Arch Virol (2002) 147: 2215–2224
DOI 10.1007/s00705-002-0875-6

Archives of Virology
Printed in Austria

Nucleotide sequence, genome organisation and phylogenetic analysis of Indian citrus ringspot virus

Brief Report

G. Rustici[*], **R. G. Milne,** and **G. P. Accotto**

Istituto di Virologia Vegetale, Consiglio Nazionale delle Ricerche, Torino, Italy

Received February 26, 2002; accepted June 25, 2002
Published online September 12, 2002

Summary. The sequence of the single-stranded RNA genome of Indian citrus ringspot virus (ICRSV) consists of 7560 nucleotides. It contains six open reading frames (ORFs) which encode putative proteins of 187.3, 25, 12, 6.4, 34 and 23 kDa respectively. ORF1 encodes a polypeptide that contains all the elements of a replicase; ORFs 2, 3 and 4 compose a triple-gene block; ORF5 encodes the capsid protein; the function of ORF6 is unknown. Phylogenetic analysis of the complete genome and each ORF separately, and database searches indicate that ICRSV, though showing some similarities to potexviruses, is significantly different, as in the presence of ORF6, the genome and CP sizes, and particle morphology. These differences favour its inclusion in a new virus genus.

*

Indian citrus ringspot virus (ICRSV) is associated with a serious disease of citrus, especially Kinnow mandarin, in India [10]. The virus has flexuous filamentous particles 650 nm in length, with an ssRNA genome of about 7500 nucleotides and a capsid protein (CP) of 34 kDa. Two open reading frames (ORFs) have been identified at the 3′-end of the genome, encoding the CP and a 23 kDa polypeptide of unknown function [11].

Here we report the complete sequence and genome organisation of an isolate (K1) of ICRSV, the same as used previously [11], and the results of phylogenetic analysis of the putative viral proteins. The virus was propagated in *Phaseolus vulgaris* cv. Saxa; purification and viral RNA extraction was as described [11].

[*]Present address: The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, U.K.

[**]The GenBank accession number of the sequence reported in this paper is AF406744.

A step by step walking strategy was used to clone the region of the genome upstream of the 3′-end previously sequenced (pICR7; [11]). The cDNA extending past the known sequence was synthesised using the ICRSV-specific primer ICR-CLN(-) (5′-AGT GCA GCG GCA AAT GTG AG-3′) and the Universal Riboclone cDNA Synthesis System (Promega) according to the manufacturer's instructions. Double-stranded cDNA was cloned in the *Eco*RI site of pBLUESCRIPT KS+ plasmid, and recombinant plasmids were used to transform competent *E. coli* DH5α cells. cDNA clone pICR10 was sequenced using both universal and virus-specific primers. This served as the basis for the next steps of walking along the genome, performed using the primers ICR-CLN2(-) (5′-ATG TTC GCG TCC TGT GAT TG-3′) and ICR-CLN3(-) (5′-ACG TTG CCG CGT TCA GTG-3′). An additional clone (pICR23), corresponding to an upstream region of RNA later uncovered, was obtained by RT-PCR using the specific primers ICR19-1950FW (5′-ACA TGC CTC GCG CAA ACT C-3′) and ICR18-990FW (5′-GTC CCA TAT CAT TGA GTG CG-3′). PCR-amplified fragments were purified and then cloned using the PCR-Script Amp cloning kit (Stratagene), according to the manufacturer's instructions.

For determining the 5′-end of the sequence, the specific primer ICR19-1470RV (5′-GTG CCA CCT ACG TAT TTC TG-3′) was used to prime cDNA synthesis; the first-strand was tagged with poly(A) and then used as template for a RACE-PCR (rapid amplification of cDNA ends [2]) using ICR19-1470RV, RACE-ADTTT (5′-GAC TCG AGT CGA CAT CGA $(T)_{17}$-3′) and RACE-ADP (5′-GAC TCG AGT CGA CAT CGA-3′) as primers. The amplified fragments were then directly sequenced.

