

# Molecular characterization of the largest and smallest genome segments, S1 and S12, of *Rice gall dwarf virus*

Heng-Mu Zhang · Jian Yang · Xing Xin ·  
Jian-Ping Chen · Michael J. Adams

Received: 16 November 2006 / Accepted: 13 December 2006 / Published online: 3 August 2007  
© Springer Science+Business Media, LLC 2007

**Abstract** The nucleotide sequences of segments S1 and S12 of a Chinese isolate of *Rice gall dwarf virus* (RGDV) were determined. This provides the first complete sequences of these segments. The complete sequence of S1, the largest genome segment of RGDV, was 4,505 nucleotides in length and was predicted to encode a large protein of 1,458 amino acids with a calculated molecular mass of nearly 166.2 kDa. The protein was related to that encoded by S1 of *Rice dwarf virus* (RDV; 50% identity and 67% similarity) and (to a lesser extent) to some large proteins of other reoviruses. It appears to be an RNA-dependent RNA polymerase (RdRp) and is probably present in particles as a minor core protein. S12, the smallest genome segment of RGDV, was 853 nucleotides in length, encoding a single major protein of 206 amino acids with a calculated molecular mass of nearly 23.6 kDa. This protein, though a little larger than those of RDV S11 and *Wound tumor virus* (WTV) S12 in size, showed some similarity to them, especially in the conserved N-terminal

region and may have RNA-binding properties. Despite having a common host plant, RDV and RGDV were not more closely related to one another than either of them was to WTV. Phylogenetic analysis of the RdRp showed that members of the genus *Phytoreovirus* were more closely related to those of the genus *Rotavirus* than to any other genus within the family *Reoviridae*.

**Keywords** Rice gall dwarf virus · Phytoreovirus · RdRp · Reoviridae · Phylogeny · Sequence analysis

## Introduction

*Rice gall dwarf virus* (RGDV), which was first described in the early 1980s [1–4], is one of the three recognized species in the genus *Phytoreovirus* within the family *Reoviridae*. RGDV causes a severe disease of rice and is responsible for significant yield losses in rice crops in China and South-East Asia. Field-infected plants of rice (*Oryza sativa*) are typically stunted with small galls along the undersides of the leaf blades and outer sides of the leaf sheaths [1, 2, 4].

Rice gall dwarf virus shares many features with *Rice dwarf virus* (RDV), another phytoreovirus. Both viruses naturally infect rice in China and South-East Asia and are propagatively transmitted to rice plants in a persistent manner and transovarially by leafhoppers (Delphacidae) [5, 6]. Both viruses multiply in their insect vectors, in which they cause no apparent symptoms, as well as in their plant hosts, in which they induce severe growth abnormalities, including dwarfing. In morphology and architecture of virions, RGDV and RDV, as well as *Wound tumor virus* (WTV) (the type member of the genus *Phytoreovirus*), have angular spikeless icosahedral

---

The nucleotide sequence data reported in this article have been submitted to the GenBank nucleotide sequence database and have been assigned the accession numbers DQ333946 and DQ494209.

---

H.-M. Zhang (✉) · J. Yang · X. Xin · J.-P. Chen  
Zhejiang Provincial Key Laboratory of Plant Virology,  
Key Laboratory of Animal and Plant Virology,  
Ministry of Agriculture, Institute of Virology and  
Biotechnology, Zhejiang Academy of Agricultural Sciences,  
Hangzhou 310021, P.R. China  
e-mail: zhhengmu@tsinghua.org.cn

M. J. Adams  
Plant–Pathogen Interactions Division, Rothamsted Research,  
Harpenden, Hertfordshire AL5 2JQ, UK  
e-mail: mike.adams@bbsrc.ac.uk

double-shelled particles approximately 65–70 nm in diameter and contain 12 segments of double-stranded RNA (dsRNA) [7, 8] and at least six structural proteins including four core structural proteins and two outer capsid proteins [9, 10]. The two rice viruses are also similar in the molecular sizes and the molar ratios of their structural proteins; interestingly, the double-shelled virus-like particles, which are indistinguishable by electron microscope from native RDV and RGDV virions, are reassembled successfully with heterogenous core and outer capsid proteins of RGDV and RDV [11, 12].

Although RGDV and RDV resemble each other biologically and morphologically, they are clearly distinct in terms of their serology, pathogenicity (symptomatology, and tissue distribution), biochemical and molecular biological properties. Antiserum against RGDV particles does not react with RDV particles and vice versa. In infected rice plants, RGDV is essentially restricted to phloem-related cells and induces tumors derived from the phloem while RDV is systemic in infected plants and does not cause hyperplasia [13]. RGDV particles retain p2 protein, a minor outer capsid protein, and infectivity irrespective of carbon tetrachloride (CCl<sub>4</sub>) treatment while RDV particles lose P2 and infectivity by this treatment [10, 14, 15]; The optimum temperature for the activity of the RNA-dependent RNA polymerase is 25°C in RGDV but 35°C in RDV [16].

As expected, these similarities and differences between RDV and RGDV are reflected in the molecular biological properties of their genomes. They are therefore very suitable for analytical comparisons to correlate genome structure and function. In the case of RDV, the nucleotide sequence of each of the 12 genomic RNAs has been determined [17], which makes it the only member of the genus so far sequenced completely, and the functions of some viral genes have been identified [15, 18–20]. In contrast, sequence information from RGDV is more limited and only available for the S2–S3/S5/S8–S11 genome segments [10, 11, 21–24]. In the early comparative studies of S2/S3/S5/S8–S11 of RGDV and RDV, it was shown that some viral genes, especially structural proteins, P2, P3, P5, and P8, are well conserved and have similar location and roles. However, little is known about the viral proteins encoded by the other genome segments, mainly because of the lack of sequence information. As a part of our efforts to gain insight into the correlation between phenotypic and genomic properties of RGDV and RDV, we have now analyzed the complete nucleotide sequences of S1 and S12, the largest and smallest of the RGDV genome segments, and discuss the possible functions of their encoded gene products and the relationship between RGDV, RDV and other reoviruses.

