y I S D M 1138

AUUUUUAUUGACAUUUGUUAUGGAAACAGCAUUAUUUAAGCAUUGAUGCAUGAUGAUCUUGGUGAAAUAUUAUUGUUAUUGCAAGGUAUGCAUUGAUGAUUUUUA 3600
Q F Y Y D T L L P G N S T I L N E Y D A V T M N L R E N N L N V K D C T I D F S 1178

AUCUGUUAUGUCCCGAGACAACAAGAGUUUUAACACCAUGUUAUUGAUCUGUUGAUCUUCACGUCGUGGUCUUCUUGAAGAAACCUUUGGCAUUGAUAUUAAGAAAU 3720
K S V S V P R Q Q Q E F F T P V I R T A A E R P R S R G L L E N L V A M I K R N 1218

UUAACUCUCCAGAUUAUAGCGGUAUUAUGAUAUGAUAUCUGCCGAUCUUGUUAUUAAGUUUUGGUAUCUUAUUGAUAUUGAUAUUGAUAUUGAUAUUGAUAUUGAUAUUG 3840
F N S P D L T G I L D T A E L V V N K F W D A Y I I D E L S G G A N V T P M 1258

CUUCAGUUGCCUUUCAGGUGGUAUGGCUAAGCAGGAGAAAGUAUCUUAUGGUCAUUAUGCUGAUUUUGAUCUUGUUAUUGCCGCUUUAUGAUAUUAACAACAUUGAUUAAGGCGC 3960
T S D A F H R W M A K Q E K S T I G Q L A D F V D F V D L P A I D Q Y K H M I K A 1298

AACCUAAACAGAAACUUGAUCUCAGUCCUACAAUGAGUAUGGCUUUCAGACUUAUUGUUUAUCACUCAAAGCAGAUAAUGCAUUUUGGCUUUCUUAUUGCAGAUUAACUAGAC 4080
Q P K Q K L D L S P Q D E Y A A L Q T I V Y H S K Q I N A I F G P L F S E L T R 1338

AGCUUUUAGAGAGGUAAGACAGUAGCAAAUUUUUUUAUCACCCGGAAAACGCCGAAACAGAUUAGAGUUUUUCAGAUUUGAUAUUGAAGGCUUUGAAGGUAUUGAAGU 4200
Q L L E R I D S S K F L F Y T R K T P E Q I E E F F S D L D S T V P M E V L E L 1378

ACAUUUAAGUAGCAAAUCCAGAAAGAUUUCUUGUUGUUGAGUUAUUGUUGGGAAGAAACUUGGCUUUAUUGGCUUUCUGAAGAAUUGGAAACAGGGAUUAAGGAAGA 4320
D I S K Y D K S Q N E F H C A V E Y L I W E K L G L N G F L E E V W K Q G H R K 1418

CAUCCUUGAAGAUUACACCGCAGGAUUAAGACUUGUUUAUGGUAUCAAAGGAAAAGCGGAGUUAACUACUUUAUUGGCAUAUAGCAGUUAUUAUUGCUGUUGUUGGUGGUA 4440
T S L K D Y T A G I K T C L W Y Q R K S G D V T T F I G N T V I I A A C L A S M 1458

UACCUAUGGAUAGGAAUUAUUAAGCUGGUUUUGUUGGAGUAGUUAUGUUA 4560
I P M D K V I K A A F C G D D S M L Y I P K G L D L P D I Q S G A N L M W N F E 1498

CAAAUUAUUAUGUAGGCUUACCGAUUUUUUGCGGAGGUUA 4680
A K L Y R K R Y G Y F C G R V I I H H D R G A I V Y Y D P V K L I S K L G C K H 1538

