Phylogenetic analysis of some large double-stranded RNA replicons from plants suggests they evolved from a defective single-stranded RNA virus

Mark J. Gibbs,1 Ryuichi Koga,2 Hiromitsu Moriyama,2† Pierre Pfeiffer3 and Toshiyuki Fukuhara2

1 Bioinformatics, Research School of Biological Sciences, The Australian National University, GPO Box 475, Canberra 2601, Australia
2 Laboratory of Molecular Cell Biology, Faculty of Agriculture, Tokyo University of Agriculture and Technology, Fuchu, Tokyo 183, Japan
3 Institute of Plant Molecular Biology, 12 Rue du Général Zimmer, 67000 Strasbourg, France

Sequences were recently obtained from four double-stranded (ds) RNAs from different plant species. These dsRNAs are not associated with particles and as they appeared not to be horizontally transmitted, they were thought to be a kind of RNA plasmid. Here we report that the RNA-dependent RNA polymerase (RdRp) and helicase domains encoded by these dsRNAs are related to those of viruses of the alpha-like virus supergroup. Recent work on the RdRp sequences of alpha-like viruses raised doubts about their relatedness, but our analyses confirm that almost all the viruses previously assigned to the supergroup are related. Alpha-like viruses have single-stranded (ss) RNA genomes and produce particles, and they are much more diverse than the dsRNAs. This difference in diversity suggests the ssRNA alpha-like virus form is older, and we speculate that the transformation to a dsRNA form began when an ancestral ssRNA virus lost its virion protein gene. The phylogeny of the dsRNAs indicates this transformation was not recent and features of the dsRNA genome structure and translation strategy suggest it is now irreversible. Our analyses also show some dsRNAs from distantly related plants are closely related, indicating they have not strictly co-speciated with their hosts. In view of the affinities of the dsRNAs, we believe they should be classified as viruses and we suggest they be recognized as members of a new virus genus (Endornavirus) and family (Endoviridae).

Introduction

Complete sequences have been obtained from large double-stranded (ds) RNA species found in cultivated rice (Oryza sativa ssp. Japonica; Moriyama et al., 1995), wild rice (Oryza rufipogon; Moriyama et al., 1999) and broad bean (Vicia faba cv. 447; Pfeiffer, 1998) and a partial sequence has been obtained for one of two possible species found in kidney bean (Phaseolus vulgaris cv. black turtle soup; Wakarchuk & Hamilton, 1990). Similar large dsRNAs have been found in a few other plant species, and together these dsRNAs have been called ‘endogenous dsRNAs’ (Moriyama et al., 1996, 1999). They have an unusual combination of properties. They are similar to cryptoviruses (Ghabrial et al., 1995) in that they are efficiently transmitted through seed, no horizontal spread has been observed in the field, no potential vectors have been identified and none is associated with disease symptoms, except for one associated with sterility (Pfeiffer et al., 1993). However, unlike cryptoviruses, which produce particles and have a genome consisting of two dsRNAs each about 2–3 kb long, none of the endogenous dsRNAs is associated with particles and each is more than 10 kb long. Consistent with a lack of particles, none of the endogenous dsRNAs is mechanically transmissible.

Work on the dsRNA from O. sativa showed that it is not encoded by the host and that its replication is regulated to produce a constant low concentration in all tissues except pollen (Fukuhara et al., 1993; Moriyama et al., 1996, 1999). Some of the other dsRNAs have also been found at a constant, low concentration in their hosts, suggesting that this is a group characteristic (Wakarchuk & Hamilton, 1985; Gabriel et al., 1987; Fairbanks et al., 1988; Valverde & Fontenot, 1990;
Zabalgogeazcoa & Gildow, 1992). This regulation, the apparent non-infectious nature of the dsRNAs, and their lack of particles led Fukuhara et al. (1993) to suggest that these organisms are a kind of RNA plasmid. Studies on the V. faba dsRNA showed that it does not even spread from cell to cell except at cell division (Duc et al., 1984), supporting this view. Fukuhara et al. (1993) also suggested that the endogenous dsRNAs might be related to hypoviruses as these fungal viruses similarly lack particles and are transmitted vertically but not horizontally except by hyphal fusion. Hypoviruses also have large dsRNA genomes and they too were once considered to be a kind of RNA plasmid (Brown & Finnegan, 1989; Shapira et al., 1991).