## Materials and methods

### Source of virus isolate and extraction of genomic dsRNA

Naturally infected rice plants with typical RGDV symptoms, including dwarf symptoms and small galls along the under sides of the leaf blades and outer sides of the leaf sheaths, were collected from Guangxi province in August 2004 and stored at –80°C. The viral genomic dsRNAs were extracted from purified virus particles or directly from the infected leaves of rice plant using the method described by Uyeda et al. [25]. The purified genomic dsRNAs were separated and then S1 and S12, respectively the largest and smallest of the genome segments, were purified using 1% low-melting temperature agarose gels [26].

### cDNA library construction

cDNA libraries were constructed essentially as described previously [27, 28] with minor modifications. In brief, each purified genomic segment was denatured [27] and used as a template for first strand cDNA synthesis by Superscript Reverse Transcriptase (MBI) in the presence of 100 ng of 9-nucleotide random primers. The resulting cDNA was purified, annealed, repaired, ligated into pGEM T easy vector (Promega), and transformed into competent *E. coli* TG1 cells for cloning. The transformants were screened using the ampicillin resistance and  $\alpha$ -complementation methods [26]. Recombinant plasmid DNA was isolated by the alkaline lysis method [26], analyzed by agarose gel electrophoresis, and 5 clones for S12 and 16 clones for S1 from the respective cDNA libraries, containing insertions of 200–1,000 bp, were used for sequencing.

### Amplification of internal regions by RT-PCR and terminal regions by ligation-RT-PCR

When all the sequences were assembled, four fragments for S1 (800, 376, 248, and 814 nt) and two for S12 (370 and 387 nt) were obtained. To amplify the remaining internal sequences, specific primers (Table 1), were designed on the basis of sequences determined and these were used to amplify the internal regions by RT-PCR. After cloning and sequencing of these amplified products, we obtained two contiguous sequences covering most of the respective genome segments, one of approximately 4.5 kb for S1 and the other of approximately 0.85 kb for S12.

To ensure that complete terminal sequences were obtained, primer zhm-1 (Table 1) was first ligated to both 3'-ends of the viral RNA. After cDNA synthesis, the purified cDNA was used as a template for PCR with primer zhm-2

**Table 1** Primers used for amplification of genome segments S1 and S12 of RGDV

Primer name	Direction	Position (nt)	Sequence (5'- to 3'-)	T <sub>m</sub> (°C)
G1-1	+	552–573	GAT AAT GCG ACG CGA ATA CAA T	62
G1-2	+	665–686	CAA CTT ACG AAC TAA GGA GGA G	64
G1-3	+	1682–1703	TGG GAT AGG TAA TCT ACG GTT G	64
G1-4	+	1735–1756	TTA GAA CCA CGT CAA AAT CTG C	62
G1-5	+	3162–3183	GAT CTA GTC GAG GCT GGG TTA A	66
G1-6	+	3205–3226	TTG AGC AAA ATG AGA TCG TGT C	62
G1-7	–	3830–3809	AGT TAT TGT CAA CCC ATT CCT C	62
G1-8	–	3784–3763	GAT CCT CCA TAT CCT TGA TTC T	62
G1-9	–	3253–3232	CGA AGG AAA TGT CGC ATC AAA G	64
G1-10	–	3178–3157	CCA GCC TCG ACT AGA TCC AAT A	66
G1-11	–	1756–1735	GCA GAT TTT GAC GTG GTT CTA A	62
G1-12	–	1702–1681	AAC CGT AGA TTA CCT ATC CCA G	64
G1-13	+	3484–3505	TGC CAT ACT CCA CCA TAA ATC A	62
G1-14	–	726–705	ATA AAG CGG GAA CTG GAT GGT A	62
G12-1	+	173–192	GTG TCT GTC CCT CGG CTT TG	62
G12-2	–	651–630	TTA CCT CGG TCT TCG TTT ACT T	62
zhm-1			PO <sub>4</sub> -CTC TTC CCC TCC CTC CTC-NH <sub>2</sub>	60
zhm-2			GAG GAG GGA GGG GAA GAG	60

(Table 1, complementary to zhm-1) and one of the four segment-specific internal primers, G1-13, G1-14, G12-1, and G12-2 (Table 1). The resulting fragments were cloned and sequenced. Experimental details were as previously described [27, 28].

#### Sequencing and sequence analysis

Recombinant plasmid DNA used for sequencing was prepared using the QIAprep spin mini prep kit (Qiagen Ltd), and the inserts were sequenced entirely on both strands using the BigDye Terminator v3.1 Cycle Sequencing Kit (Perkin Elmer Applied Biosystems, Foster, USA) on an ABI PRISM 3730 DNA Sequencer (Perkin Elmer Applied Biosystems, Foster, USA) with universal primers T7 and SP6. Sequence assembly and analysis was performed using the DNAMAN version 4.0 program (Lynnon BioSoft, Quebec, Canada). Using the strategy described, a strong consensus sequence was obtained from 3 to 14 clones in each region. Most of the clones had identical sequence in the overlapping regions. In total, 11 differences were observed in internal regions, of which six were C/T substitutions and five A/G, but only one G/A substitution, at nt 2455 in S1 produced an amino acid change (Gly805 > Glu). To assess the frequency of these differences, at least seven more independent clones obtained by RT-PCR amplification were sequenced; at least six out of seven sequences were in agreement with the consensus sequence.

Search for proteins homologous to the predicted proteins was performed with the BLAST program [29]. To construct

a phylogenetic tree for putative RdRp genes, amino acid sequences were obtained from SwissProt databases (<http://www.sanbi.ac.za/mrc/Databases/SWISSPROT.htm>) or deduced from EMBL/NCBI/DDBJ DNA databases and aligned using Clustal W, version 1.6 [30], and Multalin, version 5.4.1 [31]. Aligned sequences were used to build an evolutionary tree by the Neighbor-Joining method [32]. Phylogenetic and molecular evolutionary analyses were done using MEGA version 3.1 [33].