UCAAAUCUUGGUAUUAUUGGAAAGAUUA 4800
I K S L D H L E E F R I S L C D V S A S L N N C A Y Y G Q L N D A I A E V H K T 1578

CAGUAUUGGUGUUGUUGGUUUUGUUAUUGUUA 4920
A V N G S F A F C S I V K Y L S D K N L F R T L F Y N G S S T K G * 1611

I N L S A S E K L L P S A L T A V K S V R I S K Y D K I I S Y E N D T L S D I D 53
AUUAUUUAUUAUGGCUUUA 5040

L L K G V K L V E N G Y V C L A G L V V T G E W N L P D N C K G G V S I C L V D 93
UUGCUUUAAGGUGAAGUUAUGAUGAAGUUGGUAUUGUUGCCUUGCGGCUUGCUGCAUGGUAUUGAUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUA 5160

K R M K R A N E A T L G S Y H T S A C K K R F T F K I I P N Y S V T T A D A L K 113
AAACGUAUUAAGGCUUUA 5280

Fig. 1-2.

ORSV genome sequence

```

G I W Q V M T N I R G V E M E K G F C P L S L E F V S I C V V Y L N N I K L G L 153
GGAAUUGCCAAGUUAUGACAAAUUAGAGGGUGGAAAGGAAAGGGUUUUUGCCCAUUAUCUUGGAGUUGUUCUUAUUGCGUGUGUAUUUAAAUAUAAAACUCGGACUG 5400

R E K I L N V T E G G P T E L T E A V V D E F V E K Y P M A A R L K S F R S V N 193
AGAGAGAAGAUUUUGAACGUAACAGAAAGGAGCCACCCGAACUUCAGAAAGCAGUUGUUGAUGAUCUGCGGAGAAAGUUCUUAUGCGUCUAGGCUCAAGUCUUUUUCGUCUGUCAAC 5520

K K K P S N S S K F V N G K S R L N S R N K L N Y E N G D S D V G I S V V D D I 233
AAAAAGAAACCUAGUAACAGUUCUAAAGUUGUUAUUGGAAAUUCUAGACUUAUAGUCGUAUAAAUAAUAAUUGAAAUGGUAUGGUAUGUUGCUUAUUGUUGUUGAUGAUUU 5640

V V G N G V S D I R I D D D C E S F D A Q S D S Y *
GUGGUUGGUAUUGGUGUAUGAUAUUCGUAUUGAUGAUGUGGAGUCAUUUGACGCACAUAUCGUAUUGAUAUUGAAUUGCUUACACUUAUUCAGACCCGUCUAAAGCUGGCUUAU 5760
                                     31k←-
                                     S Y T I T D P S K L A Y 12
                                     ↳18k

UAAGCUCGGCUUGGGCUGACCCCAAUUCACUGAUAACCUUUUACCAAUUCUUGGUAUUCAGUUAUUCACACAACAAGCUCGAAACAACUGUUCUAAACAGCAGUUGUCGUAUGUUGGC 5880
L S S A W A D P N S L I N L C T N S L G N Q F Q T Q Q A R T T V Q Q Q P A D V W 52

AGCCGUGUUCUACUUGACCAUGGUCUAGGUUCCGUCGACGCCUGGUUACUUCAGAGUUAUUCGCUUAGAUCCUUAUUAUAGAUCCUUAUAAUUAUUAUUGGUAUUCUUGUAUCUCGUA 6000
Q P V P T L T S R F P A G A G Y F R V Y R Y D P I L D P L I T F L M G T F D T R 92

AUAGAUAUUCGAGGUAGAAAUCGCCAGAAUCCGACAACUACGGAAACAUUAGAUCCAUCGUAUGAGUUAUGAUGCAACUGAUGCAUAUAGAAUUGCUUUAUUAUUAUUAUUAUUAU 6120
N R I I E V E N P Q N P T T T E T L D A T R R V D D A T V A I R S A I N N L L N 132

AGUAGUUAUGGGAAACUGGUUAUGUACAUAUCAGUCUUAUUGGAGACGAUAUUCGGACUUAUUGGACCUUCCUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 6240
E L V R G T G M Y N Q V S F E T I S G L T W T S S *
                                     18k←-

GUGCAUACGAUUAUGCAUUGUGUUAUCCUCCAUAAAUCGAAGGDUUUUACUGCAGAUUUGAUGGUUUCUACGGUGAAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 6360
AUAGUGUUUAUCCUCCAUAAAUCGAAGGDUUUUUGUCAGACGCCUGUAAGGAGUGGUCACCUUAACGACACAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 6480
UGUCCUCCAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 6600
GUAGAGGCCCA 6611
    
```