Here we report an analysis of the sequences of the endogenous dsRNAs that clarifies their relationships, provides new information on their evolution and supports a clear argument for their classification. We also provide new evidence that supports grouping a wide set of viruses, including the dsRNAs, into the previously debated alpha-like virus supergroup.

**Methods**

The non-redundant coding and nucleotide sequence databases were searched using the programs BLASTP, TBLASTN and PSI-BLAST (Altschul et al., 1997) and Expect values (E values) were calculated by these programs for pairs of aligned sequences. The evolutionary relatedness (homology) of sequences was judged firstly using these E values. E values are estimates of the probability of finding the same degree of similarity by chance given the composition of the aligned sequences and the database. E values less than $1 \times 10^{-10}$ usually indicate homology and those less than $1 \times 10^{-6}$ almost always indicate homology (Altschul et al., 1997; Bork & Gibson, 1996). The similarity of pairs of sequences was also measured in percent identity plots made using the program PLOTSIMILARITY (Devereux et al., 1984) and using the program ALIGN (Dayhoff et al., 1983), which was used to make a second test of relatedness. ALIGN performs a Monte Carlo procedure to calculate a Z-score which is largely unbiased by sequence composition and length. Usually a Z-score of 3 or more would be taken to be significant, but because of the constraints on real proteins, alignments of their sequences produce biased scores centred on 3 rather than 0 (Barton & Sternberg, 1987). Thus, it is generally accepted that Z-scores of 5 to 6 indicate relatedness (Barton, 1996) when produced by this method. We did 100 alignments from randomized sequences for each pairwise comparison made with the program ALIGN and used the MDM$_s$ distance matrix (Dayhoff et al., 1978) and a gap penalty of 4 in the calculations.

Multiple alignments of amino acid and nucleotide sequences were made with the program CLUSTALW (Thompson et al., 1994). Maximum likelihood trees were found from the amino acid multiple alignments by quartet puzzling using the program PUZZLE version 4 (Strimmer & von Haeseler, 1996) after positions including gaps had been excluded. Likelihoods were calculated using the BLOSUM 62 substitution matrix (Henikoff & Henikof, 1992) and a gamma distribution of rates of change for variable sites with a shape parameter estimated from the data using a neighbour-joining tree. Maximum likelihood and most parsimonious trees were also found from the aligned nucleotide sequences by heuristic searching with the program PAUP version 4.0d4 (written by David L. Swofford) after positions including gaps and third codon positions had been excluded. Bootstrap values were calculated from neighbour-joining trees inferred from 1000 bootstrap samples.

**Results and Discussion**

**Intra-group homology and phylogeny**

Each of the three completely sequenced dsRNAs encodes a single long open reading frame which we will call the long protein (LP) gene. Comparisons of the LP amino acid sequences from the O. sativa and V. faba dsRNAs showed that the likely helicase and RNA-dependent RNA polymerase (RdRp) domains correspond to two stretches of higher identity (Fig. 1, bottom curve). These two regions were also identified in database searches (Fig. 1). An E value of $3 \times 10^{-22}$ was

![Fig. 1. Percent identity plots found using a window 100 amino acid residues long for an alignment of the long proteins (LPs) of the two Oryza dsRNAs (top curve) and for an alignment of the LPs of the O. sativa and V. faba dsRNAs (lower curve). The upper curve is scaled to the upper x-axis and the lower curve is scaled to the lower x-axis. The conserved GKT/S motif (marked with open triangles) is close to the beginning of the helicase domain, and the conserved GDD motif (marked with open circles) is close to the end of the RdRp domain. Black filled blocks mark similar regions between the V. faba and O. sativa dsRNAs identified by a BLASTP search. E values obtained from the alignments of these regions are shown. Note that the LPs of the Oryza dsRNAs are about 4600 amino acid residues long and the LP of the V. faba dsRNA is 5825 residues long.](image-url)
Phylogeny of large dsRNAs from plants

obtained for an alignment 670 residues long between the regions that include the helicase motifs from the V. faba and O. sativa dsRNA LPs. An E value of $2 \times 10^{-75}$ was obtained for an alignment 370 residues long between the regions that include the RdRp motifs from these same two LP sequences. These two database search results independently indicated that the Oryza and V. faba dsRNAs share a common ancestor.