## Results and discussion

### Genomic analysis of RGDV S1

The complete nucleotide sequence of genome segment S1 (accession number DQ494209) was 4,505 nucleotides in length with a 5'-untranslated region (UTR) of 41 nt, a large open reading frame (ORF) of 4,376 nt (from nts 42 to 4,418), and a 3'-UTR of 87 nt. Its G/C content was 38.2%, the lowest value in all the published RGDV genome segments except for S5 (data not shown). The extreme 5'- and 3'-ends of the sense strand had the sequence 5'-GGCAUUUUU...UUGAU-3', identical to those in S2 and S3. Although the conserved terminal sequences mismatched one or more nucleotide with those of all other published RGDV genome segments, they still all support the consensus RGDV terminal sequences, namely 5'-GGXA.....UGAU-3' (X = U or C), and also conform to the consensus terminal sequences of the genus *Phytoreovirus*

proposed by Omura (17), namely 5'-GGXA.....XGAU-3', which are genus-specific and thought to be a package signal for viral RNA [34].

A perfect 10 nucleotides inverted repeat, 5'- ... UUUUGAGCCA (nt 7–16). ... UGGCUCAAAA... (nt 4489–4499)-3', was identified immediately adjacent to the conserved terminal sequences, which is segment-specific and thought to act as a signal to specify a particular genome segment [34]. The MFOLD program predicted that these inverted repeats were able to form stable secondary structures displaying a panhandle, a stem loop and a non-base-paired 3'-tail that presumably act as replication and packaging signals. As expected, the shape of the stem loop was distinct from those predicted for each of the other sequenced segments of RGDV (data not shown).

Upstream of the AUG putative start codon (nts 42–44), there was only one out-of-frame UGA stop codon (nts 10–12) and no minicistron. There was no earlier in-frame AUG upstream of position 42 that allowed continuous translation through the rest of the sequence, indicating that the ORF could not begin at a more distal site. By contrast, in RDV S1 there is a minicistron (nts 6–29) (D90198 and U73201) or an in-frame start codon (nts 6–8) (D10222) upstream of the putative start codon (nts 36–38). Translation of the only one major ORF, which is the largest ORF in the reported RGDV genome, yields a putative large protein (designed RGDV p1) of 1,458 amino acids with a calculated molecular mass of nearly 166.2 kDa and a pI of 7.0. The Mr of RGDV p1 is consistent with the size of a minor structural protein, which was thought to be located in the inner part of the core particle [9].

Further analysis indicates substantial similarities between RGDV p1 and the minor core capsid protein of RDV (RDV p1, encoded on S1) which has a Mr of 164.1 kDa, and which is thought to be a viral RdRp [35]. The p1 proteins of RGDV and RDV have similar numbers of amino acid residues (1458 in RGDV and 1444 in RDV) similar amino acid compositions (data not shown), and similar pI, (7.0 in RGDV and 7.1 in RDV).

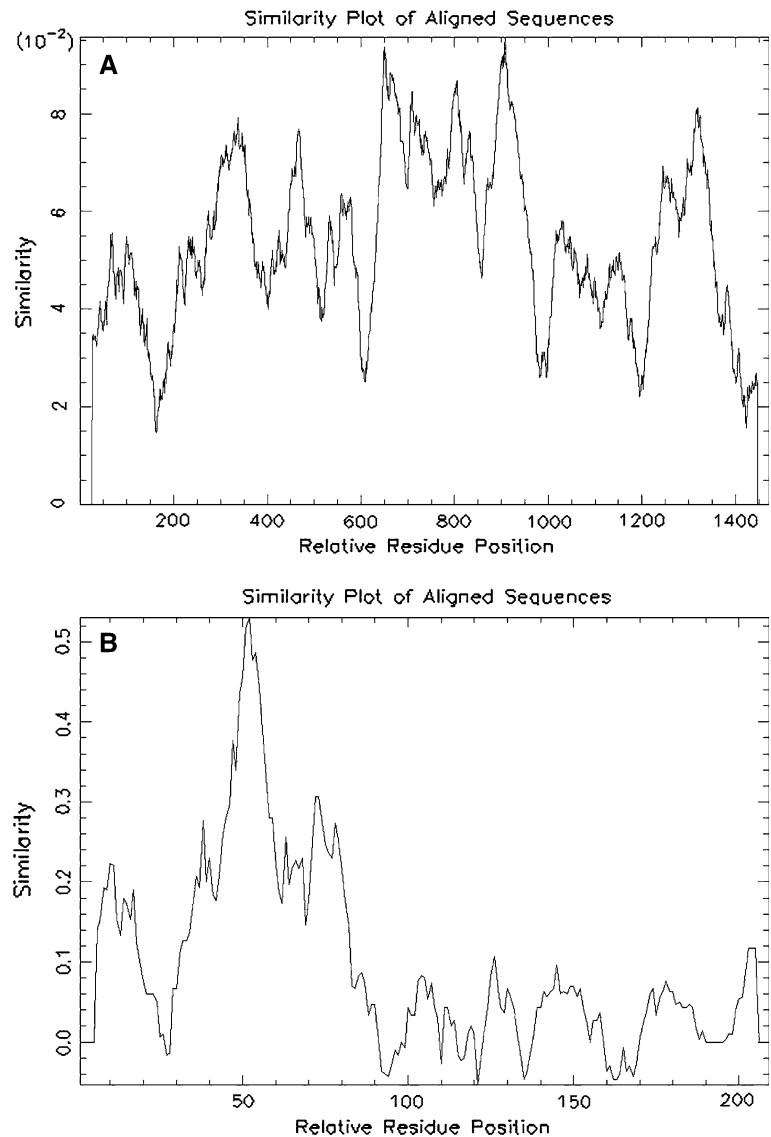
An alignment of the amino acid sequences of RGDV p1 and RDV p1 had 34 gaps, of which 22 were located at the position aa 973–994 while all the remaining gaps were located at the N- and C-termini. The two proteins were clearly related, with 50% amino acid identity over the entire length (69% similarity). Many stretches with more than 5 aa showing complete sequence identity were found in the central region (aa 44–1340) while in the region of aa 466–923, there were at least eight stretches with 12–41 aa identical amino acids (data not shown). As shown in Fig. 1A, particularly high homology was found in the region (aa 620–965), where the amino acid sequence identity was up to 70% and the similarity up to 84%, strongly supporting the view that the p1 protein encoded by RGDV