The cloning strategy is shown in Fig. 1. The cDNA clones and the PCR fragment corresponding to the 5′-end were sequenced in both orientations and used to assemble the complete sequence. Sequences were assembled with the PC/Gene program (IntelliGenetics, Mountain View, CA) and ORFs were generated with the ORF Finder Tool at the National Center for Biotechnology Information (NCBI) website (http://www.ncbi.nlm.nih.gov).

The assembled sequence consisted of 7560 nt, excluding the 3′-poly(A) tail, in good agreement with the 7500 nt estimate obtained by gel electrophoresis [11].

Analysis of the sequence showed the presence of six ORFs on the positive strand (Fig. 1), a 5′ untranslated region (UTR) of 78 nt and a 3′ UTR of 40 nt, followed by a poly(A) tail. No significant ORFs were found on the negative strand. The putative polypeptides encoded by the different ORFs are: ORF1 (nt 79–5055), 1658 amino acids (aa), M_r 187.3 kDa; ORF2 (nt 5063–5740), 225 aa, M_r 25 kDa; ORF3 (nt 5718–6047), 109 aa, M_r 12 kDa; ORF4 (nt 5974–6156), 60 aa, M_r 6.4 kDa; ORF5 (CP; nt 6179–7156), 325 aa, M_r 34 kDa, and ORF6 (nt 6856–7524), 222 aa, M_r 23 kDa. This genome organization resembles that of other filamentous viruses, particularly those in the genera *Carlavirus* and *Allexivirus*, which possess ORFs 1 to 6, though their ORF6 is about half the size of ICRSV ORF6 (Fig. 1).

Functional domains in the amino acid sequences derived from the ORFs were sought using the Conserved Domain Database (CDD) [5] at the NCBI website. ORF1 contained a viral helicase 1 and an RNA-dependent RNA polymerase 2