Database searches with the TBLASTN program using LP sequences from the Oryza dsRNAs identified similarities with the partial sequence of a dsRNA from P. vulgaris. An alignment of the P. vulgaris polypeptide with the LP from the O. sativa dsRNA yielded an E value of $1 \times 10^{-10}$. By contrast, database searches with the LP sequence from the V. faba dsRNA did not detect similarities with the P. vulgaris dsRNA sequence. We confirmed this difference in affinities using the Monte Carlo procedure. Z-scores of 12.5 and 11.2 were obtained for comparisons between amino acid sequences encoded by the P. vulgaris and Oryza dsRNAs, but a Z-score of only 4.2 was obtained for the comparison of the P. vulgaris dsRNA sequence and the equivalent sequence from the V. faba dsRNA. A maximum-likelihood tree inferred from these sequences concurred with this result (Fig. 2, top). As our sequence comparisons showed that the V. faba and Oryza dsRNAs have a common ancestor, and that the P. vulgaris and Oryza dsRNAs are similarly related, we conclude that the P. vulgaris and V. faba dsRNAs are also related. A multiple alignment supported this conclusion as it showed that 22% of the amino acid residues were strictly or strongly conserved in the equivalent P. vulgaris, Oryza and V. faba dsRNA LP sequences. Our conclusion is also supported by the work of Wakarchuk & Hamilton (1985), who showed that the P. vulgaris dsRNA shares several properties with the V. faba and Oryza dsRNAs. Together these data suggest the genome of the P. vulgaris dsRNA is similar to those of the fully sequenced dsRNAs and on this basis we suggest the P. vulgaris dsRNA should be placed in the same taxonomic family.

Experiments suggest that the dsRNAs are only vertically transmitted through seed, either through ovules (Duc et al., 1984) or through both ovules and pollen (Moriyama et al., 1996). Attempts have been made to transmit them mechanically, by grafting, using aphids and using dodder, but all have failed (Turpen et al., 1988; Valverde et al., 1990; Zabalgogaezcoa & Gildow, 1992). An implication of this likely limit on spread is that the dsRNAs should co-speciate with their hosts. Endogenous dsRNAs from closely related plants

clump pecluvirus; PVM, Potato M carlavirus; PVV, Phaseolus vulgaris endornavirus; PVX, Potato X potexvirus; RBDV, Raspberry bushy dwarf idaeovirus; RuBV, Rubella rubivirus; SBWMV, Soil-borne wheat mosaic furovirus; SINV, Sindbis alphavirus; TMV, Tobacco mosaic tobamovirus; TRV, Tobacco rattle tobamovirus; TSV, Tobacco streak ilarvirus; TYMV, Turnip yellow mosaic tymovirus; VFV, Vicia faba endornavirus. The inset (top) shows a mid-point rooted tree for the dsRNAs inferred from the partial sequence from the P. vulgaris dsRNA and the equivalent sequences from the other dsRNAs. Terminal nodes for the dsRNAs have been labelled with the acronyms we have suggested.
should be more closely related to each other than to endogenous dsRNAs from distantly related plants. Hence, the relationships between the *Oryza*, *P. vulgaris* and *V. faba* dsRNAs were unexpected: *Phaseolus* and *Vicia* belong to the legume subfamily *Papilionoideae* and are only distant relatives of the *Oryza* species.

**RNA-dependent RNA polymerase affinities**

Surprisingly BLASTP database searches revealed similarities between the RdRp-like regions in the LP sequences and the RdRp sequences of several single-stranded (ss) RNA viruses. E values as low as $7 \times 10^{-10}$ were obtained for alignments about 300 residues long with RdRp sequences from clustroviruses. Alignments with RdRp sequences from *Hepatitis E virus* (HEV) and *Alfalfa mosaic alfamovirus* (AMV) yielded the next lowest E values, starting at $1 \times 10^{-4}$ and $2 \times 10^{-4}$ respectively.

It has been suggested that the clustroviruses and alfamoviruses belong to the ‘alpha-like supergroup’ of positive-sense ssRNA viruses. This supergroup is said to also include viruses from the following genera: *Alphavirus*, *Benyvirus*, *Bromovirus*, *Crinivirus*, *Cucumovirus*, *Capillovirus*, *Carlavirus*, *Furovirus*, * Hordeivirion*, *Idaeovirus*, *Marafivirus*, *Pecluvirus*, *Pomovirus*, *Potexvirus*, *Rubivirion*, *Tetravirus*, *Tobamovirus*, *Tobravirus*, *Trichovirus*, *Tymovirus* and *Vitivirus*. The supergroup was defined on the basis of several features (Goldbach & de Haan, 1994; Koonin & Dolja, 1993), but only some of these are shared by all the proposed members: (i) a positive-strand RNA genome with a 5’ cap, (ii) production of a subgenomic RNA encoding a virion protein, (iii) homologous RdRp and helicase amino acid sequences. Viruses that are not assigned to the supergroup share the first two of these features and hence, the supergroup is probably best defined only by the proposed homology of the helicase and RdRp sequences.