S1 is a viral RdRp located within the core particle, as a functional homolog of RDV p1. However, the N-terminal 200 amino acids and the C-terminal 120 amino acids were much less conserved (31–35% identity and 50% similarity) than the central regions, probably reflecting the slight differences in the regulatory mechanisms for the two RdRps, for example in optimum temperatures.

By comparison with the proteins encoded by other genome segments, the putative viral RNA polymerase was one of most conserved viral proteins between RDV and RGDV, with values close to that of the major outer capsid protein Vp8 (51% identical and 68% similar) encoded by S8 and that of the other minor core protein Vp5 (50% identical and 67% similar), which was considered to be a viral guanylyltransferase and is encoded by S5. This contrasts with the values for the major inner core protein Vp3 (41% identical and 62% similar), which is considered to form the basic scaffold of the core particle and is encoded by S3, for those of the minor outer capsid protein Vp2 (38% identical and 57% similar) encoded by S2, and those of all the other non-structural proteins (22–36% identical and 44–60% similar). Further, comparison of the respective hydrophobicity profiles of the p1 proteins of RDV and RGDV confirm the similarity of the two proteins (data not shown), also supporting the proposed functional conformity.

Homology searches within the Swiss-Prot, GenPept, PIR and PDB databases were performed for the overall RGDV p1 sequence and for stretches of residues using protein-protein BLAST (BLASTP [29]). The results showed that, in addition to similarities to RDV p1, the deduced amino acid sequence of RGDV p1 showed some relationships in localized regions (20–24% identical and ~37–41% similar) with the RdRps of some animal reoviruses, including Chuzan virus (CZV), Yunnan obivirus (YNV), avian rotavirus (ARV), bluetongue virus (BTV), and Kadipiro virus (KDV). These similarities are difficult to detect in full-length sequence alignments but the five conserved RdRp motifs can be found and aligned between RGDV RdRp and the RdRps of other reoviruses (Fig. 2). These motifs were found in RGDV within the region 639–887 of the protein: The motif RxxRxI, found at position 641–646, is believed to participate in binding ribonucleotide triphosphates (rNTP) ensuring the faithful selection of the correct NTP by the reovirus RdRp [37]; the motif DxxxxD was found at position 720–725, in which the highly conserved aspartate residues are thought to be involved in magnesium coordination and possibly sugar selection [38]; the motif SGxxxT was found at position 804–809, of which importance for RdRp enzyme activity has been demonstrated in other viruses by site-directed mutagenesis, even conservative substitutions, of the most highly conserved residues abolishing or drastically inhibiting RNA polymerase

**Fig. 1** Similarity plots of aligned sequences among RGDV and RDV p1 (A) and RGDV p12, RDV p11, and WTV p12 (B). Similarity scores were averaged over a running window of 50 (A) or 10 (B) amino acids



activity [39] although the specific function of these conserved residues remain to be elucidated; the motif GDD, well known as the core motif of RdRp, was at position 845–847, in which the first aspartate is thought to be involved in the binding of the divalent cations  $Mg^{2+}$  and/or  $Mn^{2+}$ ; and the motif xxKxx was found at position 883–887. These similar features strongly support the hypothesis that the protein encoded by RGDV S1 is a viral RdRp.

It is well-known that the RNA-directed RNA polymerase (RdRp) is an essential protein encoded in the genomes of all RNA-containing viruses with no DNA stage. Sequence analysis has confirmed the presence of RdRp in all fully sequenced reoviruses and in the case of the genus *Phytoreovirus* enzyme activity has been demonstrated experimentally for RGDV [16] and also for the other members of the genus, WTV and RDV [40, 41]. A characteristic of reoviruses generally is that the RdRp is a

minor core protein; transcription and replication occur within a core complex, an exquisite nano-scale machine for transcription and replication. Our results with RGDV are consistent with this pattern.

#### Genomic analysis of RGDV S12

The complete nucleotide sequence of genome segment S12 (accession number DQ333946), the smallest of the RGDV genome segments, was 853 nt in length with a 5'-UTR of 30 nt, a major ORF of 620 nt and a 3'-UTR of 202 nt. Its G/C content was about 5% more than that of RGDV S1. The extreme 5'- and 3'-ends of the sense strand had the sequence 5'-GGUAUUUUU... UGAU-3', which is identical to the terminal sequences of RGDV S8–S11 and also conforms to the consensus terminal sequences of the genus *Phytoreovirus* as described above. An imperfect 9 nucleotides