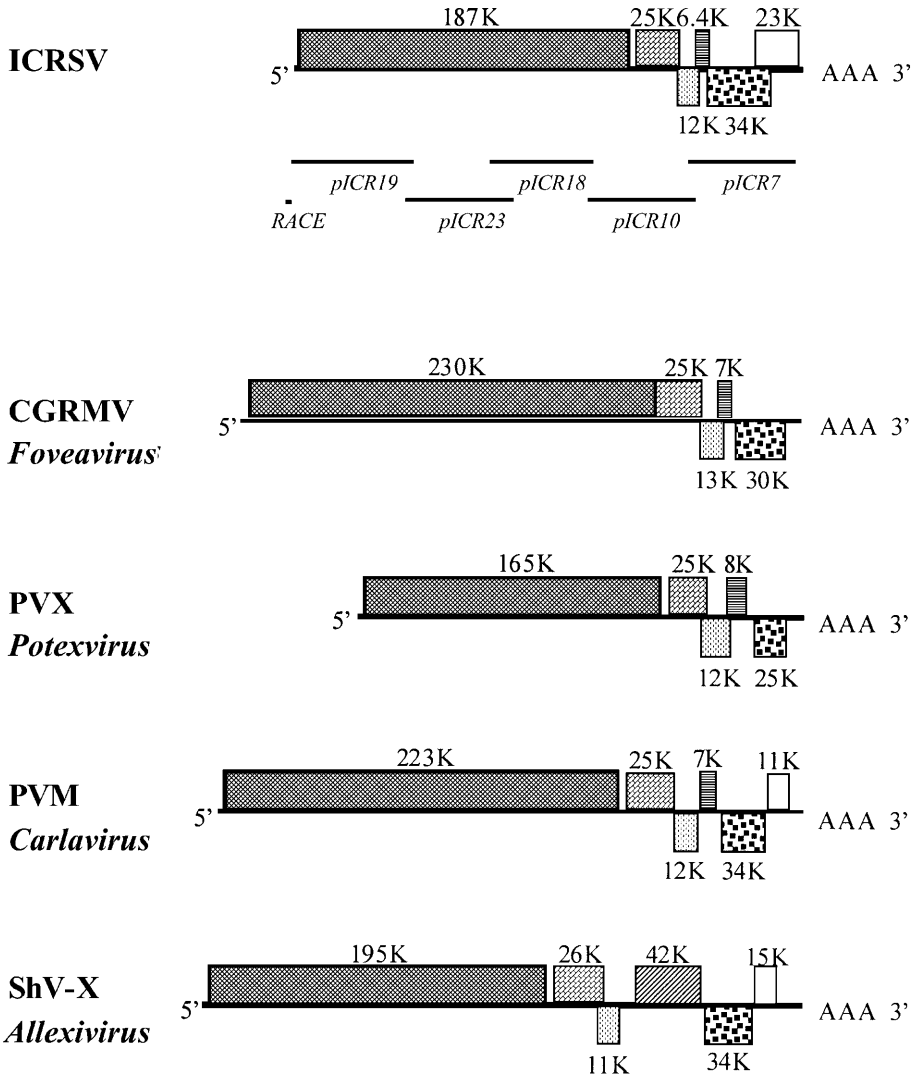

Fig. 1. Comparison of the genome structures of ICRSV, CGRMV, PVX, PVM and ShV-X. For ICRSV, the cloning strategy is also indicated. Boxes with the same pattern represent comparable ORFs. For virus names in full, see text

(RdRp) domain. The conserved motifs I–IV of the putative methyltransferase domain of "Sindbis-like" viruses [8] were located in the N-terminal region (aa 61–230). The NTP-binding helicase motif GxxxxGKS/T (x stands for any aa residue), and the other five motifs reported in nucleic acid helicases [12] were found in the central part (aa 930–1160). The conserved blocks for the RdRp [4] were in the region between aa 1350 and 1585. The highly conserved GDD motif was located at aa 1534–1536. The ORF1 protein was thus similar to the RdRps of potexviruses, allexiviruses, and, to a lesser extent, other filamentous viruses. The above information indicates that ICRSV ORF1 encodes the replicase.

The three partially overlapping ORFs 2, 3 and 4 formed a triple gene block (TGB), a common feature of several groups of plant viruses including

foveaviruses [3, 15], carlaviruses and potexviruses [9]. The TGB is considered to be involved in cell-to-cell movement [6]. The ORF2 protein contained a viral helicase 1 domain that covers almost the entire sequence; the NTP-binding helicase motif GxGKS/T was identified at aa 30–34. The ORF3 and ORF4 proteins contained large domains present in proteins of plant viruses in a number of families, according to the CDD.

ORF5 encoded the CP, and ORF6, encoding a putative protein of unknown function, showed limited similarity with nucleic acid-binding regulatory proteins in allexi- and carlaviruses, as reported by [11].

The conserved motif ACTTAA, found in potex- and carlaviruses at 32–65 nt-$(A)_n$ from the 3′-end, is present in ICRSV at nt 7526–7531. This motif is thought to play a role in the synthesis of negative-sense viral RNAs [14].

To clarify the relationship of ICRSV with other viruses, phylogenetic analyses were performed using the complete nucleotide sequence and the amino acid sequences of each ORF.

The following filamentous RNA viruses were included in the analysis. Allexiviruses: Garlic virus A (GarV-A, Acc. No. AB010300), Garlic virus B (GarV-B, U89243), GarV-C (AB010302), Shallot virus X (ShV-X, M97264); capilloviruses: Apple stem grooving virus (ASGV, D14995), Cherry virus A (CVA, X82547); carlaviruses: Potato virus M (PVM, D14449), Blueberry scorch virus (BBScV, L25658), Garlic latent virus (GarLV, Z68502); foveaviruses: Apple stem pitting virus (ASPV, D21829), Rupestris stem pitting associated virus (RSPaV, AF057136), Cherry green ring mottle virus (CGRMV, AF017780, tentative member); potexviruses: Potato virus X (PVX, X55802), Strawberry mild yellow edge virus (SMYEV, D12517), WClMV (X16636), Narcissus mosaic virus (NMV, D13747), Cymbidium mosaic virus (CymMV, AF016914), Cassava common mosaic virus (CsCMV, U23414), Papaya mosaic virus (PapMV, D13957), Potato aucuba mosaic virus (PAMV, S73580), Plantago asiatica mosaic virus (PlAMV, Z21647), Clover yellow mosaic virus (ClYMV, D29630), Banana mosaic virus (BaMV, D26017), Foxtail mosaic virus (FoMV, M62730); trichoviruses: Apple chlorotic leaf spot virus (ACLSV, M58152), Cherry mottle leaf virus (CMLV, AF170028); vitiviruses: Grapevine virus A (GVA, X75433), Grapevine virus B (GVB, X75448); and the unclassified Citrus leaf blotch virus (CLBV, AJ318061) and Banana mild mosaic virus (BanMMV, AF314662).