Zanotto *et al.* (1996) tested the support from RdRp sequence alignments for several proposed supergroups of RNA viruses and found only weak support for the alpha-like supergroup. We repeated part of this work by compiling a dataset that included at least one RdRp sequence from each genus in the supergroup for which sequence is available, and then by obtaining Z-scores by the Monte Carlo procedure. The RdRp-like regions of the endogenous dsRNAs were included in the analysis to test their relationships, as was the RdRp sequence of HEV. Sequences from carmo-like and picorna-like viruses were also included as they were likely to be outliers. As shown in Table 1, Z-scores of more than 6 were obtained for many of the comparisons. These scores proved the relatedness of almost all the RdRp sequences from viruses previously assigned to the alpha-like supergroup and showed that the RdRp sequences of the dsRNAs and HEV are related to alpha-

---

**Table 1. Z-scores from pairwise comparisons of RdRp sequences**

The first three columns show scores from comparisons involving the three major clusters (see text). The lowest scores obtained from within-subset comparisons are shown in brackets. The remaining scores in the first three columns are the highest scores obtained from comparisons between RdRp sequences from the clusters and ones from ungrouped viruses and outgroups. Scores that clearly indicate relatedness are shown in bold. All comparisons were made using a sequence starting about 170 residues on the N-terminal side of the GDD motif and ending about 60 residues on its C-terminal side. Sequences from the five carmo-like viruses named in Fig. 2 were used, as were sequences from the following picorna-like viruses: *Cowpea mosaic comovirus*, *Hepatitis A hepatovirus*, *Human rhinovirus 89*, *Rice tungro spherical waikavirus* and *Tobacco ringspot nepovirus*. A complete list of acronyms is given in the legend to Fig. 2.

<table>
<thead>
<tr>
<th>Virus (Group)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Alfamoviruses etc</td>
<td>(10.9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Capillovirus etc</td>
<td>(9.0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. dsRNAs</td>
<td>12.4</td>
<td>7.5</td>
<td>(20.0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. SINV</td>
<td>10.9</td>
<td>6.9</td>
<td>5.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. RuBV</td>
<td>7.3</td>
<td>5.6</td>
<td>8.9</td>
<td>5.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. HEV</td>
<td>7.3</td>
<td>8.6</td>
<td>9.4</td>
<td>4.8</td>
<td>5.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. BNYVV</td>
<td>5.6</td>
<td>2.4</td>
<td>2.0</td>
<td>4.0</td>
<td>4.2</td>
<td>5.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. HaSV</td>
<td>6.0</td>
<td>6.1</td>
<td>4.6</td>
<td>4.0</td>
<td>4.4</td>
<td>10.4</td>
<td>4.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Carmo-like viruses</td>
<td>&lt;3.0</td>
<td>&lt;3.0</td>
<td>&lt;3.0</td>
<td>&lt;3.0</td>
<td>&lt;3.0</td>
<td>&lt;3.0</td>
<td>&lt;3.0</td>
<td>&lt;3.0</td>
<td>&lt;3.0</td>
</tr>
<tr>
<td>10. Picorna-like viruses</td>
<td>&lt;2.6</td>
<td>&lt;2.6</td>
<td>&lt;2.6</td>
<td>&lt;2.6</td>
<td>&lt;2.6</td>
<td>&lt;2.6</td>
<td>&lt;2.6</td>
<td>&lt;2.6</td>
<td>&lt;3.1</td>
</tr>
</tbody>
</table>
like virus sequences. Only the affinities of the RdRp sequence of *Beet necrotic yellow vein benyovirus* (BNYVV), which was previously assigned to the alpha-like supergroup, were doubt-ful. The highest Z-score obtained for a comparison between the BNYVV sequence and an alpha-like virus sequence was only 5.8. Low Z-scores were obtained from comparisons between alpha-like virus RdRp sequences and those of carmo-like and picorna-like viruses, supporting the view that the alpha-like virus RdRp sequences are a discrete natural grouping.