RGDV	AWRPVRPI...DCSSWDQT...SGLRDTFFMNSVQN...FQVAGDDAI...PQKTV
RDV-C	AWRPVRPI...DCSSWDQT...SGLRDTFFMNSVQN...FQVAGDDAI...PQKTV
RDV-H	AWRPVRPI...DCTSWDQT...SGLRDTFFMNSVQN...FQVAGDDAI...PQKTV
RDV-A	AWRPVRPI...DCSSWDQT...SGLRDTFFMNSVQN...FQVAGDDAI...PQKTV
RRSV	IGRRQRAI...DASVQASV...SQPFTHVHTFTL...LTVQGDTR...GFKVS
RBSDV	IDRRARVI...DMKGMDAH...TGFFATSAQHTLFL...HSVMGDDVL...LSRMY
MRCV	IDRRARVI...DMKGMDAH...SGFFATSAQHTLFL...HSVMGDDVL...LSRIF
FDV	IDRRARVI...DMKGMDAH...SGFFATSAQHTLFL...HSVMGDDV...LSRIF
NLRV	IDRRGRII...DMSGMDAH...SGLFATSGQHTMFL...NYVMGDDIF...YSKYS
MRV-1	VQRRPRSI...DASITWDF...SGSTATSTEHTANN...YVCQDDGL...GWKYD
AHSV	PIKSTRIV...DYSEFDTH...SGENSTLIANSLHN...EQYVGDDTL...MSKTL
BTV	PIKATRTI...DYSEYDTH...SGENSTLIANSMHN...EQYVGDDTL...PSKTM
CZV	PIKATRTI...DYSDYDTH...SGENSTLVANSMHN...EQYVGDDTL...ASKTI
PHSV	PIKSTRII...DYSEYDTH...SGENTTLVWNSIHN...EMVVGDDML...EAKTT
SCRV	PVKPTREV...DMSNFDKH...SGENSTLVANSLHN...EMVVGDDTL...PAKTT
YUOV	PIKATRII...DYSEYDQH...SGENTTLVWNSVHN...ETVVGDDML...PAKTT
HuRV-A	PGRRTRII...DVSQWDSS...SGEKQTKAANSIAN...IRVDGDDNY...KVKAL
BoRV-A	PGRRTRII...DVSQWDSS...SGEKQTKAANSIAN...IRVDGDDNY...KVKAL
SiRV-A	PGRRTRII...DVSQWDSS...SGEKQTKAANSIAN...IRVDGDDNY...KVKAL
MuRV-B	PARQTRGI...DVSQWDAS...SGEKTTKIGNSPAN...LRVDGDDNY...RVKAL
PoRV-C	PGRRTRAI...DVSQWDAS...SGEKQTKIMNSIAN...IRVDGDDNY...KVKAL
CTFV	VGRRPRVI...DVKGMDSS...SGLLNTADQHTPLG...GSVLGDDQV...ETKFL
EYAV	VGRRPRVI...DVKGMDSS...SGLLNTADQHTPLG...GSVLGDDQV...ETKFL
GCRV	VQRRRSI...DASITYNY...SGSTATSTEHTANN...YVCQDDGI...YLKLY
CSV	VQRRRSI...DASITYAA...SGSTATSTEHTANN...YVCQDDGM...YLKLY
GSRV	VQRRRSI...DASITYNY...SGSTATSTEHTANN...YVCQDDGI...YLKLY
CPV-1	SDRRQRAI...DVAGMDAS...SGRADTSTHHTVLL...IKILGDDIM...LSQNS
CPV-14	IDRRQRAI...DISGMDAS...SGRADTSTHHTVLL...MRILGDDV...YKNS
CPV-15	IDRRQRGI...DVKGMDAS...SGRADTSAHHTMLL...LSAMGDDLI...ASRNS
MYRV-1	QQRPRVI...DIIGMDAS...SGWYNTSSQHTMLG...GRVLGDDIL...FLKQK
MYRV-3	QQRRRRII...DVAGMDAT...SGKFSTSSQHTTIG...SSVLGDDLQ...FLKQK
BAV	GTRAKRII...DPGQFDS...SGELTTQTRNTTIN...DNKVGDDSV...HIKIW
KDV	GTRDKRII...DFSQYDSS...SGELTTLQLRNTDPN...ENIVGDDKY...HIKIW
LNSV	GTRAKRII...DFAQFDSS...SGELTTLQLRNTATN...ENIGDDKN...HIKIF
DpRV-1	LGKNGGI...DSSRPDAV...SGIYVKLEAIHNQT...DFPDGDDVK...LGKRY
APRV	IERRQRGI...DVKGMDTH...SGLITTSNHHHTQML...LMILGDDVS...ESRNF
OpBuRV	VDRGRVI...DVKGMDAA...SGKFSTNAQHSFIN...SLISGDDLT...LSKYS
ESRV	GGRATRII...DYSTWDRQ...SGELTTFQSNHVTN...INIVGDDVN...HIKQY
MPRV	PARILRLI...DASKLDQH...SGALITAEKNTVTS...DYLWGGDAY...LTKYV
consensus	--R--R-I...D-S--d--...SG--T--S--...GDD--...-K-

**Fig. 2** Conserved amino acid sequences specific for RNA-dependent RNA polymerases of reoviruses. The motifs are those described in Nakashima et al. [36]. The sequences used for comparison were: RGDV, genus *Phytoreovirus* (DQ494209, in this study); *Rice dwarf virus*—China isolate (RDV-C), genus *Phytoreovirus* (U73201); *Rice dwarf virus*—H isolate (RDV-H), genus *Phytoreovirus* (D10222); *Rice dwarf virus*—A isolate (RDV-A), genus *Phytoreovirus* (U90198); *Rice ragged stunt virus* (RRSV), genus *Oryzavirus* (U66714); *Rice black-streaked dwarf virus*—China (RBSDV), genus *Fijivirus* (AJ294757); *Mal de Rio Cuarto virus* (MRCV), genus *Fijivirus* (AF499925); *Fiji disease virus* (FDV), genus *Fijivirus* (AY029520); *Nilaparvata lugens reovirus* (NLRV), genus *Fijivirus* (D49693); *Mammalian orthoreovirus* subgroup 1, serotype Dearing 3 (MRV-1), genus *Orthoreovirus* (M24734); *African horse sickness virus* serotype 9 (AHSV), genus *Orbivirus* (U94887); *Bluetongue virus* (BTV), genus *Orbivirus* (X12819); *Chuzan virus* (CZV), genus *Orbivirus* (AB018086); *Peruvian horse sickness virus* (PHSV), genus *Orbivirus* (DQ248057); *St. Croix River virus* (SCRV), genus *Orbivirus* (AF133431); *Yunnan orbivirus* (YUOV), genus *Orbivirus* (AY701509); *Human rotavirus* (HuRV-A), genus *Rotavirus* (A) (AB022765); *Bovine rotavirus* (BoRV-A), genus *Rotavirus* (A) (J04346); *Simian rotavirus* SA11 (SiRV-A), genus *Rotavirus* (A) (X16830); *Murine rotavirus* IDIR (MuRV-B), genus *Rotavirus* (B)