Figure 2 shows a phylogenetic tree generated by multiple alignment of the full nucleotide sequences using ClustalW [13], and displayed using TreeView [7]. The tree shows that ICRSV clusters with the potexviruses, though as an outlier. Overall homology with the two nearest potexviruses (NMV and PAMV) is about 47%. The next most related genus is *Allexivirus*.

The BLAST programs [1] available at the NCBI website were used to search for homologies in DNA and protein databases. The 5′UTR and 3′UTR of ICRSV did not show significant homology to the corresponding regions of other viruses, except for a weak similarity of the 5′UTR to potexviruses.

The putative aa sequences of each of the six ORFs of ICRSV were then analysed in relation to similar sequences in other viruses for which data are

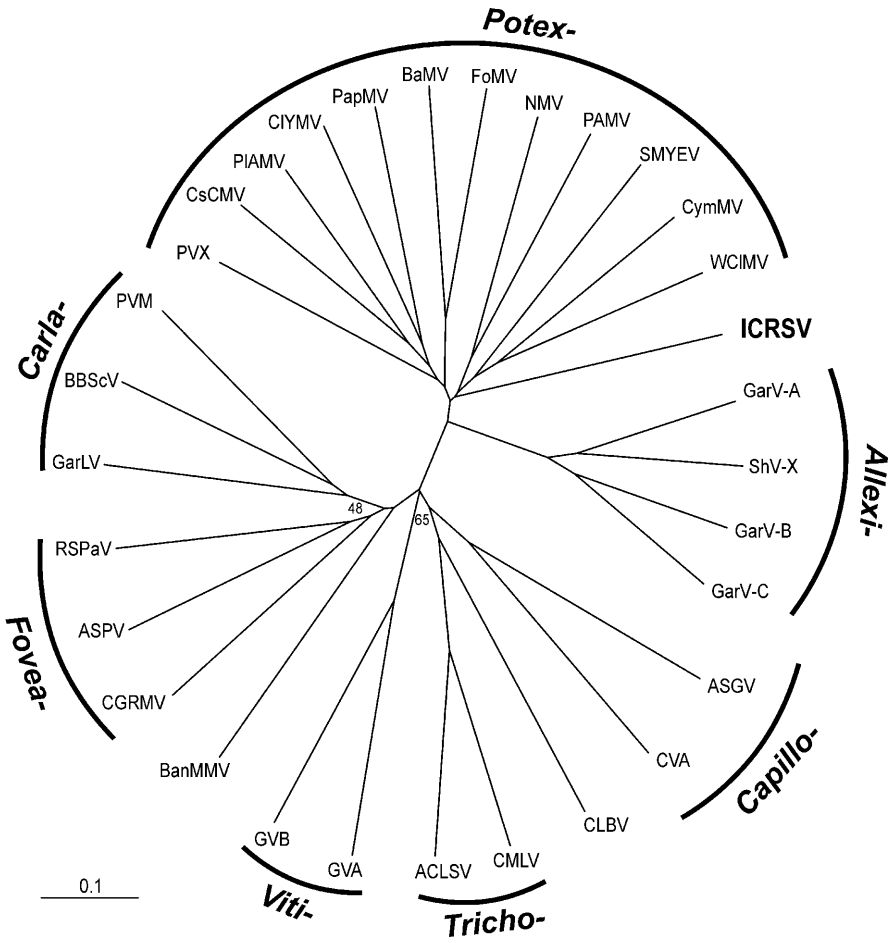

Fig. 2. Unrooted phylogenetic tree based on full nucleotide sequences of selected filamentous RNA viruses (see text for details), using the program ClustalW with gap open penalty of 10 and gap extension penalty of 5, and 1000 bootstrap replications. Branch lengths are proportional to estimated divergence. All bootstrap values exceeded 90%, except those indicated. For virus names in full, see text

available, using three approaches. Trees were constructed using ClustalW (Fig. 3a, b and c), overall percentage homologies were calculated (Table 1) and BLAST searches were conducted to detect regions of close homology.