Fig. 1-3.

Fig. 1. Nucleotide sequence of ORSV Cy-1 RNA and encoded amino acid sequences for each ORF. ATG initiation codons are underlined, and approximate molecular weight of the translation product is indicated. Asterisks indicate termination codons.

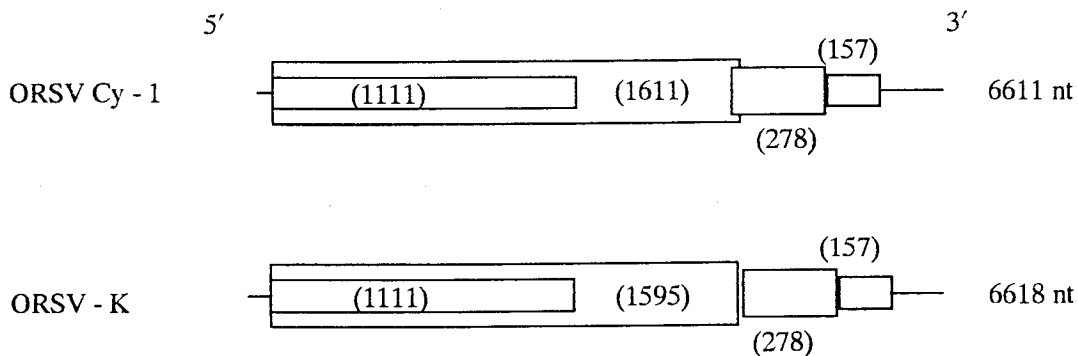


Fig. 2. Comparison of ORFs of ORSV Cy-1 and ORSV-K deduced from the nucleotide sequences. Figures in parentheses present the number of amino acid residues. Data from GenBANK, accession number : X82130 (ORSV-K).

1984), which is present from base 20 to 22 in both strains.

We previously reported on the 3' non-coding region of ORSV Cy-1 (Isomura *et al.* 1990). The 3' non-coding region of ORSV Cy-1 RNA is different from those of the other tobamoviruses. The 3' non-coding region of ORSV Cy-1 RNA is composed of 414 nt, but the 3' non-coding regions of TMV-V, TMV-L, TMGMV, and CGMMV are 204, 202, 210, and 176 nt, respectively (Goelet *et al.* 1982, Ohno *et al.* 1984, Solis and García-Arenal 1990, Ugaki *et al.* 1991). ORSV Cy-1 has 3 repeated domains, each of 35 nt, which were very similar. Such repeated sequences are not present in the other tobamoviruses. It is unknown if all 3 copies within the ORSV Cy-1 sequence are required for replication of ORSV Cy-1 RNA. ORSV-K has also 3 repeated domains, each of 35 nucleotides, which were very similar (Ryu *et al.* 1995).

The 31 K cell-to-cell movement protein of ORSV Cy-1 is the same number of amino acid residues as that of ORSV-K (Table 1). The ORSV Cy-1 31 K protein has 98% amino acid sequence similarity to that of ORSV-K. We previously reported the sequences of 31 K and 18 K protein genes (Isomura *et al.* 1990, Isomura *et al.* 1991). The putative 126 K and 183 K proteins correspond to the 130 K and 180 K proteins of TMV-V and TMV-L, respectively (Goelet *et al.* 1982, Ohno *et al.* 1984). The former protein of ORSV Cy-1 has 1111 amino acids, which is the same number as that of ORSV-K. The 183 K read-through product of ORSV Cy-1 is 16 amino acids longer than that of ORSV-K. The amino acid sequence identity of the ORSV Cy-1 183 K and 126 K proteins is approximately 94% and 94% with the corresponding 183 K and 126 K proteins of ORSV-K.