inverted repeat, 5' ... UUUUUCUUG (nt 5–13). ... CGA-GAAAAA (nt 839–847)...-3', was identified immediately adjacent to the conserved terminal sequences in genomic segment S12, which was segment-specific in its sequence and predicted secondary structure (MFOLD). The major ORF of RGDV S12 started at the first AUG initiation codon (nt 31–33) and stopped by a UAA stop codon at nt 649–651. The ORF potentially encodes a protein of 206 amino acids with a predicted Mr of 23.6-kDa and a pI of 10.2.

(M97203); *Porcine rotavirus* (PoRV-C), genus *Rotavirus* (C) (M74216); *Colorado tick fever virus* Florio N-7180 strain (CTFV), genus *Coltivirus* (AF133428); *Eyach virus* (EYAV), genus *Coltivirus* (AF282467); *Grass carp reovirus* (GCRV), genus *Aquareovirus* (AF284502); *Chum salmon reovirus* (CSV), genus *Aquareovirus* (AF418295); *Golden shiner reovirus* (GSRV), genus *Aquareovirus* (AF403399); *Lymantria dispar cypovirus 1* (CPV-1), genus *Cyporeovirus* (AF389463); *Lymantria dispar cypovirus 14* (CPV-14), genus *Cyporeovirus* (AF389452); *Trichoplusia ni cytoplasmic polyhedrosis virus 15* (CPV-15), genus *Cyporeovirus* (AF291683); *Cryphonectria parasitica mycoreovirus-1/9B21* (MYRV-1), genus *Mycoreovirus* (AF277888); *Rosellinia anti-rot virus* (MYRV-3), genus *Mycoreovirus* (AB102674); *Banna virus* (BAV), genus *Seadornavirus* (AF133430); *Kadipiro virus* (KDV), genus *Seadornavirus* (AF133429); *Liaoning virus* (LNSV), genus *Seadornavirus* (AY701339); *Diadromus pulchellus idnoreovirus 1* (DpRV-1), genus *Idnoreovirus* (X80481); *Aedes pseudoscutellaris reovirus* (APRV), proposed genus *Dinovernavirus* (DQ087277); *Operophtera brumata reovirus* (OpBuRV), Unassigned *Reoviridae* (DQ192235); *Eriochair sinensis reovirus* (ESRV) S1, proposed genus *Cardoreovirus* (AY542965); *Micromonas pusilla reovirus* (MPRV) S2, proposed genus *Mimoreovirus* (DQ126102)

When a homology search of the Swiss-Prot, PIR, GenPept and PDB databases was performed for the 23.6-kDa polypeptide of RGDV using BLASTP, we found the protein, just as has been identified for other RGDV non-structural proteins including p10 and p11 [22, 23], only shared significant (but low) similarity to two other phyto-reoviral proteins, namely Pns11, a 20.0-kDa non-structural protein of RDV, and Pns12, a 19.2-kDa non-structural protein of WTV. Other viral proteins, including those

belong to family *Reoviridae* so far registered in the database, did not exhibit any significant sequence similarity to the RGDV protein. Multiple sequence alignment further revealed that the deduced amino acid sequences of the three phyto-reovirus proteins could be aligned over their entire length, with 25–26% identity and 44–48% similarity between them, although the putative polypeptide encoded by RGDV S12 was 23 amino acids longer than that of RDV S11 and 28 amino acids longer than that of WTV S12 at the C-terminus. The N-terminal regions were much more conserved than the C-termini (Fig. 1B). Particularly in the region of aa 31–80, the amino acid sequence identity was up to 46–52% and the similarity was up to 70–74%. Phyto-reoviruses appear to be unique within the family *Reoviridae* in that they can multiply both in plants and in invertebrates and can be transmitted from infective females to their progeny through the eggs. Among proteins encoded by species of the genus *Phytoreovirus*, Pns11 of RDV and Pns12 of WTV have been detected both in host plants and in vector insects [42, 43] and are believed to be homologous [17]. It was suggested that the RDV Pns11 could play an important role in virus replication and/or genome assortment within the virus infection cycle [44]. The sequence similarity suggests that the 23.6-kDa polypeptide encoded by S12 of RGDV might be their counterpart and so might have similar functions in the propagation of the virus in both hosts.