ORF1. This clearly clustered with potexviruses (Fig. 3a). Percentage homology with potexviruses fell within the same range as homologies between individual potexviruses (Table 1) but homology with allexi-, carla- and foveaviruses was well below the range found within each genus. Similarity searches with BLAST confirmed this indication, since the highest scores were detected with WClMV, SMYEV, NMV, PVX, PAMV and other potexviruses, followed by allexiviruses. Thus it appears that ICRSV ORF1 is most similar to that of potexviruses.

ORF2. This ORF also clustered with ORF2s of potexviruses, though sequence divergence within the genus is high (not shown). According to percent homologies,

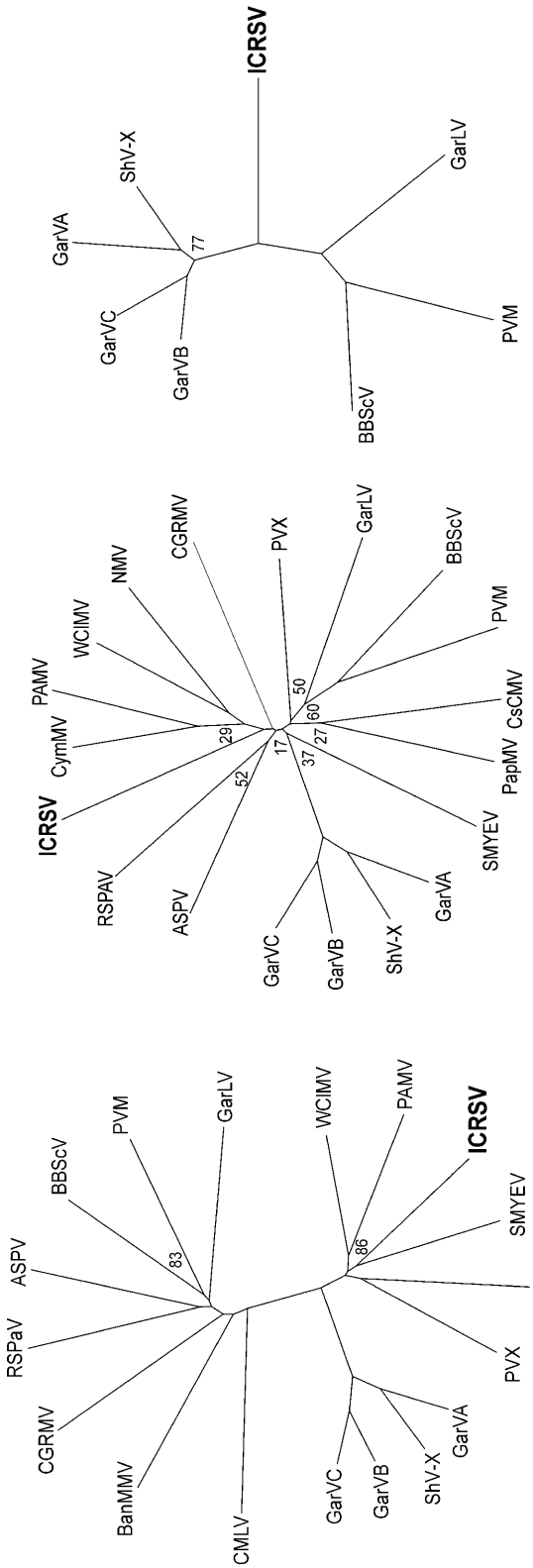

Fig. 3. Phylogenetic analysis of amino acid sequences of the RdRp (**a**), CP (**b**) and ORF6 (**c**) of selected filamentous RNA viruses (see text for details). Trees were constructed by the neighbour-joining method using the ClustalW with gap open penalty of 10, gap extension penalty of 0.1 and BLOSUM protein weight matrix. Bootstrap values (1000 bootstrap replications) exceeded 90%, except those indicated. Branch lengths are proportional to estimated divergence. For virus names in full, see text