Phylogenetic trees were constructed for the RdRp amino acid sequences (Fig. 2). Strong support was found in the trees for two major clusters that included RdRp sequences from the following viruses: (1) alfamoviruses, bromoviruses, clustro-viruses, criniviruses, cucumoviruses, furoviruses, hordeiviruses, icaeoviruses, iarviruses, pelcluviruses, pomoviruses, tobamoviruses and tobaviruses and (2) capilloviruses, carlaviruses, marafiviruses, potexviruses, trichoviruses, tymoviruses and vitiviruses. There was also strong support for grouping the sequences from the dsRNAs, which we have designated cluster 3 in the tree and Table 1. The sequences of BNYVV, *Helicoverpa armigera stunt tetravirus* (HaSV), HEV, *Rubella rubivirus* (RuBV) and *Sindbis alphavirus* (SINV) were grouped in the maximum likelihood tree, but there was only weak support for this cluster and it was not found in trees inferred by other methods. Clusters 1, 2 and 3 (defined above) were supported by the Monte Carlo procedure results (Table 1), but these results did not support the grouping of BNYVV, HaSV, HEV, RuBV and SINV. The Monte Carlo results suggested that the RdRps of the dsRNAs are most closely related to those of cluster 1 and this possibility was supported by the database searches, but the phylogenetic analysis did not resolve the relationships between the clusters. The differences between our Monte Carlo randomization results and those of Zanotto et al. (1996) probably result from comparing slightly different regions of the RdRp sequences and using different parameters for the alignments. We used a region of sequence identified by the program BLASTp as high-scoring. This region began about 170 residues on the N-terminal side of the GDD motif and ended about 60 residues on its C-terminal side. Zanotto et al. (1996) tested various regions of sequence identified in earlier publications.

Zanotto et al. (1996) not only challenged the relatedness of the alpha-like viruses but also that they are monophyletic. By contrast, our phylogenetic analysis and Monte Carlo randomization results support the monophyly of the RdRp sequences from the alpha-like supergroup if this grouping includes the sequences from the dsRNAs and HEV (Fig. 2 and Table 1). One database search result challenged this conclusion: an alignment between an alpha-like RdRp sequence and that of the carmo-like virus *Maize chlorotic mottle machlomovirus* yielded an E value that could be significant ($8 \times 10^{-5}$), but we could not discern a clear phylogenetic signal that could link the carmo-like and alpha-like viruses (Table 1), and the carmo-like RdRp sequences are clearly outliers to the supergroup in the tree (Fig. 2). It is important to note that the actual estimates for the lengths of the branches leading to the carmo-like outlier sequences are double those shown (Fig. 2). Plainly our conclusions are contrary to some of those of Zanotto et al. (1996), but we have only been concerned with the alpha-like supergroup. The fact that none of our database searches and analyses supported links to sequences other than those from the mentioned viruses and dsRNAs supports the theory of Zanotto et al. (1996) that there is no phylogenetic signal that could be used to group all extant RdRp sequences.

**Helicase sequence affinities**

A profile search using the PSI-BLAST and the helicase-like sequences from the *V. faba* and *O. sativa* dsRNAs detected similarities with helicase sequences from viruses in every genus in the alpha-like supergroup for which sequence is available. Comparisons with tobamovirus sequences produced the highest E values, $2 \times 10^{-8}$, in the first iteration. Reciprocal searches made with the helicase sequence of *Tobacco mosaic tobamovirus* identified the *V. faba* dsRNA helicase in the first iteration with an E value of $5 \times 10^{-8}$. Z-scores generated from a helicase sequence dataset compiled from alpha-like virus sequences were lower than those from the RdRp sequence dataset. A comparison of the *V. faba* dsRNA helicase with the helicase domain of *Raspberry bushy dwarf icaeovirus* yielded a Z-score of 7.1, as did a comparison with the helicase domain of *Beet soil-borne furovirus*. Tobamoviruses, icaeoviruses and furoviruses are placed in cluster 1 in the alpha-like virus RdRp tree (Fig. 2) and hence, these results support the notion that the dsRNAs may be most closely related to this subset within the supergroup.