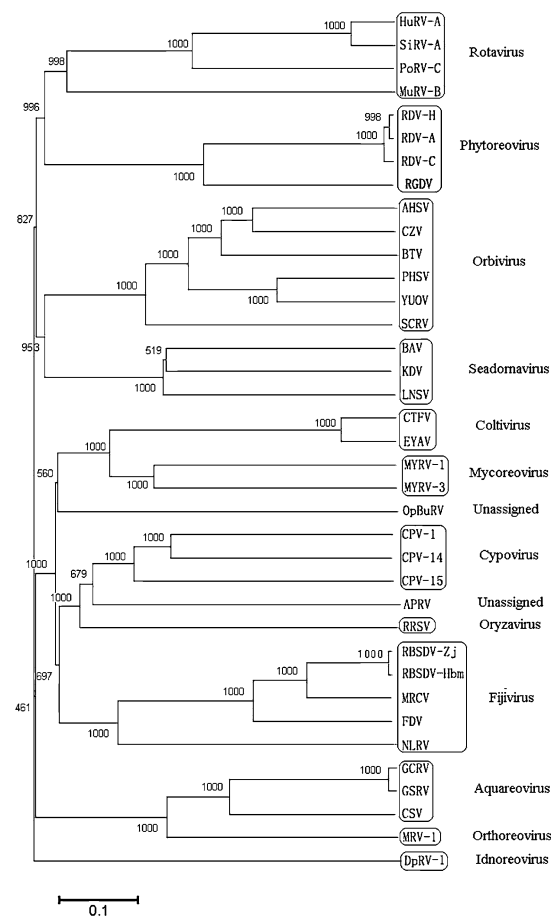
The C-terminus of RGDV p12 is also extremely hydrophilic and contains several (at least 7) basic regions. This is similar to the corresponding proteins of RDV and WTV although actual sequence conservation is small in this region. The RGDV p12, like pns11 of RDV and pns12 of WTV, has a pI of 10.2. At neutral pH, RGDV p12 would carry an overall charge of +14 while RDV pns11 and WTV pns12 would carry 16 and 14 positive charges, respectively. Recently, it has been reported that RDV pns11 can bind to both single-stranded and double-stranded RNA in a sequence non-specific manner [44]. These related results lead us to suggest that the protein encoded on RGDV S12, similar to that of RDV S11, might be a nucleic acid-binding protein. It is tempting to speculate that these proteins could bind ssRNA and dsRNA via electrostatic interactions in vivo.

A further feature in common between RGDV S12, S11 of RDV and S12 of WTV is the relatively long 3'-UTR (202 nucleotides; almost a quarter of the segment length). This is very unusual among reoviruses but its significance is not known.

#### Phylogenetic analysis

In order to study the phylogenetic relationships within the family *Reoviridae*, a phylogenetic tree was constructed

using the complete amino acid sequences of RdRp proteins of RGDV and all other reported viruses of the family by the UPGMA method (Fig. 3). The major branches of the tree, corresponding to established or proposed genera were well supported by bootstrap analysis ( $n = 1,000$ ). A similar result was obtained by Neighbor-joining (data not shown). The topology of the tree showed that RGDV closely clustered with RDV, another member of genus *Phytoreovirus*, and was more related to members of genus *Rotavirus* than to any other genus of family *Reoviridae*, supporting the conclusion that reoviruses infecting plants and animals have a common origin. Noticeably, three genera infecting plant hosts were only distantly related to each other. The phylogenetic analysis based on the RdRp also supported the status of RGDV as an independent species of the genus. However, the tree did not include WTV, the type member of the genus *Phytoreovirus*, for which there is no RdRp sequence. Phylogenetic analysis is less informative for p12 because of the scarcity of related sequences. However, it is interesting that although RDV and RGDV have the same



**Fig. 3** Phylogenetic tree based on the amino acid sequences of the RdRp proteins of RGDV and other reoviruses by the UPGMA method. Sequences and abbreviations are as in Fig. 2

plant host, they are no more closely related to one another (25.1–26.4% identical amino acids) than either of them is to WTV (RGDV: 27.8%, RDV 25.7–30.3% identity). This pattern of relationship between RGDV, RDV and WTV can also be seen with the other known RGDV proteins including the two structural proteins, p5 and p8, and the two non-structural proteins, p10 and p11 (data not shown).

## Conclusion

The complete sequences of RGDV S1 and S12 will contribute to progress on gene product assignment and functional analyses of the encoded viral proteins. Database searches for sequence similarities and motifs have supplied clues for the elucidation of locations of viral proteins (structural or non-structural proteins). For instance, RGDV RdRp is very likely to be the gene for a core protein, based on the fact that all reoviral RdRp proteins examined to date are constituents of inner capsids [45]. RGDV p12 is very likely to be a RNA-binding non-structural protein, given that the properties of its presumed counterparts, RDV p11 and WTV p12, have been identified. It will be interesting to examine the location and role of these proteins in contrast to those of the other reoviral proteins. These comparative sequence studies, in combination with ongoing genomic and proteomic studies, are intended to better delineate the roles of individual RGDV proteins in the viral replication cycle.

**Acknowledgments** This work was funded by the Zhejiang Provincial Foundation for Natural Science (Z305165) and two special grants from the Zhejiang Provincial Department of Science and Technology (2005E10010 and 2005C22027). We thank Mrs Liu Guimei and Mr Chen Hui, Station of Plant Protection and Quarantine, Agricultural Bureau of Chenxi county, Guangxi Province, for providing virus-infected rice plants. Rothamsted Research receives grant-aided support from the Biotechnology and Biological Sciences Research Council of the United Kingdom