Table 1. Summary of similarities between ICRSV and four genera of filamentous RNA viruses: percent homology (identity) of amino acid sequence of the six ICRSV ORFs with respect to other RNA viruses; average values of homology (and range of variation) for each ORF within each genus are also presented. Multiple alignments were performed using ClustalW with gap open penalty of 10, gap extension penalty of 0.1 and BLOSUM protein weight matrix

ICRSV	potex-			allexi-			carla-			fovea-		
	WClMV	PVX	intra-potex-	GarVC	GarVA	intra-allexi-	PVM	GarLV	intra-carla-	RSPaV	ASPV	intra-fovea-
ORF1	51.6	43.1	39.7–56.0	34.5	33.4	69.6–83.0	17.8	19.4	39.4–42.6	20.0	19.8	37.8–45.5
ORF2	31.5	28.3	23.3–44.5	24.5	25.6	54.5–67.8	29.3	21.3	31.1–50.4	26.1	29.9	37.7–47.0
ORF3	36.1	33.7	27.7–50.0	37.3	38.2	54.4–68.0	38.6	42.0	44.2–49.0	30.8	40.7	32.7–44.8
ORF4	33.9	28.1	12.9–34.4	–	–	–	22.0	18.3	20.6–35.0	20.0	28.1	26.6–39.7
ORF5	35.3	29.5	23.0–56.8	19.3	21.7	63.9–73.9	20.1	20.7	38.1–46.8	24.2	21.7	28.5–38.8
ORF6	–	–	–	30.5	30.8	55.6–72.8	15.4	21.5	38.5–42.9	–	–	–

"–"= no corresponding ORF is present

it could fall within the potexviruses (Table 1). BLAST searches gave similar results, with closest affinities to potexviruses (PapMV, Lily virus X, Cactus virus X), followed by some foveaviruses (ASPV, CGRMV) and carlaviruses (Chrysanthemum virus B, PVM).

ORF3. Phylogenetic analysis of this ORF did not indicate any specific taxonomic position; it clustered with carla-, potex- and foveaviruses, while allexiviruses formed a separate homogeneous cluster (not shown). Percentage homologies (Table 1) also indicated that only allexiviruses were clearly different. BLAST search results however showed highest similarity with a carlavirus (GarLV) and an allexivirus (GarV-B).

ORF4. Similarly, the taxonomic position of ORF4 was not evident from phylogenetic trees (not shown). Individual species within each genus showed such divergence that clustering was poor (low bootstrap values). Percentage homology (Table 1) fell well within the range between individual potexviruses, but this criterion also did not exclude the carla- and foveaviruses. Using BLAST searches, highest similarity, though with low scores, was with CVB (*Carlavirus*) and WClMV, CymMV and PVX (*Potexvirus*). ORF4 of allexiviruses showed essentially no homology with the ORFs at the same location of ICRSV and of the other three genera. This is indicated as blanks in Table 1.

ORF5. Phylogenetic analysis (Fig. 3b) confirmed the similarity to potexviruses found by [11]. This is also apparent from Table 1, and BLAST searches showed that the closest sequences were the CPs of WClMV, CymMV, PAMV, all potexviruses. However, the size of the ICRSV CP (34 kDa) is much larger than that of potexviruses (18–27 kDa), and homology was limited to the C-terminal part of the protein. When the N-terminal 133 amino acids were analysed with the BLAST search, no significant similarities were found. However, using the Fasta3 program (at the website http://www.ebi.ac.uk) with a PAM70 matrix, a limited similarity with a conserved domain of *Alphaherpesvirus* glycoprotein 1 was detected.

ORF6. This had counterparts only among carla- and allexiviruses. Phylogenetic analysis (Fig. 3c) showed that this ORF does not cluster close to these genera, and Table 1 indicates that though some similarity exists with viruses in these two genera, percentage homology is less than that between individual members of each genus. The ICRSV ORF6 protein is much larger than in allexi- and carlaviruses, and any homology is confined to the C-terminal half.

The above analysis indicates that ICRSV could almost fit into the genus *Potexvirus*, although ORF3 is also compatible with *Foveavirus* and ORF4 with *Carlavirus* and *Foveavirus*. The presence of ORF6, however, makes ICRSV clearly distinct from the genus *Potexvirus*.