**The dsRNAs probably evolved from a defective alpha-like virus**

Our results show that the endogenous dsRNAs share a common ancestor with alpha-like viruses. The RdRps and helicases encoded by the dsRNAs are probably functionally similar to their alpha-like virus counterparts, but there are differences between the modes of replication of these organisms. Alpha-like virus genomic RNA is present in host cells primarily as messenger-sense ssRNA (positive-strand RNA), but no full-length positive-strand RNA from an endogenous dsRNA has been unequivocally identified (Pfeiffer et al., 1993; Fukuhara et al., 1995) and work on the *V. faba* and *P. vulgaris* dsRNAs suggests these organisms do not produce full-length ssRNAs (Pfeiffer et al., 1993; Wakarchuk & Hamilton, 1985; Lefebvre et al., 1990). Possibly more significant is the fact that each of the three fully sequenced dsRNAs includes a break (discontinuity or nick) on the coding strand but not the negative strand (Pfeiffer et al., 1993; Fukuhara et al., 1995; Moriyama et al., 1999). The conservation of the break and the fact that the LP gene continues in-frame through the break in all three molecules, suggests the break is somehow causally linked to the dsRNA form.
The relationship between the endogenous dsRNAs and the ssRNA alpha-like viruses suggests that either the dsRNAs evolved from an ancestral ssRNA virus or vice versa. The phylogenetic trees found with the RdRp and helicase sequences support the first of these options, as they show the alpha-like viruses to be far more diverse than the dsRNAs (Fig. 2). If this is true, as we believe, then as all alpha-like viruses have ssRNA genomes and encode their own virion proteins and produce particles, it is likely that an ancestor of the dsRNAs also had these features. Two alternative explanations for the difference in diversity should be considered but may be dismissed. First, the difference could be because we do not have a fair measure of the diversity of the dsRNAs, but this is unlikely in view of the similarities between all the known endogenous dsRNAs and the evolutionary distance between the Oryza and V. faba dsRNAs. Second, the dsRNAs could be evolving far slower than the alpha-like viruses, but this is unlikely as it would mean that the dsRNAs were subjected to much stronger selection pressures than the alpha-like viruses and that this difference in selection has been maintained over long periods.

We found no sequence similarities with known virion proteins, suggesting the dsRNAs do not encode a virion protein, which correlates with the fact that the dsRNAs are not associated with particles. This unusual characteristic could be related to the evolution of the dsRNA form, as the relative accumulation of the positive-sense and complementary-sense (negative) RNA strands is known to be affected when positive-strand RNA plant viruses lose the function of their virion protein genes (Nassuth & Bol, 1983; French & Ahlquist, 1988; van der Kuyl et al., 1991). The balance of positive to negative RNA produced by an ancestor of the dsRNAs may have similarly altered when it lost its virion protein gene. The transformation to a dsRNA form probably became irreversible when the dsRNAs evolved the break in the positive strand.

### Taxonomy

The endogenous dsRNAs could be classified as sub-viral agents because they lack particles, but the distinction between viruses and sub-viral agents is not entirely clear. The bacteriophage P4 and dependoviruses are classified as both sub-viral agents and viruses, and umbraviruses and hypoviruses are classified as viruses but not sub-viral agents, although they do not produce particles. One possible explanation for this confusion is that the dependoviruses, hypoviruses and umbraviruses, and the bacteriophage P4, are all related to conventional viruses (Mayo et al., 1995; Murant et al., 1995; Hillman et al., 1995). We know of no other sub-viral agents with this property and we view this situation as precedence for recognizing evolutionary relationships at this level of classification. On this basis we propose that the endogenous dsRNAs should be recognized as viruses because of their affinities to alpha-like viruses. As it is clear that the dsRNAs form a distinct group, we propose that they be nominated as members of a new virus genus and we suggest the name *Endornavirus* (endo, from Greek: within, and RNA) for this genus. The dsRNAs have no homology to recognized dsRNA viruses and because they are very different from other alpha-like viruses, we propose that they be assigned to a separate family, for which we suggest the name *Endoviridae*. To achieve continuity with the literature we also suggest that endornavirus species be named after the plant species in which they were first found. Thus, the dsRNAs for which we have sequence could be named *Plaseolus vulgaris endornavirus* (PVuV), *Oryza rufoapter endornavirus* (ORV), *Oryza sativa endornavirus* (OSV) and *Vicia faba endornavirus* (VFV).

### References


Phylogeny of large dsRNAs from plants


Received 7 June 1999; Accepted 10 September 1999