Table 1. Genome organization and sequence homology between ORSV CY-1 and K

	ORSV Cy-1	ORSV K	Sequence homology (%)	
			Total	Alignment
Total genome size	6611nt	6618nt	96	96
5' non-coding region	1-62 (62nt)	1-69 (69nt)	85	90
126 K protein cistron (No. of amino acid residues)	63-3401 (1111a.a.)	70-3408 (1111a.a.)	97 (94)	97 (94)
183 K protein cistron (No. of amino acid residues)	63-4901 (1611a.a.)	70-4858 (1595a.a.)	94 (93)	97 (94)
31 K protein cistron (No. of amino acid residues)	4879-5718 (278a.a.)	4886-5725 (278a.a.)	98 (98)	98 (98)
18 K protein cistron (No. of amino acid residues)	5721-6197 (157a.a.)	5728-6204 (157a.a.)	97 (95)	97 (95)
3' non-coding region	6198-6611 (414nt)	6205-6618 (414nt)	97	97

ORSV genome sequence

(a)

```

824      I      883
ORSV Cy-1  V D G V P G C G K T K E I L E T V N F D E D L I L V P G K E A C K M I I K R A N K S G H V R A T K D N V R T V D S F L M
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
824      883

884      II      III      943
          H L K P K T Y N K L F I D E G L M L H T G C V N F L I A L S H C R E A M V F G D T E Q I P F I N R V A N F P Y P K H F A
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
          H L K P K T Y N K L F I D E G L M L H T G C V N F L I A L S H C R E A M V F G D T E Q I P F I N R V A N F P Y P K H F G
884      943

944      IV      1003
          T L V Y D H R E V R R L S L R C P A D V T H F M N S K Y D G K V L C T N D V I R S V D A E V V R G K G V F N P K S K P L
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
          H T C L H R R E V R R L S L R C P A D V T H F M N S K Y D G K F L C T N D V I R S V D A E V V R G K G V F N P K S K P L
944      1003

1004      V      1063
          K G K I I T F T O S D K A E L K E R G Y E E V S T F G E I N T V H E I O G E T F E D V S V V R L T P T P L E L I S K S S
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
          K G K I I T F T O S D K A E L N E R G Y E E V S T F G E I N T V H E I O G E T F E D V S V V R L T P T A L E L I S K S S
1004      1063

1064      VI      1080
          P H V L V A L T R H T K S F K Y Y
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
          P H V L V A L T R H T K S F K Y Y
1064      1080
    
```

(b)

```

1376      I      1435
ORSV Cy-1  L E L D I S K Y D K S O N E F H C A V E Y L I W E K L G L N G F L E E V W K Q G H R K T S L K D Y T A G I K T C L W Y O
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1376      1435

1436      II      III      1495
          R K S G D V T T F I G N T V I I A A C L A S M I P M D K V I K A A F C G D D S M L Y I P K G L D L P D I Q S G A N L M W
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
          R K S G D V T T F I G N T V I I A A C L A S M I P M D K V I K A A F C G D D S I L D I P K G L D L P D I Q S E A N L M W
1436      1495

1496      IV      1515
          N F E A K L Y R K R Y G Y F C G R Y I I
          : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
          N F E A K L Y R K R Y G Y F C A R Y I I
1496      1515
    
```

Fig. 3. Comparison of partial amino acid sequences of the 183 K protein of ORSV Cy-1 with those of the 183 K protein of ORSV-K (Goelet *et al.* 1982). These sequences contain motifs indicated by Habili and Symons (1989). (a) The region contains nucleic acid helicase motifs; (b) the region contains RNA polymerase motifs. Motifs are indicated by solid lines.