## References

1. M. Putta, D. Chettanachit, T. Morinaka, D. Parejarearn, *Int. Rice Res. Newsl.* **5**(3), 10 (1980)
2. T. Omura, H. Inoue, T. Morinaka, Y. Saito, D. Chettanachit, M. Putta, A. Parejarearn, S. Disthaporn, *Plant. Dis.* **64**, 795 (1980)
3. C.A. Ong, T. Omura, *Int. Rice Res. Newsl.* **7**(2), 7 (1982)
4. H.Z. Fan, S.G. Zhang, X.Z. He, S.D. Xie, C.Z. Liu, L.G. Zhou, D. Zhu, X.R. Liu, *Acta Phytopath. Sin.* **13**, 1 (1983)
5. H. Inoue, T. Omura, *Plant. Dis.* **66**, 57 (1982)
6. T. Morinaka, H. Inoue, T. Omura, Y. Saito, M. Putta, D. Chettanachit, A. Parejarearn, S. Disthaporn, *Plant. Dis.* **66**, 703 (1982)
7. T. Hibi, T. Omura, Y. Saito, *J. Gen. Virol.* **65**, 1585 (1984)
8. G. Boccardo, R.G. Milne, S. Disthaporn, D. Chettanachit, M. Putta, *Intervirology* **23**, 167 (1985)
9. T. Omura, Y. Minove, M. Matsuoka, Y. Nozu, T. Tsuchizaki, Y. Saito, *J. Gen. Virol.* **66**, 811 (1985)
10. W. Maruyama, K. Ichimi, Y. Fukui, J. Yan, Y. Zhu, H. Kamiunten, T. Omura, *Arch. Virol.* **142**, 1 (1997)
11. Y. Takahashi, M. Tomiyama, H. Hibino, T. Omura, *J. Gen. Virol.* **75**, 269 (1994)
12. N. Miyazaki, K. Hagiwara, H. Naitow, T. Higashi, R.H. Cheng, T. Tsukihara, A. Nakagawa, T. Omura, *J. Mol. Biol.* **345**, 229 (2005)
13. T. Omura, H. Hibino, H. Inoue, T. Tsuchizaki, *J. Gen. Virol.* **66**, 2581 (1985)
14. J. Yan, M. Tomaru, A. Takahashi, I. Kimura, H. Hibino, T. Omura, *Virology* **224**, 539 (1996)
15. T. Omura, J. Yan, *Adv. Virus Res.* **54**, 15 (1999)
16. M. Yokoyama, Y. Nozu, J. Hashimoto, T. Omura, *J. Gen. Virol.* **65**, 533 (1984)
17. T. Omura, *Sem. Virol.* **6**, 97 (1995)
18. S. Zhu, F. Gao, X. Cao, M. Chen, G. Ye, C. Wei, Y. Li, *Plant Physiol.* **139**, 1935 (2005)
19. T. Wei, T. Shimizu, K. Hagiwara, A. Kikuchi, Y. Moriyasu, N. Suzuki, H. Chen, T. Omura, *J. Gen. Virol.* **87**, 429 (2006)
20. T. Wei, A. Kikuchi, N. Suzuki, T. Shimizu, K. Hagiwara, H. Chen, T. Omura, *Arch. Virol.* **151**, 1701 (2006)
21. H. Koganezawa, H. Hibino, F. Motoyoshi, H. Kato, H. Noda, K. Ishikawa, T. Omura, *J. Gen. Virol.* **71**, 1861 (1990)
22. H. Noda, K. Ishikawa, H. Hibino, H. Kato, T. Omura, *J. Gen. Virol.* **72**, 2837 (1991)
23. Y. Moriyasu, K. Ishikawa, A. Kikuchi, S. Imanishi, S. Tomita, K. Akutsu, T. Omura, *Virus. Genes.* **20**, 237 (2000)
24. K. Ichimi, A. Kikuchi, Y. Moriyasu, X.B. Zhong, K. Hagiwara, H. Kamiunten, T. Omura, *JARQ.* **36**, 83 (2002)
25. I. Uyeda, L. Bong-Choon, Y. Ando, in *Methods in molecular biology, vol 81. Plant virology protocols: from virus isolation to transgenic resistance*, ed. by G.D. Foster, S.C. Taylor (Humana Press, Totowa, 1998), p. 65
26. J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular cloning: a laboratory manual*. (Cold Spring Harbor Laboratory Press, New York, 1989)
27. H.M. Zhang, J.P. Chen, J.L. Lei, M.J. Adams, *Eur. J. Plant. Pathol.* **107**, 563 (2001)
28. H.M. Zhang, J.P. Chen, M.J. Adams, *Arch. Virol.* **146**, 2331 (2001)
29. S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, *Nucleic Acids Res.* **25**, 3389 (1997)
30. J.D. Thompson, D.G. Higgins, T.J. Gibson, *Nucleic Acids Res.* **22**, 4673 (1994)
31. F. Corpet, *Nucleic Acids Res.* **16**, 10881 (1988)
32. N. Saitou, M. Nei, *Mol. Biol. Evol.* **4**, 406 (1987)
33. S. Kumar, K. Tamura, M. Nei, *Briefings Bioinf.* **5**, 150 (2004)
34. J.V. Anzola, Z. Xu, T. Asamizu, D.L. Nuss, *Proc. Natl. Acad. Sci. USA* **84**, 8301 (1987)
35. N. Suzuki, *Sem. Virol.* **6**, 89 (1995)
36. N. Nakashima, M. Koizumi, H. Watanabe, H. Noda, *J. Gen. Virol.* **77**, 139 (1996)
37. Y.Z. Tao, D.L. Farsetta, M.L. Nibert, S.C. Harrison, *Cell* **111**, 733 (2002)
38. J.L. Hansen, A.M. Long, S.C. Schultz, *Structure* **5**, 1109 (1997)
39. E.K. O'Reilly, C.C. Kao, *Virology* **252**, 287 (1998)
40. D.R. Black, C.A. Knight, *J. Virol.* **6**, 194 (1970)
41. T. Kodama, N. Suzuki, *Ann. Phytopath. Soc. Japan* **39**, 251 (1973)
42. D.L. Nuss, A.J. Peterson, *J. Virol.* **34**, 532 (1980)
43. N. Suzuki, M. Sugawara, T. Kusano, H. Mori, Y. Matsuura, *Virology* **202**, 41 (1994)
44. H. Xu, Y. Li, Z.J. Mao, Y.Y. Li, Z.J. Wu, L. Qu, C.C. An, X.T. Ming, J. Schiemann, R. Casper, Z.L. Chen, *Virology* **240**, 267 (1998)



45. P.P.C. Mertens, M. Arella, H. Attoui, S. Bellonik, M. Bergoin, G. Boccardo, T.F. Booth, W. Chiu, J.M. Diprose, R. Duncan, in *Virus taxonomy: Seventh Report of the International Committee for the taxonomy of viruses*, ed. by M.H.V. Van Regenmortel, C.M. Fauquet, D.H.L. Bishop, C.H. Calisher, E.B. Carsten, M.K. Estes, S.M. Lemon, J. Maniloff, M.A. Mayo, D.J. McGeoch, C.R. Pringle, R.B. Wickner (Academic Press, New York, 2000), p. 395