Other characteristics of ICRSV also separate it from potexviruses. The genome size (7.6 Kb) is in the range of vitiviruses (7.4–7.7 Kb), larger than potexviruses (6–7 Kb) and smaller than carla- (8.3–8.5 Kb), allexi- (8.1–8.8 Kb) and foveaviruses (8.4–9.3 Kb). The size of the CP (34 kDa), on the other hand falls outside that of viti- and potexviruses (18–27 kDa) but is in the range of carla-, allexi- and foveaviruses (31–36 kDa, 28–36 kDa and 28–44 kDa respectively). In overall

genome organization, ICRSV is similar to carla- and allexiviruses, due to the presence of ORF6. It should also be noted that ICRSV differs from carla- and potexviruses in possessing more flexuous particles that display a clear helix [11], a morphology closer to that of fovea-, allexi- and capilloviruses. No biological data on ICRSV, such as host range or natural means of transmission, are available, that might give further clues to the taxonomic placement of this virus.

We conclude that although ICRSV is most similar to members of the genus *Potexvirus*, it differs from them significantly. Constitution of a new genus to accommodate ICRSV should be considered.

References

1. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402
2. Frohman MA, Dush MK, Martin GR (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc Natl Acad Sci USA 85: 8998–9002
3. Jelkmann W (1994) Nucleotide sequences of apple stem pitting virus and of the coat protein gene of a similar virus from pear associated with vein yellows disease and their relationship with potex- and carlaviruses. J Gen Virol 75: 1535–1542
4. Koonin EV (1991) The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. J Gen Virol 72: 2197–2206
5. Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, Bryant SH (2002) CDD: a database of conserved domain alignments with links to domain three-dimensional structure. Nucleic Acids Res 30: 281–283
6. Mushegian AR, Koonin EV (1993) Cell-to-cell movement of plant viruses. Insights from amino acid sequence comparisons of movement proteins and from analogies with cellular transport systems. Arch Virol 133: 239–257
7. Page RDM (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. Comput Appl Biosci 12: 357–358
8. Rozanov MN, Koonin EV, Gorbalenya AE (1992) Conservation of the putative methyltransferase domain: a hallmark of the 'Sindbis-like' supergroup of positive-strand RNA viruses. J Gen Virol 73: 2129–2134
9. Rupasov VV, Morozov SYu, Kanyuka KV, Zavriev SK (1989) Partial nucleotide sequence of potato virus M RNA shows similarities to potexviruses in gene arrangement and the encoded amino acid sequences. J Gen Virol 70: 1861–1869
10. Rustici G, Accotto GP, Noris E, Masenga V, Luisoni E, Milne RG (2000a) Indian citrus ringspot virus: a proposed new species with some affinities to potex-, carla-, fovea-, and allexiviruses. Arch Virol 145: 1895–1908
11. Rustici G, Noris E, Accotto GP, Luisoni E, Milne RG, Pant RP, Ahlawat YS (2000b) Further characterisation and detection of Indian citrus ringspot virus. In: da Graça JV, Lee RF, Yokomi RK (eds), Proc, Fourteenth Conference of the International Organisation of Citrus Virologists IOCV 2000, pp 360–362
12. Skryabin KG, Morozov SYu, Kraev AS, Rozanov MN, Chernov BK, Lukasheva LI, Atabekov JG (1988) Conserved and variable elements in RNA genomes of potexviruses. FEBS Lett 240: 33–40

13. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680
14. White KA, Bancroft JB, Mackie GA (1992) Mutagenesis of a hexanucleotide sequence conserved in potexvirus RNAs. Virology 189: 817–820
15. Zhang Y-P, Kirkpatrick BC, Smart CD, Uyemoto JK (1998) cDNA cloning and molecular characterisation of cherry green ring mottle virus. J Gen Virol 79: 2275–2281

Author's address: Dr. Gian Paolo Accotto, Istituto di Virologia Vegetale, CNR, Strada delle Cacce, 73, I-10135 Torino, Italy; e-mail: g.accotto@ifa.to.cnr.it