Habili and Symons (1989) showed that positive-strand viruses could be grouped based on amino acid sequence motifs of nucleic acid helicases and RNA polymerases, and proposed a new luteovirus supergroup. Within the 126 K protein of ORSV Cy-1 there are six areas of sequential amino acids associated with proteins having helicase activity (Fig. 3). These regions show a high level of sequence conservation within ORSV strains, Regions II, III, IV, V, and VI have complete sequence conservation between ORSV Cy-1 and ORSV-K, whereas region I have 1 conservative amino acid difference. Within the 183 K protein gene region there are four areas of conserved amino acids associated with putative viral replicase proteins (Fig. 3). These four regions show a high level of sequence conservation within ORSV strains. Region II has complete sequence conservation between ORSV Cy-1 and ORSV-K, whereas regions I, III and IV have 1, 2, and 1 conservative amino acid difference, respectively.

REFERENCES

- Avilá-Rincon, M. J., Ferrera, M. L., Alonso, E., Garcia-Luque, I. and Diaz-Ruiz, J. R. 1989. Nucleotide sequences of 5' and 3' non-coding regions of pepper mild mottle virus strain S RNA. *J. Gen. Virol.* 70 : 3025-3031.
- Dasgupta, R., Harada, F. and Kaesberg, P. 1976. Blocked 5' termini brome mosaic virus RNA. *J. Virol.* 18 : 260-267.
- García-Arenal, F., Zaitlin, M. and Palukaitis, P. 1987. Nucleotide sequence analysis of six satellite RNAs of cucumber mosaic virus: Primary sequence and secondary structure alterations do not correlate with differences in pathology. *Virology* 158 : 339-347
- Goelet, P., Lomonosoff, G. P., Butler, P. J. G., Akam, M. E., Gait, M. J. and Karn, J. 1982. Nucleotide sequence of tobacco mosaic virus RNA. *Proc. Natl. Acad. Sci. U. S. A.* 79 : 5818-5822.
- Gould, A. R. and Symons, R. H. 1982. Cucumber mosaic virus RNA3: Determination of the nucleotide sequence provides the amino acid sequences of protein 3A and viral coat protein. *Eur. J. Biochem.* 126 : 217-226.
- Habili, N. and Symons, R. H. 1989. Evolutionary relationship between luteoviruses and other RNA plant viruses based on sequence motifs in their putative RNA polymerases and nucleic acid helicases. *Nucleic Acids Res.* 17 : 9543-9555.
- Ikegami, M., Osaki, T. and Inouye, T. 1987. Single-stranded DNA in tobacco leaf curl virus. *Ann. Phytopathol. Soc. Japan.* 53 : 269-273.
- Inouye, N. 1966a. A virus disease of *Cymbidium* caused by odontoglossum ringspot virus in Japan. *Ber. Ohara Inst. landw. Biol. Okayama Univ.* 13 : 149-159.
- Inouye, N. 1966b. Virus diseases of orchids. IV. Symptoms of viruses in *Cattleya*. *Japan Orchid Soc. Bull.* 12 : 2-5 (in Japanese).
- Isomura, Y., Matumoto, Y., Murayama, A., Chatani, M., Inouye, N. and Ikegami, M. 1990. Nucleotide

ORSV genome sequence

- sequence of cell-to-cell transport protein gene of odontoglossum ringspot virus. *Nucleic Acids Res.* 18 : 7448.
- Isomura, Y., Matumoto, Y., Murayama, A., Chatani, M., Inouye, N. and Ikegami, M. 1991. Molecular cloning, sequencing and expression in *Escherichia coli* of the odontoglossum ringspot virus coat protein gene. *J. Gen. Virol.* 72 : 2247-2249.
- Jensen, D. D., and Gold, H. A. 1951. A virus ring spot of *Odontoglossum* orchid : symptoms, transmission and electron microscopy. *Phytopathology* 41 : 648-653.
- Mandele, S. 1968. Localization of unique sequence in tobacco mosaic virus ribonucleic acid. *J. Biol. Chem.* 243 : 3671-3674.
- Ohno, T., Aoyagi, M., Yamanashi, Y., Saito, H., Ikawa, S., Meshi, T. and Okada, Y. 1984. Nucleotide sequence of the tobacco mosaic virus (tomato strain) genome and comparison with the common strain genome. *J. Biochem.* 96 : 1915-1923.
- Ryu, K. H., Choi, J. K. and Park, W. M. 1995. Analysis of the coat protein gene of odontoglossum ringspot virus-Cy (Korean isolate). *Mol. Cells* 5 : 181-186.
- Sippel, A. E. 1973. Purification and characterization of adenosine triphosphate : Ribonucleic acid adenytransferase from *Escherichia coli*. *Eur. J. Biochem.* 37 : 31-40.
- Solis, I. and Garcia-Arenal, F. 1990. The complete nucleotide sequence of the genomic RNA of the tobamovirus tobacco mild green mosaic virus. *Virology* 177 : 553-558.
- Tomenius, K., Clapham, D. and Meshi, T. 1987. Localization by immunogold cytochemistry of the virus-coded 30 K protein in plasmodesmata of leaves infected with tobacco mosaic virus. *Virology* 160 : 363-371.
- Tyc, K., Kornaska, M., Gross, H. J. and Filipowicz, W. 1984. Multiple ribosome binding to the 5'-terminal leader sequence of tobacco mosaic virus RNA, assembly of the 80S ribosome mRNA complex at the AUU codon. *Eur. J. Biochem.* 140 : 503-511.
- Ugaki, M., Tomiyama, M., Kakutani, T., Hidaka, S., Kiguchi, T., Nagata, R., Sato, T., Motoyoshi, F. and Nishiguchi, M. 1991. The complete nucleotide sequence of cucumber green mottle mosaic virus (SH strain) genomic RNA. *J. Gen. Virol.* 72 : 1487-1495.

オドントグロッサムリングスポットウイルス Cy-1 株 RNA のゲノム構成および韓国株との比較

池上正人・井上成信

摘 要

オドントグロッサムリングスポットウイルス Cy-1 株 (ORSV Cy-1) の cDNA クローンを用いて完全長ゲノム RNA の塩基配列を決定した。この配列の長さは 6611 塩基であり、126 K, 183 K, 31 K および 18 K のタンパク質に対応する 4 個のオープンリーディングフレーム (ORF) を含んでいた。ORSV Cy-1 の 5' 非翻訳領域は 62 塩基であった。126 K および、その 183 K リードスルー産物をコードする ORF 内にはヘリカーゼとポリメラーゼの塩基配列モチーフが含まれていた。5' 非翻訳領域である 1 から 62 塩基目の領域には 2 個の G 残基とリボソーム結合領域 (AUU) を保持していた。ORSV Cy-1 の 3' 非翻訳領域は 414 塩基長であった。ORSV Cy-1 の遺伝子構成は、韓国株 (ORSV-K) と殆ど同じであった。しかしながら、ORSV Cy-1 の 183 K タンパク質をコードする ORF は 31 K タンパク質をコードしている ORF と一部重なっていたが、ORSV-K では重なっていなかった。ORSV Cy-1 および ORSV-K の 183 K タンパク質のアミノ酸数を比較すると、ORSV Cy-1 の方が 16 アミノ酸長かった。ORSV Cy-1 と ORSV-K の塩基配列の相同性は 96% であった。

キーワード：トバモウイルス, オドントグロッサムリングスポットウイルス,
塩基配列, ゲノム構造